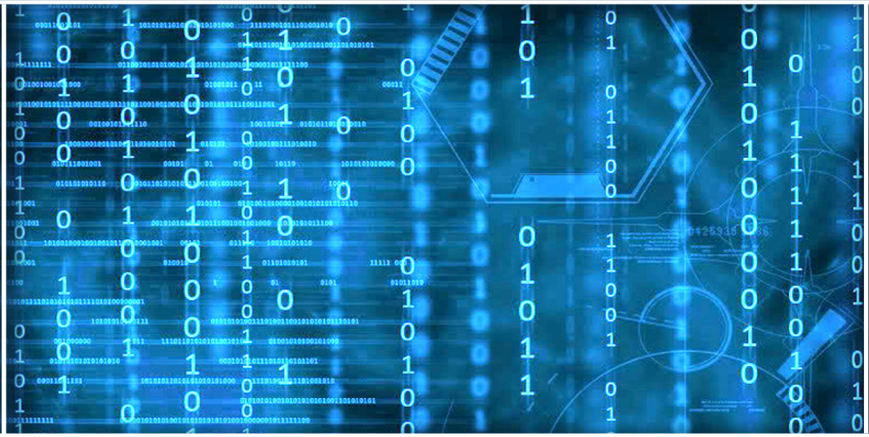


Volume 16 Issue 1

January 2025



ISSN 2156-5570(Online)

ISSN 2158-107X(Print)

Editorial Preface

From the Desk of Managing Editor...

It may be difficult to imagine that almost half a century ago we used computers far less sophisticated than current home desktop computers to put a man on the moon. In that 50 year span, the field of computer science has exploded.

Computer science has opened new avenues for thought and experimentation. What began as a way to simplify the calculation process has given birth to technology once only imagined by the human mind. The ability to communicate and share ideas even though collaborators are half a world away and exploration of not just the stars above but the internal workings of the human genome are some of the ways that this field has moved at an exponential pace.

At the International Journal of Advanced Computer Science and Applications it is our mission to provide an outlet for quality research. We want to promote universal access and opportunities for the international scientific community to share and disseminate scientific and technical information.

We believe in spreading knowledge of computer science and its applications to all classes of audiences. That is why we deliver up-to-date, authoritative coverage and offer open access of all our articles. Our archives have served as a place to provoke philosophical, theoretical, and empirical ideas from some of the finest minds in the field.

We utilize the talents and experience of editor and reviewers working at Universities and Institutions from around the world. We would like to express our gratitude to all authors, whose research results have been published in our journal, as well as our referees for their in-depth evaluations. Our high standards are maintained through a double blind review process.

We hope that this edition of IJACSA inspires and entices you to submit your own contributions in upcoming issues. Thank you for sharing wisdom.

Thank you for Sharing Wisdom!

Kohei Arai
Editor-in-Chief
IJACSA
Volume 16 Issue 1 January 2025
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

Editor-in-Chief

Dr. Kohei Arai - Saga University

Domains of Research: Technology Trends, Computer Vision, Decision Making, Information Retrieval, Networking, Simulation

Associate Editors

Alaa Sheta

Southern Connecticut State University

Domain of Research: Artificial Neural Networks, Computer Vision, Image Processing, Neural Networks, Neuro-Fuzzy Systems

Arun Kulkarni

University of Texas at Tyler

Domain of Research: Machine Vision, Artificial Intelligence, Computer Vision, Data Mining, Image Processing, Machine Learning, Neural Networks, Neuro-Fuzzy Systems

Domenico Ciunzo

University of Naples, Federico II, Italy

Domain of Research: Artificial Intelligence, Communication, Security, Big Data, Cloud Computing, Computer Networks, Internet of Things

Dr Ronak AL-Haddad

Anglia Ruskin University / Cambridge

Domain of Research : Technology Trends, Communication, Security, Software Engineering and Quality, Computer Networks, Cyber Security, Green Computing, Multimedia Communication, Network Security, Quality of Service

Elena Scutelnicu

"Dunarea de Jos" University of Galati

Domain of Research: e-Learning, e-Learning Tools, Simulation

In Soo Lee

Kyungpook National University

Domain of Research: Intelligent Systems, Artificial Neural Networks, Computational Intelligence, Neural Networks, Perception and Learning

Renato De Leone

Università di Camerino

Domain of Research: Mathematical Programming, Large-Scale Parallel Optimization, Transportation problems, Classification problems, Linear and Integer Programming

Xiao-Zhi Gao

University of Eastern Finland

Domain of Research: Artificial Intelligence, Genetic Algorithms

CONTENTS

Paper 1: Advanced Machine Learning Approaches for Accurate Migraine Prediction and Classification

Authors: Chokri Baccouch, Chaima Bahar

PAGE 1 – 11

Paper 2: A Comparative Study of Predictive Analysis Using Machine Learning Techniques: Performance Evaluation of Manual and AutoML Algorithms

Authors: Karim Mohammed Rezaul, Md. Jewel, Anjali Sudhan, Mifta Uddin Khan, Maharage Roshika Sathsarani Fernando, Kazy Noor e Alam Siddiquee, Tajnuva Jannaf, Muhammad Azizur Rahman, Md Shabiul Islam

PAGE 12 – 31

Paper 3: Detection of DDoS Cyberattack Using a Hybrid Trust-Based Technique for Smart Home Networks

Authors: Oghenetajiri Okporokpo, Funminiyi Olajide, Nemitari Ajenka, Xiaoqi Ma

PAGE 32 – 41

Paper 4: Forecasting the Emergence of a Dominant Design by Classifying Product and Process Patents Using Machine Learning and Text Mining

Authors: Koji Masuda, Yoshinori Hayashi, Shigeyuki Haruyama

PAGE 42 – 48

Paper 5: Control Interface for Multi-User Video Games with Hand or Head Gestures in Directional Key-Based Games

Authors: Oscar Ramirez-Valdez, César Baluarte-Araya, Rodrigo Castillo-Lazo, Italo Ccoscco-Alvis, Alexander Valdiviezo-Tovar, Alexander Villafuerte-Quispe, Dylan Zuñiga-Huraca

PAGE 49 – 60

Paper 6: Teaching Programming in Higher Education: Analyzing Trends, Technologies, and Pedagogical Approaches Through a Bibliometric Lens

Authors: Mariuxi Vinuesa-Morales, Jorge Rodas-Silva, Cristian Vidal-Silva

PAGE 61 – 68

Paper 7: Harnessing the Power of Federated Learning: A Systematic Review of Light Weight Deep Learning Protocols

Authors: Haseeb Khan Shinwari, Riaz Ul Amin

PAGE 69 – 78

Paper 8: SEC-MAC: A Secure Wireless Sensor Network Based on Cooperative Communication

Authors: Yassmin Khairat, Tamer O. Diab, Ahmed Fawzy, Samah Osama, Abd El- Hady Mahmoud

PAGE 79 – 87

Paper 9: Digital Twin Model from Freehanded Sketch to Facade Design, 2D-3D Conversion for Volume Design

Authors: Kohei Arai

PAGE 88 – 95

Paper 10: Marked Object-Following System Using Deep Learning and Metaheuristics

Authors: Ken Gorro, Elmo Ranolo, Lawrence Roble, Rue Nicole Santillan, Anthony Ilano, Joseph Pepito, Emma Sacan, Deofel Balijon

PAGE 96 – 106

Paper 11: Hawk-Eye Deblurring and Pose Recognition in Tennis Matches Based on Improved GAN and HRNet Algorithms

Authors: Weixin Zhao

PAGE 107 – 118

Paper 12: A Highly Functional Ensemble of Improved Chaos Sparrow Search Optimization Algorithm and Enhanced Sun Flower Optimization Algorithm for Query Optimization in Big Data

Authors: Mursubai Sandhya Rani, N. Raghavendra Sai

PAGE 119 – 134

Paper 13: IT Spin-Offs Challenges in Developing Countries

Authors: Mahmoud M. Musleh, Ibrahim Mohamed, Hasimi Sallehudin, Hussam F. Abushawish

PAGE 135 – 142

Paper 14: Multi-Factors Analysis Using Visualizations and SHAP: Comprehensive Case Analysis of Tennis Results Forecasting

Authors: Yuan Zhang

PAGE 143 – 152

Paper 15: Exploring Diverse Conventional and Deep Linguistic Features for Sentiment Analysis of Online Content

Authors: Yajun Tang

PAGE 153 – 162

Paper 16: An AI-Driven Approach for Advancing English Learning in Educational Information Systems Using Machine Learning

Authors: Xue Peng, Yue Wang

PAGE 163 – 171

Paper 17: Investigating Immersion and Presence in Virtual Reality for Architectural Visualization

Authors: Athira Azmi, Sharifah Mashita Syed Mohamad

PAGE 172 – 179

Paper 18: Data Mining MRO-BP Network-Based Evaluation Effectiveness of Music Teaching

Authors: Yifan Fan

PAGE 180 – 189

Paper 19: Employing Data-Driven NOA-LSSVM Algorithm for Indoor Spatial Environment Design

Authors: Di Wang, Hui Ma, Tingting Lv

PAGE 190 – 200

Paper 20: Enhancing Customer Churn Prediction Across Industries: A Comparative Study of Ensemble Stacking and Traditional Classifiers

Authors: Nurul Nadzirah bt Adnan, Mohd Khalid Awang

PAGE 201 – 208

Paper 21: Hotspots and Insights on Quality Evaluation of Study Tours: Visual Analysis Based on Bibliometric Methodology

Authors: Meihua Deng

PAGE 209 – 220

Paper 22: Internet of Things (IoT) Driven Logistics Supply Chain Management Coordinated Response Mechanism

Authors: Chong Li

PAGE 221 – 232

Paper 23: Big Data Analytics of Knowledge and Skill Sets for Web Development Using Latent Dirichlet Allocation and Clustering Analysis

Authors: Karina Djunaidi, Dine Tiara Kusuma, Rahma Farah Ningrum, Puji Catur Siswipraptini, Dina Fitria Murad

PAGE 233 – 244

Paper 24: Optimizing Multi-Dimensional SCADA Report Generation Using LSO-GAN for Web-Based Applications

Authors: Fanxiu Fang, Guocheng Qi, Haijun Cao, He Huang, Lingyi Sun, Jingli Yang, Yan Sui, Yun Liu, Dongqing You, Wenyu Pei

PAGE 245 – 256

Paper 25: Optimizing Decentralized Exam Timetabling with a Discrete Whale Optimization Algorithm

Authors: Emily Sing Kiang Siew, San Nah Sze, Say Leng Goh

PAGE 257 – 265

Paper 26: Fusion of Multimodal Information for Video Comment Text Sentiment Analysis Methods

Authors: Jing Han, Jinghua Lv

PAGE 266 – 274

Paper 27: Enhancing Stock Market Forecasting Through a Service-Driven Approach: Microservice System

Authors: Asaad Algarni

PAGE 275 – 282

Paper 28: Improved Whale Optimization Algorithm with LSTM for Stock Index Prediction

Authors: Yu Sun, Sofianita Mutalib, Liwei Tian

PAGE 283 – 295

Paper 29: Multinode LoRa-MQTT of Design Architecture and Analyze Performance for Dual Protocol Network IoT

Authors: Rizky Rahmatullah, Hongmin Gao, Ryan Prasetya Utama, Pupu Dani Prasetyo Adi, Jannat Mubashir, Rachmat Muwardi, Widar Dwi Gustian, Hanifah Dwiyaniti, Yuliza

PAGE 296 – 303

Paper 30: Machine Learning-Based Fifth-Generation Network Traffic Prediction Using Federated Learning

Authors: Mohamed Abdelkarim Nimir Harir, Edwin Ataro, Clement Temaneh Nyah

PAGE 304 – 313

Paper 31: CN-GAIN: Classification and Normalization-Denormalization-Based Generative Adversarial Imputation Network for Missing SMES Data Imputation

Authors: Antonius Wahyu Sudrajat, Ermatita, Samsuryadi

PAGE 314 – 322

Paper 32: An Agile Approach for Collaborative Inquiry-Based Learning in Ubiquitous Environment

Authors: Bushra Fazal Khan, Sohaib Ahmed

PAGE 323 – 333

Paper 33: M-COVIDLex: The Construction of a Domain-Specific Mixed Code Sentiment Lexicon

Authors: Siti Noor Allia Noor Ariffin, Sabrina Tiun, Nazlia Omar

PAGE 334 – 347

Paper 34: Optimized Hybrid Deep Learning for Enhanced Spam Review Detection in E-Commerce Platforms

Authors: Abdulrahman Alghaligah, Ahmed Alotaibi, Qaisar Abbas, Sarah Alhumoud

PAGE 348 – 357

Paper 35: Optimization of Fourth Party Logistics Routing Considering Infection Risk and Delay Risk

Authors: Guihua Bo, Sijia Li, Mingqiang Yin, Mingkun Chen, Xin Liu

PAGE 358 – 369

Paper 36: Convolutional Neural Network and Bidirectional Long Short-Term Memory for Personalized Treatment Analysis Using Electronic Health Records

Authors: Prasanthi Yavanamandha, D. S. Rao

PAGE 370 – 379

Paper 37: Spam Detection Using Dense-Layers Deep Learning Model and Latent Semantic Indexing

Authors: Yasser D. Al-Otaibi, Shakeel Ahmad, Sheikh Muhammad Saqib

PAGE 380 – 387

Paper 38: A Deep Learning for Arabic SMS Phishing Based on URLs Detection

Authors: Sadeem Alsufyani, Samah Alajmani

PAGE 388 – 396

Paper 39: Jordanian Currency Recognition Using Deep Learning

Authors: Salah Alghyaline

PAGE 397 – 404

Paper 40: Foreground Feature-Guided Camouflage Image Generation

Authors: Yuelin Chen, Yuefan An, Yonsen Huang, Xiaodong Cai

PAGE 405 – 411

Paper 41: Adoption of Generative AI-Enhanced Profit Sharing Digital Systems in MSMEs: A Comprehensive Model Analysis

Authors: Mardiana Andarwati, Galandaru Swalaganata, Sari Yuniarti, Fandi Y. Pamuji, Edward R. Sitompul, Kuku Yudhistiro, Puput Dani Prasetyo Adi

PAGE 412 – 422

Paper 42: DeepLabV3+ Based Mask R-CNN for Crack Detection and Segmentation in Concrete Structures

Authors: Yuewei Liu

PAGE 423 – 431

Paper 43: Multi-Objective Optimization of Construction Project Management Based on NSGA-II Algorithm Improvement

Authors: Yong Yang, Jinrui Men

PAGE 432 – 444

Paper 44: Performance Evaluation of Efficient and Accurate Text Detection and Recognition in Natural Scenes Images Using EAST and OCR Fusion

Authors: Vishnu Kant Soni, Vivek Shukla, S. R. Tandan, Amit Pimpalkar, Neetesh Kumar Nema, Muskan Naik

PAGE 445 – 453

Paper 45: AI-Powered Learning Pathways: Personalized Learning and Dynamic Assessments

Authors: Mohammad Abrar, Walid Aboraya, Rawad Abdel Khaliq, Kabali P Subramanian, Yousuf Al Husaini, Mohammed Al Husaini

PAGE 454 – 462

Paper 46: Feature Reduction and Anomaly Detection in IoT Using Machine Learning Algorithms

Authors: Adel Hamdan, Muhannad Tahboush, Mohammad Adawy, Tariq Alwada'n, Sameh Ghwanmeh

PAGE 463 – 470

Paper 47: Network Security Based on GCN and Multi-Layer Perception

Authors: Wei Yu, Huitong Liu, Yu Song, Jiaming Wang

PAGE 471 – 480

Paper 48: An Ensemble Semantic Text Representation with Ontology and Query Expansion for Enhanced Indonesian Quranic Information Retrieval

Authors: Liza Trisnawati, Noor Azah Binti Samsudin, Shamsul Kamal Bin Ahmad Khalid, Ezak Fadzrin Bin Ahmad Shaubari, Sukri, Zul Indra

PAGE 481 – 489

Paper 49: A Review of Reinforcement Learning Evolution: Taxonomy, Challenges and Emerging Solutions

Authors: Ji Loun Tan, Bakr Ahmed Taha, Norazreen Abd Aziz, Mohd Hadri Hafiz Mokhtar, Muhammad Mukhlisin, Norhana Arsad

PAGE 490 – 502

Paper 50: Towards Transparent Traffic Solutions: Reinforcement Learning and Explainable AI for Traffic Congestion

Authors: Shan Khan, Taher M. Ghazal, Tahir Alyas, M. Waqas, Muhammad Ahsan Raza, Oualid Ali, Muhammad Adnan Khan, Sagheer Abbas

PAGE 503 – 511

Paper 51: Strategic Supplier Selection in Advanced Automotive Production: Harnessing AHP and CRNN for Optimal Decision-Making

Authors: Karim Haricha, Azeddine Khat, Yassine Issaoui, Ayoub Bahnasse, Hassan Ouajji

PAGE 512 – 524

Paper 52: Understanding Art Deeply: Sentiment Analysis of Facial Expressions of Graphic Arts Using Deep Learning

Authors: Fei Wang

PAGE 525 – 534

Paper 53: A Hybrid Transformer-ARIMA Model for Forecasting Global Supply Chain Disruptions Using Multimodal Data

Authors: Qingzi Wang

PAGE 535 – 543

Paper 54: Marine Predator Algorithm and Related Variants: A Systematic Review

Authors: Emmanuel Philibus, Azlan Mohd Zain, Didik Dwi Prasetya, Mahadi Bahari, Norfadzlan bin Yusup, Rozita Abdul Jalil, Mazlina Abdul Majid, Azurah A Samah

PAGE 544 – 568

Paper 55: The Current Challenges Review of Deep Learning-Based Nuclei Segmentation of Diffuse Large B-Cell Lymphoma

Authors: Gei Ki Tang, Chee Chin Lim, Faezahtul Arbaeyah Hussain, Qi Wei Oung, Aidy Irman Yazid, Sumayyah Mohammad Azmi, Haniza Yazid, Yen Fook Chong

PAGE 569 – 583

Paper 56: User Interface Design of Digital Test Based on Backward Chaining as a Measuring Tool for Students' Critical Thinking

Authors: I Putu Wisna Ariawan, P. Wayan Arta Suyasa, Agus Adiarta, I Komang Gede Sukawijana, Nyoman Santiyadnya, Dewa Gede Hendra Divayana

PAGE 584 – 590

Paper 57: Early Alzheimer's Disease Detection Through Targeting the Feature Extraction Using CNNs

Authors: D Prasad, K Jayanthi, Pradeep Tilakan

PAGE 591 – 602

Paper 58: Enhancement of Coastline Video Monitoring System Using Structuring Element Morphological Operations
Authors: I Gusti Ngurah Agung Pawana, I Made Oka Widyantara, Made Sudarma, Dewa Made Wiharta, Made Widyia Jayantari

PAGE 603 – 611

Paper 59: Application of MLP-Mixer-Based Image Style Transfer Technology in Graphic Design

Authors: Qibin Wang, Xiao Chen, Huan Su

PAGE 612 – 621

Paper 60: Integrating Blockchain and Edge Computing: A Systematic Analysis of Security, Efficiency, and Scalability

Authors: Youness Bentayeb, Kenza Chaoui, Hassan Badir

PAGE 622 – 632

Paper 61: Enhancing COVID-19 Detection in X-Ray Images Through Deep Learning Models with Different Image Preprocessing Techniques

Authors: Ahmad Nuruddin bin Azhar, Nor Samsiah Sani, Liu Luan Xiang Wei

PAGE 633 – 644

Paper 62: Deep Learning-Based Automatic Cultural Translation Method for English Tourism

Authors: Jianguo Liu, Ruohan Liu

PAGE 645 – 653

Paper 63: A Novel Metric-Based Counterfactual Data Augmentation with Self-Imitation Reinforcement Learning (SIL)

Authors: K. C. Sreedhar, T. Kavya, J. V. S. Rajendra Prasad, V. Varshini

PAGE 654 – 661

Paper 64: Segmentation of Nano-Particles from SEM Images Using Transfer Learning and Modified U-Net

Authors: Sowmya Sanan V, Rimal Isaac R S

PAGE 662 – 677

Paper 65: Application of Big Data Mining System Integrating Spectral Clustering Algorithm and Apache Spark Framework

Authors: Yuansheng Guo

PAGE 678 – 686

Paper 66: Large Language Models for Academic Internal Auditing

Authors: Houda CHAMMAA, Rachid ED-DAOUDI, Khadija BENAZZI

PAGE 687 – 694

Paper 67: Enhanced Facial Expression Recognition Based on ResNet50 with a Convolutional Block Attention Module

Authors: Liu Luan Xiang Wei, Nor Samsiah Sani

PAGE 695 – 711

Paper 68: YOLO-WP: A Lightweight and Efficient Algorithm for Small-Target Detection in Weld Seams of Small-Diameter Stainless Steel Pipes

Authors: Huaishu Hou, Yukun Sun, Chaofei Jiao

PAGE 712 – 722

Paper 69: Determination of Pre Coding Elements and Activities for a Pre Coding Program Model for Kindergarten Children Using the Fuzzy Delphi Method (FDM)

Authors: Siti Naimah Rahman, Norly Jamil, Intan Farahana Abdul Rani, Hafizul Fahri Hanafi

PAGE 723 – 732

Paper 70: A Novel Internet of Things and Cloud Computing-Driven Deep Learning Framework for Disease Prediction and Monitoring

Authors: Bo GUO, Lei NIU

PAGE 733 – 740

Paper 71: Comparison of Artificial Neural Network and Long Short-Term Memory for Modelling Crude Palm Oil Production in Indonesia

Authors: Brodjol Sutijo Suprih Ulama, Robi Ardana Putra, Fausania Hibatullah, Mochammad Reza Habibi, Mochammad Abdillah Nafis

PAGE 741 – 747

Paper 72: Enhanced Jaya Algorithm for Quality-of-Service-Aware Service Composition in the Internet of Things

Authors: Yan SHI

PAGE 748 – 755

Paper 73: Enhancing Facial Expressiveness in 3D Cartoon Animation Faces: Leveraging Advanced AI Models for Generative and Predictive Design

Authors: Langdi Liao, Lei Kang, Tingli Yue, Aiting Zhou, Ming Yang

PAGE 756 – 767

Paper 74: A Lightweight Anonymous Identity Authentication Scheme for the Internet of Things

Authors: Zhengdong Deng, Xuannian Lei, Junyu Liang, Hang Xu, Zhiyuan Zhu, Na Lin, Zhongwei Li, Jingqi Du

PAGE 768 – 775

Paper 75: Comparative Analysis of Feature Selection Based on Metaheuristic Methods for Human Heart Sounds Classification Using PCG Signal

Authors: Motaz Farooq A Ben Hamza, Nilam Nur Amir Sjarif

PAGE 776 – 791

Paper 76: Developing an Integrated Platform to Track Real Time Football Statistics for Somali Football Federation (SFF)

Authors: Bashir Abdinur Ahmed, Husein Abdirahman Hashi, Abdifatah Abdilatif Ahmed, Abdikani Mahad Ali

PAGE 792 – 797

Paper 77: Elevator Abnormal State Detection Based on Vibration Analysis and IF Algorithm

Authors: Zhaoxiu Wang

PAGE 798 – 808

Paper 78: LFM Book Recommendation Based on Fusion of Time Information and K-Means

Authors: Dawei Ji

PAGE 809 – 818

Paper 79: A Proposed Approach for Agile IoT Smart Cities Transformation– Intelligent, Fast and Flexible

Authors: Othman Asiry, Ayman E. Khedr, Amira M. Idrees

PAGE 819 – 829

Paper 80: A Novel Optimization Strategy for CNN Models in Palembang Songket Motif Recognition

Authors: Yohannes, Muhammad Ezar Al Rivian, Siska Devella, Tinaliah

PAGE 830 – 841

Paper 81: A Novel Hybrid Algorithm Based on Butterfly and Flower Pollination Algorithms for Scheduling Independent Tasks on Cloud Computing

Authors: Huiying SHAO

PAGE 842 – 850

Paper 82: Task Scheduling in Fog Computing-Powered Internet of Things Networks: A Review on Recent Techniques, Classification, and Upcoming Trends

Authors: Dongge TIAN

PAGE 851 – 861

Paper 83: A System Dynamics Model of Frozen Fish Supply Chain

Authors: Leni Herdiani, Maun Jamaludin, Iman Sudirman, Widjajani, Ismet Rohimat

PAGE 862 – 873

Paper 84: Methodological Review of Social Engineering Policy Model for Digital Marketing

Authors: Wenni Syafitri, Zarina Shukur, Umi Asma' Mokhtar, Rossilawati Sulaiman

PAGE 874 – 885

Paper 85: Comprehensive Bibliometric Literature Review of Chatbot Research: Trends, Frameworks, and Emerging Applications

Authors: Nazruddin Safaat Harahap, Aslina Saad, Nor Hasbiah Ubaidullah

PAGE 886 – 896

Paper 86: Application of Collaborative Filtering Optimization Algorithm Based on Semantic Relationships in Interior Design

Authors: Kai Zhao, Lei Wang

PAGE 897 – 905

Paper 87: Hybrid Clustering Framework for Scalable and Robust Query Analysis: Integrating Mini-Batch K-Means with DBSCAN

Authors: Sridevi K N, Rajanna M

PAGE 906 – 912

Paper 88: Modified Moth-Flame Optimization Algorithm for Service Composition in Cloud Computing Environments

Authors: Yeling YANG, Miao SONG

PAGE 913 – 922

Paper 89: Enhanced Task Scheduling Algorithm Using Harris Hawks Optimization Algorithm for Cloud Computing

Authors: Fang WANG

PAGE 923 – 933

Paper 90: PCE-BP: Polynomial Chaos Expansion-Based Bagging Prediction Model for the Data Modeling of Combine Harvesters

Authors: Liangyi Zhong, Mengnan Deng, Maolin Shi, Ting Lou, Shaoyang Zhu, Jingwen Zhan, Zishang Li, Yi Ding

PAGE 934 – 943

Paper 91: Detecting Emotions with Deep Learning Models: Strategies to Optimize the Work Environment and Organizational Productivity

Authors: Cantuarias Valdivia Luis Alberto de Jesús, Gómez Human Javier Junior, Sierra-Liñan Fernando

PAGE 944 – 953

Paper 92: Sentiment and Emotion Analysis with Large Language Models for Political Security Prediction Framework

Authors: Liyana Safra Zaabar, Adriana Arul Yacob, Mohd Rizal Mohd Isa, Muslihah Wook, Nor Asiakin Abdullah, Suzaimah Ramli, Noor Afiza Mat Razali

PAGE 954 – 960

Paper 93: Text-to-Image Generation Method Based on Object Enhancement and Attention Maps

Authors: Yongsen Huang, Xiaodong Cai, Yuefan An

PAGE 961 – 968

Paper 94: Enhanced Traffic Congestion Prediction Using Attention-Based Multi-Layer GRU Model with Feature Embedding

Authors: Sreelekha M, Midhunchakkaravarthy Janarthanan

PAGE 969 – 983

Paper 95: Robust Joint Detection of Coronary Artery Plaque and Stenosis in Angiography Using Enhanced DCNN-GAN

Authors: M. Jayasree, L. Koteswara Rao

PAGE 984 – 995

Paper 96: Design and Research of Accounting Automation Management System Based on Swarm Intelligence Algorithm and Deep Learning

Authors: Dan Gui, Wei Ma, Wanfei Chen

PAGE 996 – 1005

Paper 97: Enhancing Road Safety: A Multi-Modal Drowsiness Detection System for Drivers

Authors: Guirrou Hamza, Mohamed Zeriab Es-Sadek, Youssef Taher

PAGE 1006 – 1011

Paper 98: Decoding Face Attributes: A Modified AlexNet Model with Emphasis on Correlation-Heterogeneity Relationship Between Facial Attributes

Authors: Abdelaali Benaiss, Otman Maarouf, Rachid El Ayachi, Mohamed Biniz, Mustapha Oujaoura

PAGE 1012 – 1026

Paper 99: Evaluation of Eye Movement Features and Visual Fatigue in Virtual Reality Games

Authors: Yuwei Ji

PAGE 1027 – 1038

Paper 100: High-Accuracy Vehicle Detection in Different Traffic Densities Using Improved Gaussian Mixture Model with Cuckoo Search Optimization

Authors: Nor Afiqah Mohd Aris, Siti Suhana Jamaian

PAGE 1039 – 1052

Paper 101: PSR: An Improvement of Lightweight Cryptography Algorithm for Data Security in Cloud Computing

Authors: P. Sri Ram Chandra, Syamala Rao, Naresh K, Ravisankar Malladi

PAGE 1053 – 1058

Paper 102: Optimizing Feature Selection in Intrusion Detection Systems Using a Genetic Algorithm with Stochastic Universal Sampling

Authors: RadhaRani Akula, GS Naveen Kumar

PAGE 1059 – 1068

Paper 103: Optimizing Route Planning for Autonomous Electric Vehicles Using the D-Star Lite Algorithm

Authors: Bhakti Yudho Suprpto, Suci Dwijayanti, Desi Windisari, Gatot Aria Pratama

PAGE 1069 – 1078

Paper 104: Stacking Regressor Model for PM2.5 Concentration Prediction Based on Spatiotemporal Data

Authors: Mitra Unik, Imas Sukaesih Sitanggang, Lailan Syaufina, I Nengah Surati Jaya

PAGE 1079 – 1086

Paper 105: Feature Substitution Using Latent Dirichlet Allocation for Text Classification

Authors: Norsyela Muhammad Noor Mathivanan, Roziyah Mohd Janor, Shukor Abd Razak, Nor Azura Md. Ghani

PAGE 1087 – 1098

Paper 106: Multilabel Classification of Bilingual Patents Using OneVsRestClassifier: A Semiautomated Approach

Authors: Slamet Widodo, Ermatita, Deris Stiawan

PAGE 1099 – 1106

Paper 107: Dolphin Inspired Optimization for Feature Extraction in Augmented Reality Tracking

Authors: Indhumathi S, Christopher Clement J

PAGE 1107 – 1116

Paper 108: Empirical Analysis of Variations of Matrix Factorization in Recommender Systems

Authors: Srilatha Tokala, Murali Krishna Enduri, T. Jaya Lakshmi, Koduru Hajarathaiah, Hemlata Sharma

PAGE 1117 – 1138

Paper 109: Efficient Tumor Detection in Medical Imaging Using Advanced Object Detection Model: A Deep Learning Approach

Authors: Taoufik Saidani

PAGE 1139 – 1145

Paper 110: Efficient Anomaly Detection Technique for Future IoT Applications

Authors: Ahmad Naseem Alvi, Muhammad Awais Javed, Bakhtiar Ali, Mohammed Alkathami

PAGE 1146 – 1158

Paper 111: GRACE: Graph-Based Attention for Coherent Explanation in Fake News Detection on Social Media

Authors: Orken Mamyrbayev, Zhanibek Turysbek, Mariam Afzal, Marassulov Ussen Abdurakhimovich, Ybytayeva Galiya, Muhammad Abdullah, Riaz Ul Amin

PAGE 1159 – 1171

Paper 112: Intelligent Fault Diagnosis for Elevators Using Temporal Adaptive Fault Network

Authors: Zhiyu Chen

PAGE 1172 – 1182

Paper 113: High-Precision Multi-Class Object Detection Using Fine-Tuned YOLOv11 Architecture: A Case Study on Airborne Vehicles

Authors: Nasser S. Albalawi

PAGE 1183 – 1190

Paper 114: AI-Driven Image Recognition System for Automated Offside and Foul Detection in Football Matches Using Computer Vision

Authors: Qianwei Zhang, Lirong Yu, WenKe Yan

PAGE 1191 – 1198

Paper 115: Deep Q-Learning-Based Optimization of Path Planning and Control in Robotic Arms for High-Precision Computational Efficiency

Authors: Yuan Li, Byung-Won Min, Haozhi Liu

PAGE 1199 – 1207

Paper 116: Android Malware Detection Through CNN Ensemble Learning on Grayscale Images

Authors: El Youssofi Chaymae, Choug dali Khalid

PAGE 1208 – 1217

Paper 117: Cross-Domain Health Misinformation Detection on Indonesian Social Media

Authors: Divi Galih Prasetyo Putri, Savitri Citra Budi, Arida Ferti Syafiandini, Ikhlasul Amal, Revandra Aryo Dwi Krisnandaru

PAGE 1218 – 1224

Paper 118: Comparison of Machine Learning Algorithms for Malware Detection Using EDGE-IIoTSET Dataset in IoT

Authors: Jawaher Alshehri, Almaha Alhamed, Mounir Frikha, M M Hafizur Rahman

PAGE 1225 – 1238

Paper 119: Building Detection from Satellite Imagery Using Morphological Operations and Contour Analysis over Google Maps Roadmap Outlines

Authors: Arbab Sufyan Wadood, Ahthasham Sajid, Muhammad Mansoor Alam, Mazliham MohD Su'ud, Arshad Mehmood, Inam Ullah Khan

PAGE 1239 – 1255

Paper 120: Exploring Machine Learning in Malware Analysis: Current Trends and Future Perspectives

Authors: Noura Alyemni, Mounir Frikha

PAGE 1256 – 1268

Paper 121: SM9 Key Encapsulation Mechanism for Power Monitoring Systems

Authors: Chao Hong, Peng Xiao, Pandeng Li, Zhenhong Zhang, Yiwei Yang, Biao Bai

PAGE 1269 – 1277

Paper 122: A Review of Analyzing Different Agricultural Crop Yields Using Artificial Intelligence

Authors: Vijaya Bathini, K. Usha Rani

PAGE 1278 – 1290

Paper 123: LMS-YOLO11n: A Lightweight Multi-Scale Weed Detection Model

Authors: YaJun Zhang, Yu Xu, Jie Hou, YanHai Song

PAGE 1291 – 1300

Paper 124: DBYOLOv8: Dual-Branch YOLOv8 Network for Small Object Detection on Drone Image

Authors: Yawei Tan, Bingxin Xu, Jiangsheng Sun, Cheng Xu, Weiguo Pan, Songyin Dai, Hongzhe Liu

PAGE 1301 – 1309

Paper 125: Eagle Framework: An Automatic Parallelism Tuning Architecture for Semantic Reasoners

Authors: Haifa Ali Al-Hebshi, Muhammad Ahtisham Aslam, Kawther Saeedi

PAGE 1310 – 1322

Paper 126: Imbalance Datasets in Malware Detection: A Review of Current Solutions and Future Directions

Authors: Hussain Almajed, Abdulrahman Alsaqer, Mounir Frikha

PAGE 1323 – 1335

Paper 127: Artificial Intelligence in Financial Risk Early Warning Systems: A Bibliometric and Thematic Analysis of Emerging Trends and Insights

Authors: Muhammad Ali Chohan, Teng Li, Suresh Ramakrishnan, Muhammad Sheraz

PAGE 1336 – 1351

Paper 128: DBFN-J: A Lightweight and Efficient Model for Hate Speech Detection on Social Media Platforms

Authors: Nourah Fahad Janbi, Abdulwahab Ali Almazroi, Nasir Ayub

PAGE 1352 – 1361

Paper 129: Exploring the Best Machine Learning Models for Breast Cancer Prediction in Wisconsin

Authors: Abdullah Al Mamun, Touhid Bhuiyan, Md Maruf Hassan, Shahedul Islam Anik

PAGE 1362 – 1368

Paper 130: A Machine Learning-Based Analysis of Tourism Recommendation Systems: Holistic Parameter Discovery and Insights

Authors: Raniah Alsaifi, Rashid Mehmood, Saad Alqahtany

PAGE 1369 – 1382

Advanced Machine Learning Approaches for Accurate Migraine Prediction and Classification

Chokri Baccouch*¹, Chaima Bahar²

LIGM, University Gustave Eiffel, Marne-la-Vallée, France^{1,2}

LR-Sys'Com-ENIT, Communications Systems LR-99-ES21, National Engineering School of Tunis, University of Tunis, Tunisia¹
MACS Laboratory LR16ES22, National Engineering School of Gabes, University of Gabes, Tunisia²

Abstract—Migraine is a neurovascular disorder with a prevalence that exceeds 1 billion individuals worldwide, but it has long been recognized to have unique diagnostic challenges due to its heterogeneous pathophysiology and dependence on subjective assessments. As has been extensively documented by a number of international law bodies, migraine in the workplace has been identified as a significant issue that requires urgent attention. Migraine defined by episodic, unilateral and debilitating symptoms including aura, nausea incurs a high socioeconomic burden in disability. Mechanisms such as altered cortical excitability and trigeminal system activation, although researched to a high extent, are still inadequately understood. Deep learning and machine learning (ML) hold tremendous potential for transforming diagnosis and classification of migraine. This study evaluates several machine learning (ML) models such as gradient boosting, decision tree, random forest, k-Nearest Neighbors (KNN), support vector machine (SVM), logistic regression, multi-layer perceptron (MLP), artificial neural network (ANN), and deep neural network (DNN) for multi-class classification of migraine. By employing advanced preprocessing techniques and publicly obtainable datasets, the study addresses the challenge of identifying different types of migraines that may share common variables. In this study, several machine learning (ML) models including gradient boosting, decision tree, random forest, k-Nearest Neighbors show that for multi-class migraine classification MLP and Gradient Boosting had good performance in most models, but did perform poorly in complex subcategories like Typical Aura with Migraine. Both attained high accuracies (96.4% and 97%, respectively). KNN and Logistic Regression, two traditional models, performed well at basic classifications but poorly at more complex situations; Neural networks (ANN and DNN) showed much flexibility towards data complexities. These results underscore how important it is to align model selection with data properties and provide avenues for improving performance through regularization and feature engineering. This strategy illustrates how AI-powered solutions can revolutionize the way we manage, treat, and prevent migraines across the globe.

Keywords—Headache classification; migraine; migraine diagnosis; migraine classification

I. INTRODUCTION

Migraine is a complex and common neurovascular disease that poses many challenges to accurate diagnosis and effective treatment. More than 90% of people in the world are affected by headache disorders in general [1], but migraine stands out for its effects on the brain, body, and quality of life in particular. They are among the most common causes of neurological consultations, and treatment costs in countries such as China approach an annual 672.7 billion yuan [2]. Although migraines are not directly life-threatening, they significantly impair work

performance, physical health, mental well-being, and overall quality of life [3].

The multifaceted nature of migraine, a chronic medical condition with overlapping legal and social dimensions, is well-documented. The impact of this condition on individuals' rights, professional and personal lives is significant and thus requires a comprehensive response that combines advanced healthcare, legal protection, and technological innovation. Internationally, international laws play a crucial role in regulating the treatment of chronic and complex diseases such as migraine, which have a profound impact on patients' quality of life and functioning. In this context, the International Covenant on Economic, Social and Cultural Rights (ICESCR) is a significant instrument, as it recognizes in [4] the right of individuals to health, a comprehensive right that is not limited to the provision of treatment but extends to the right to access basic healthcare services, including migraine treatment. This right is essential not only to improve the state of health of patients but also to restore the ability to lead a normal professional and social life [5, 6].

The Convention on the Rights of Persons with Disabilities (CRPD) further underscores the imperative to ensure that individuals with disabilities, including those afflicted by chronic migraine, have access to essential health services [7]. The CRPD obliges state parties to formulate comprehensive health policies that guarantee the provision of specialised medical care, taking into account individual differences in diagnosis and treatment. The utilisation of innovative technology, including machine learning techniques, has the potential to enhance the accuracy of diagnosis and personalise treatment regimens, ensuring that all patients receive the timely and optimal care they require. Legislative frameworks, such as the Americans with Disabilities Act (ADA) in the United States, play a pivotal role in safeguarding individuals with migraine from discrimination in the workplace. This legislation stipulates the provision of reasonable accommodations, such as flexible work schedules, quiet work environments, and the ability to work from home, thereby ensuring that individuals with migraine can continue to perform their professional duties in a manner that is both conducive to their well-being and effective in their roles [5,8].

Migraine, the top cause of functional disability among people aged 15 to 55, can severely hamper productivity and routine activities during episodes. The often-unpredictable nature of migraine attacks increases the anxiety and dysfunction related to them by the uncertainty of when they might occur. Conven-

tional treatment strategies either interrupt migraine during an attack or reduce their frequency through preventative measures. Preventive medication on high-risk days, as well as abortive treatment that is most effective early in the migraine cycle, has moved from proof-of-concept studies with more promising early data. This highlights the need for prediction-based solutions in migraine management. In the International Classification of Headache Disorders (ICHD). In [9], headaches are classified into three categories: primary headaches (e.g., migraine, tension-type headaches, and trigeminal autonomic cephalalgias), secondary headaches, and cranial neuropathies or facial pain disorders.

Causes of migraines are myriad, including diet, lifestyle, genetics, and physiology (Fig. 1). The role of dietary triggers, such as caffeine, alcohol, and food additives, together with lifestyle factors, such as stress, poor sleeping patterns, and lack of exercise, cannot be underestimated in precipitating attacks. Additional risk factors include having a genetic predisposition to the condition and heightened susceptibility driven by physiological as well as biochemical factors, from hormonal fluctuations to neurotransmitter imbalances, which leads to the onset of migraines. Migraines have a multi-faceted pathophysiology, with two key components being triggers (stimuli causing attacks) and prodromal symptoms, cognitive, sensory, behavioral, or physical changes, that can occur 1 to 48 hours before an attack, and serve as critical but difficult-to-measure indicators of imminent migraines due to their highly subjective nature and methodological biases. Indeed, neurophysiological changes (e.g. changes in autonomic tone) can inform on this prodromal phase. Migraine and tension-type headaches are the most common types of primary headaches worldwide, with 10% and 40% respectively, while cluster headaches are rare, with an estimated prevalence of 0.1% [10, 11].

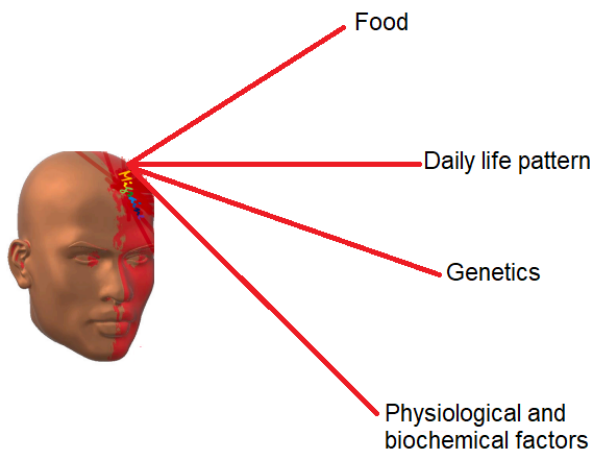


Fig. 1. Factors that trigger migraine.

Intractable migraines, with a 16% annual incidence in the general population, are the second most prevalent cerebral disease in the world and rank as the leading cause of disability worldwide, even more than all neurological diseases combined [12]. Migraines are generally divided into three categories: migraines with aura, migraines without aura, and chronic migraines. Migraines with aura occur in up to 25% of cases and are characterised by transient visual, speech, or neurological

abnormalities lasting no longer than an hour [13]. Migraines without aura, on the other hand, appear as unilateral, moderate-to severe-intensity, pulsatile pain, often with accompanying nausea and vomiting, photophobia, and phonophobia, and can last from 4 to 72 h if untreated [14]. Migraine is classified into episodic (less than 15 headache days per month) and chronic types (defined as 15 or more headache days per month with eight or more headache days with features of fully developed migraine), the latter being more frequent and having an unfavorable impact on daily life [15].

Digital technologies on the rise allow new opportunities in migraine management. Direct translation can occur through wearable technology and mobile health technologies for headache characteristics, prodromic symptoms, and physiological changes. However, because so much complexity is involved in the neurobiological processes underlying migraine, prediction can be difficult. Accurate prediction requires sophisticated models capable of integrating and interpreting complex flows of biological and physiological data. Machine learning (ML) appears to be a promising solution, as it can process and analyze complex and heterogeneous data types. Machine learning could streamline migraine detection, prediction, and classification processes, enhance diagnostic accuracy, optimize treatments, and ultimately reduce the financial and societal burden associated with migraine management.

This study aims to explore the revolutionary potential of machine learning (ML) to evolve migraine attack prediction, particularly in resource-limited settings with limited access to state-of-the-art medical technology. Even when they can be valuable, conventional diagnostics such as MRI (Magnetic Resonance Imaging), PET(Positron Emission Tomography) and CT(Computed Tomography) scans are pricey and require specialist knowledge, which places them at a disadvantage in developing countries. When using ML algorithms that provide a cost-efficient and high-throughput alternative, reliable systems for diagnosing and predicting the onset of migraine are now widely available.

Many advanced machine learning techniques were evaluated in this study, such as artificial neural networks (ANN), deep neural networks (DNN), multi-layer perceptron (MLP), logistic regression, k-nearest neighbors (KNN), support vector machines (SVM), gradient boosting, decision trees, and random forests. Models showed promising results with a model accuracy of 97.12% (MLP), 96.40% (Gradient Boosting), and 96.04% (Decision Tree). Such results showcase this exciting potential for AI-powered approaches to revolutionize headache management and improve patient outcomes globally.

The remainder of the paper is organized as follows: the “Related Work” section re-vIEWS prior research, while the “Materials and Methods” section outlines the proposed methodology and dataset. The “Experiments” section details the conducted experiments and their findings. Finally, the “Conclusion and Future Work” section summarizes the key results and outlines potential directions for future research.

II. RELATED WORK

Recent advances in artificial intelligence (AI) can yield complex predictive algorithms able to predict migraine episodes. Smith et al. For instance, [16] used supervised

machine learning methods such as random forests and deep learning networks on longitudinal data collected by biometric tracking devices or mobile applications. A great example was given by the team at the National Institutes of Health, who showed that neural network models were able to predict future migraine attacks 85% of the time by including trigger factors (stress, sleep, food, etc.) into the model. This underscores the need for machine learning to pre-emptive migraine treatment, and real-time data collection.

Several studies have explored whether the combination of multimodal data (e.g. genetic, environmental, and behaviour) can augment prediction accuracy. The model of Zhang et al. [4] was able to achieve better sensitivity and specificity than using conventional methods by integrating information from genetic profiles, lifestyle surveys, and wearable sensors. The results show how combining data from multiple sources can help us learn more about and predict migraine episodes.

There has also been ongoing research regarding advancement in early detection of migraines using brain imaging techniques as well as biomarkers. For example, during the premonitory phase of migraines, Garcia and his team found cues that predicted a migraine episode was coming [17] and uncovered discrepancies in patients' levels of neurotransmitters; that is, glutamate and serotonin. Using functional magnetic resonance imaging (fMRI), their study suggested that there are different patterns of brain activity before and during migraine attacks, with encouraging potential prospects for real-time detection. Adding to this, Garcia et al. [17] conducted further research on serum biomarkers in migraine patients, including increased levels of glutamate and serotonin. This opens up the potential of diagnosis by biomolecular profile and hints at neurological mechanisms. Their techniques help ensure early intervention strategies, which become more accurate by providing a non-biased way to identify migraines.

Continuous monitoring devices like smartwatches and fitness trackers have captured interest due to their ability in terms of potentially mitigating migraines. According to Lee et al. when physiological data (e.g. heart rate and stress) were collected in real time, they gave a 78% chance of detecting the initial symptoms for migraine [18]. These results highlight the benefits of regular physiologic monitoring in migraine therapy and may become a paving method for wearable technology-assisted preventive treatment fighting techniques.

Conventional approaches to migraine classification are based on the International Classification of Headache Disorders' (IHS) symptoms criteria. Recent studies, however, aim to improve this strategy by adding more precise clinical features. Müller et al. [19], for example, discovered a subtype of migraine associated with increased sensory sensitivities and sleep disturbances, opening the door to more individualized treatment choices.

The classification of migraines has been further transformed by genomic advancements. A meta-analysis of genetic research by Johnson et al. [20] found many genetic loci linked to heightened migraine risk. Their study suggested a genetic risk classification by combining genetic data with clinical information. This in turn permitted both patient stratification and treatment individualization to the genetic profile of each host.

The study in [21] used five different supervised machine learning methods which aimed to define group of symptoms described by participants as migraines. For classification and deployment, we used Weka data mining tool. The results indicated that, of all the models tested, naïve bayes would be more suitable and easier.

An investigation [22] used brain signals captured through an EEG and a computer-aided diagnostic (CAD) system to classify various forms of migraines. This system accomplished classification using deep learning models as follows: VGG16, ResNet101, and DenseNet121.

A method to integrate EEG in an online migraine detection tool for support of clinical decision making was also presented by a respective study [23]. The EEG dataset consisted of recordings from 21 healthy volunteers and 18 migraine patients. The results showed that the Bi-LSTM method with 128 channels outperformed other models, including Random Forests (RF), Linear Discriminant Analysis (LDA), and Support Vector Machines (SVM), with the maximum accuracy of 95.99%.

Furthermore, 400 patients' clinical data that had been annotated by domain experts was employed in a different study [24]. The 24 most pertinent factors were chosen after the researchers first collected data based on symptoms. Then, to categorize migraines, an Artificial Neural Network (ANN) and other conventional machine learning models were used. The ANN model outperformed other algorithms including SVM, Logistic Regression (LR), Decision Trees, and k-Nearest Neighbors (KNN) with a 97% classification accuracy for migraines.

In [25], different machine learning techniques were used to examine somatosensory evoked potential components in the frequency and temporal domains for migraine categorization. Among these were Logistic Regression (LR), Linear Discriminant Analysis (LDA), Random Forests (RF), k-Nearest Neighbors (KNN), Extreme Gradient Boosting (XGBoost), Support Vector Machines (SVM), and Multilayer Perceptrons (MLP). The models were able to differentiate between interictal or ictal migraine conditions and healthy controls with an accuracy of over 88%.

In detecting the two classes of headaches and differentiating between healthy controls and migraine sufferers, the CNN method based on an initiation module outperformed the conventional support vector machine, which had an accuracy of 83.67%, with a greater accuracy of 86.18% [26]. A feature selection technique was used in a different study [27] to enhance the migraine group's classification. With accuracies rising from 67% to 93%, 90% to 95%, and 93% to 94%, respectively, this method improved the performance of the Naive Bayes, SVM, and Adaboost classifiers. In a similar vein, a study by [28] that used EEG signals to diagnose migraines early revealed that the artificial neural network (ANN) outperformed logistic regression and support vector machines (SVM) with an accuracy of 88%. Hemoglobin changes in the prefrontal cortex (PFC) were observed using functional near-infrared spectroscopy (fNIRS) during a mental arithmetic (MAT) task. The specificities and sensitivities were 75% and 100% for chronic migraine (CM) and 100% and 75% for medication overuse headaches (MOH), respectively. Based on the findings,

it seems that fNIRS and machine learning work better together to classify migraines [7].

In the medical industry, data mining techniques are essential. Data exploration classification methods such as Naïve Bayes, KNN, SVM, and random forests were used in the study [29]. Among these, Naïve Bayes emerged as the best classifier, with an accuracy of 0.905 and a precision of 0.475. A medical case study on hemodynamic parameter monitoring of actual patients is presented as a practical scenario to monitor real patients' life parameters using the WBSN (Wireless Body Sensor Network). N4SID models (Numerical Subspace State-Space System Identification) were built with a low false positive rate and an average forecasting horizon of 47 minutes [30]. Finally, we used one of machine learning techniques to distinguish healthy subjects with migraineurs by combining three functional measures from rs-fMRI [31].

Ufuk et al. [32] proposed the use of deep neural networks (DNN) for diagnosing migraines, achieving an accuracy of 95%. They used eight attributes to diagnose three types of migraines (with aura, without aura, and chronic migraine). Ferroni [33] suggested using a decision support system (DSS) to diagnose medication-overuse migraine, with an accuracy of 82%. In another study [34], a DSS was proposed for diagnosing primary headaches, achieving an accuracy of 80%. The authors compared four machine learning techniques: Bagging, Naïve Bayes, Boosting, and Random Forest. Rober Keight [36] proposed using decision trees (DST) to diagnose primary headache types using 9 machine learning classifiers, achieving an accuracy of 95%. Hao Yang [35] used convolutional neural networks (CNN) for migraine classification from MRI, achieving an accuracy of 99%. Akben [36] implemented an artificial neural network (ANN) for migraine diagnosis, with an accuracy of 83.3%. Akben [37] also used an SVM classifier to diagnose migraines, achieving an accuracy of 85%. Subasi [38] tested different versions of the Random Forest method for migraine diagnosis, obtaining an accuracy of 85.95%. De la Hoz [39] used an ANN for migraine diagnosis, achieving an accuracy of 88%. Yolanda Garcia [27] proposed feature selection for migraine diagnosis, achieving an accuracy of 90%. Even more recently, researchers have focused on the use of MRI and fMRI images for the detection and classification of migraines [40–42].

In [43], the study presented the design and development of an ML decision support system aimed at providing diagnosis of tension headaches and migraines. The results obtained with the logistic regression model were found to be the best among all. The accuracy level raised to 0.84 with a stand against models such as gradient boosting algorithms and random forests.

The Table I describes the parameters that were used during related works.

III. MATERIALS AND METHODS

The provision of preparation of the data takes a long time and uses relatively powerful computational resources, with the straightforward methods of deep learning/machine learning. Therefore, getting some relevant information depends on an effective machine learning system. Designing further this machine-learning architecture is therefore quite compli-

TABLE I. PARAMETERS USED DURING RELATED WORKS

Study	Techniques Used	Dataset/Attributes	Accuracy
[23]	Bi-LSTM, SVM, LDA, Random Forest	EEG signals from 18 migraine patients and 21 controls	95.99%
[24]	ANN, SVM, Logistic Regression, Decision Trees, KNN	Clinical data from 400 patients	97%
[25]	SVM, RF, KNN, XGBoost, LDA, MLP, Logistic Regression	Somatosensory evoked potential features	88%
[26]	CNN with initiation module	EEG signals	86.18%
[27]	Naïve Bayes, SVM, Adaboost	Feature selection applied to migraine group	Increased from 67%-94%
[28]	ANN	EEG signals, fNIRS	88%
[29]	Naïve Bayes, KNN, SVM, Random Forest	Data exploration techniques for classification	90.5% (Naïve Bayes)
[32]	DNN	8 attributes for diagnosing 3 types of migraines	95%
[33]	DSS (Decision Support System)	Medication-overuse migraine data	82%
[35]	DSS	Primary headache data	80%
[24]	DST (Decision Trees)	9 machine learning classifiers for primary headache types	95%
[36]	CNN	MRI data for migraine classification	99%
[37]	ANN	-	83.3%
[38]	SVM	-	85%
[39]	Random Forest	Different versions of Random Forest	85.95%
[40]	ANN	-	88%
[43]	Logistic Regression, Gradient Boosting, Random Forest	Symptom-based data for headache classification	84%

cated. Customizing or tuning a model involves adjusting the parameters of the classifier.

In this study, different models were trained using a variety of machine-learning algorithms. These were adjusted and optimized afterward for the dataset in order to enhance the quality of classification. As shown in Fig. 2, the algorithms that have been considered include Gradient Boosting, Decision Tree, Random Forest, k-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression, Multi-Layer Perceptron (MLP), Artificial Neural Networks (ANN), and Deep Neural Networks (DNN).

A. Dataset

An in-depth examination of contributing causes and related symptoms is made possible by the migraine database that is supplied, which provides a thorough and exhaustive overview of the many components of this ailment. Individuals' ages, which range from 18 to 70 years old, are a crucial component of the demographic data since they enable investigation of the effects of migraines on various age groups. This dataset is notable for its comprehensive examination of the features of migraine episodes, recording variables like attack duration (which can vary from 30 minutes to 72 hours) and frequency (which can range from 1 to 10 attacks per month), providing a more accurate picture of symptom severity and recurrence. Additionally, a scale from 1 to 10 is used to assess the pain's intensity, with 10 being the most severe agony.

Additional information is given on where the pain may occur, which assists with spotting recurring patterns, such as

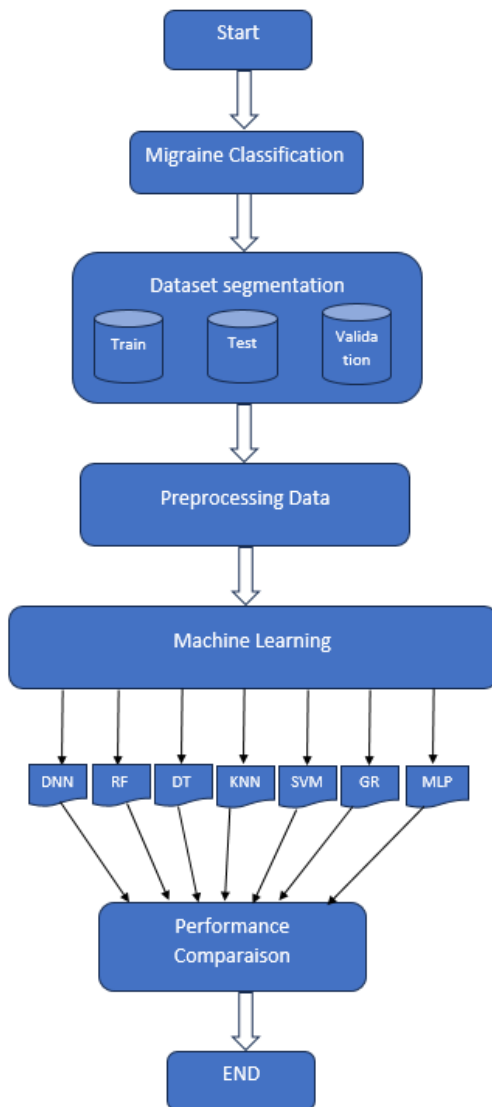


Fig. 2. Proposed system flowchart for migraine classification.

a preponderance of unilateral pain (either right or left) or pain felt on the forehead, neck, or temples. The database has also captured those associated conditions, nausea and vomiting, typical signs of migraine. The other statistic showing that 70% of migraineurs feel sick and about 50% vomit during a headache creates a more nuanced clinical impression of the side effects.

In addition to the actual pain, this database includes phenomenological and other sensory characteristics of migraines: phonophobia (hypersensitivity to sound), photophobia (hypersensitivity to light), and other visual anomalies (blurriness, aura, etc.). The extremely important features of this disease dealing with the sensory aspect are the symptoms, and about 60% of migraineurs experience photophobia and phonophobia during their attacks.

This dataset, with its rich pool of variables, affords taking a deep look into migraine research, leading to empirical observations to determine the associations among variables.

Having this data will allow for extensive studies related to how age, frequency of attacks, intensity, and concomitant symptoms influence the severity of migraines. The scientific and medical community would greatly benefit from this resource, as it is of prime importance to improve diagnostics, design personalized treatments, and take more focused approaches to treatment in clinical practice.

1) *Database preprocessing:* Data preparation is an important step before applying machine learning models in order to obtain really reliable results. This comprises a number of crucial sub-steps in the context of our analysis of migraine data, including noise reduction, inconsistent data repair, error detection, and data conversion into useful numerical variables.

We began our efforts by doing much cleaning of the data: extreme values and missing information were removed, mismatches resolved, and missing numbers imputed. For example, any rows where there was missing data on features such as age, severity of pain, or frequency of attack were either deleted or imputed. To further ensure data validity and consistency, problematic cases of data entry (like somewhat unbelievable values of negative ages and migraine attacks lasting more than 72 hours) were fixed.

The input was then transformed into numerical variables so that machine learning algorithms could process it more easily. To make the data interpretable for the analytic models, some variables, such as pain intensity, were left on a numeric scale (from 1 to 10), while other variables, like related symptoms (nausea, vomiting, phonophobia), were converted into binary variables.

These methods have strengthened data quality and made our dataset a better candidate for machine learning models while simultaneously guaranteeing a balanced representation of the various classes (such as severe and non-severe migraines). This pre-emphasis technique is what allows a model to achieve the highest performance possible during training and yield analyses that are more reliable and precise of variables linked to migraines.

The study relied on an initial corpus of 1,386 clinical records of Tunisian patients suffering from various pathologies associated with migraines. Several machine learning classifiers, including KNN, SVM, RF, DT, LG, MLP, ANN, and DNN, were applied. The proposed analysis used the diagnosed condition and migraine symptoms as input data.

This analysis focused on 23 variables, including age, visual disturbances, dizziness, and vomiting that represent the common clinical symptomology during an acute headache. In addition, the variable identified as diagnostic was included to signify the type of migraine classically referred to. This variable is the diagnosis of the migraine type that was made by the doctor on the basis of the patient's medical history and reported symptoms. The symptomatic variables record the manifestations such as nausea or lightheadedness.

The Feature Importance Analysis is shown in Fig. 3, with enrolling age, visual disturbances, intensity of pain, and phonophobia among the class-leading features in migraine classification. On the contrary, there are worthless factors such as ataxia, diplopia, and dysarthria that have little influence on classification and could be disregarded; this would increase

efficiency and thereby help simplify the model. This procedure emphasizes the priority of emphasizing the necessary components while eliminating the least important ones to boost the performance of classifications.

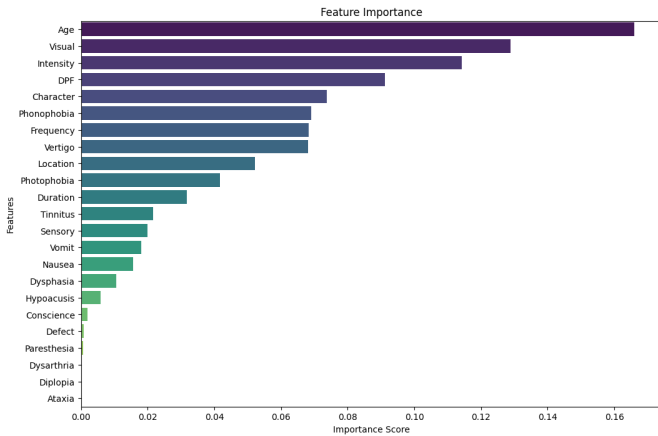


Fig. 3. Feature importance analysis for migraine classification.

B. Classification Models

The algorithms testing on the migraine classification dataset after applying the basic preprocessing methods, various machine learning methods including GB, LG, SVM, KNN, DT, RF, MLP, ANN as well as deep neural network DNN were applied to the dataset.

The parameters utilized in the experiment are described in Table II.

TABLE II. HYPER-PARAMETERS FOR DIFFERENT MODELS

Model	Hyper-parameter	Value
multirow DNN	Number of epochs	100
	Activation function	relu
	Optimizer	Adam
	Model	Sequential (first layer)
	Number of neurons at first dense layer	512
	Hidden layer	2
	Classification function	softmax
	Loss function	categorical-cross entropy
SVM	Kernel	Linear
	Class	sklearn.svm.SVC
	Regularization parameters C	1
	Probability	True
KNN	Neighbors range	(1,15,1)
	Weights	Uniform
	Metric distance	Euclidean
RF	n_estimators	100
	max_depth	50
	min_samples_split	5
	min_samples_leaf	2
	max_features	sqrt
	random_state	42

IV. EXPERIMENTS AND RESULTS

The outcomes for migraine classification using various machine learning models are presented in this section. Specifically, we focus the assessment on the classifiers of complexity namely Multi-Layer Perceptron (MLP), Deep Neural Networks

(DNN), and Artificial Neural Networks (ANN). These classifiers were examined in detail with a view that performance assessment based on a variety of training sets would reveal their capability in maintaining accuracy, robustness, applicability in migraine diagnosis and attacks management with a large dataset. In-person examination and assessment reveal to a great degree the applicability of such techniques in real life including the merits and demerits.

A. ANN Model

Despite the fact that both belong to the same family of neural networks, ANN and DNN stand separated by layers-of-Depth and complexity, which will influence their efficiency in predicting, detecting, and classifying migraines. ANN, having just one or two hidden layers, is more preferred in classification, such as classifying migraineurs from clinical tabular data, because ANN works better with small datasets due to its lesser training data requirements and resistance to overfitting. Fig. 4 illustrates the basic architecture of ANN.

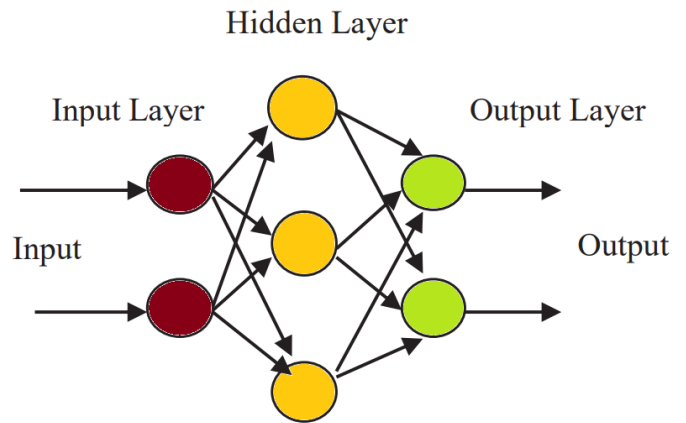


Fig. 4. Basic model of ANN.

The accuracy curves (Fig. 5) show a steady improvement in performance over the 100 epochs, reaching a high level and stabilizing around 95% for both the training and validation sets. The model does not exhibit significant signs of overfitting, as the validation accuracy closely follows the training accuracy. This indicates that the dense neural network (ANN) has effectively learned to generalize without being restricted solely to the training data.

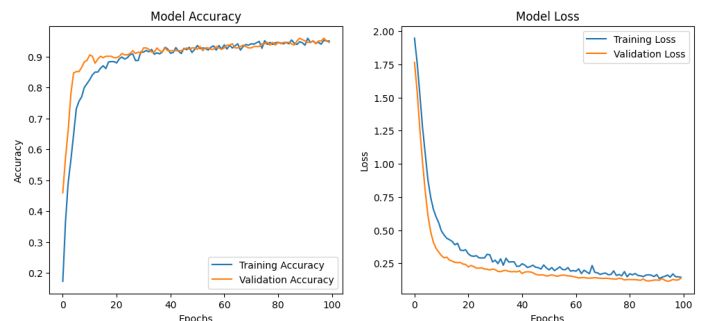


Fig. 5. Accuracy and loss graph of ANN model.

The loss curve (Fig. 5) demonstrates a steady decline, converging toward a low value, indicating the effective training of the ANN model. The validation loss closely mirrors the training loss, confirming good generalization without noticeable overfitting or underfitting.

The confusion matrix (Fig. 6) provides a detailed assessment of the model's classification performance across different categories. Correct predictions are highlighted along the main diagonal, with most classes, including Basilar-type aura, Familial hemiplegic migraine, Migraine without aura, Other, and Typical aura without migraine, exhibiting near-perfect accuracy. However, some misclassifications are noted, particularly for Sporadic hemiplegic migraine, which shows moderate confusion with the Other category, and for Typical aura with migraine, where a few samples are incorrectly classified as Basilar-type aura or Familial hemiplegic migraine.

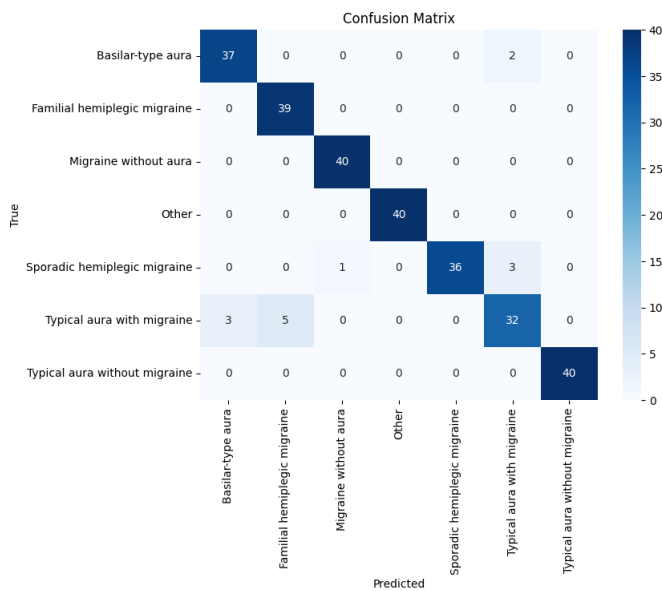


Fig. 6. Confusion matrix of ANN model.

The ANN model provides relatively good performance, and that can be seen by clear separability among most of the classes and fewer misclassifications there. Additional support for effective learning and generalization is provided by accuracy and loss curves. However, remaining challenges relate more to classes with overlapping characteristics, such as migraines with or without aura and the different types of hemiplegic migraine. Class imbalances or an insufficient distinction among classes in the dataset could be responsible for the said problems.

B. DNN Model

Deep neural networks (DNN) become particularly suited for such works because of their more comprehensive and deeper architecture and their capability to furnish complex information, for instance, time-series information collected by Internet of Things sensors for other physiological signals (ECG, vectorcardiogram), or functional MRI images. The ability of these models to perform well in complex tasks—such as discerning migraine types (e.g. aura versus non-aura) and amalgamating data from various sources to give timely

warning of migraine attacks—is astonishing. An ability to capture complex interactions proves helpful to tackling migraine classification’s different challenging problems. However, the efficiency of DNN usually depends on the availability of large datasets and the employment of complex regularization techniques to minimize the risks of overfitting. Thus, this emphasizes how critical it is to have proper data preparation and model optimization to fully leverage the advantage of DNN in this field.

The architecture of the DNN model used in the classification of migraine types is described in its Fig. 7. It has four layers—an input layer, two hidden layers, and an output layer. This architecture can manage the complexity of the migraine-related datasets and give extremely high classification results.

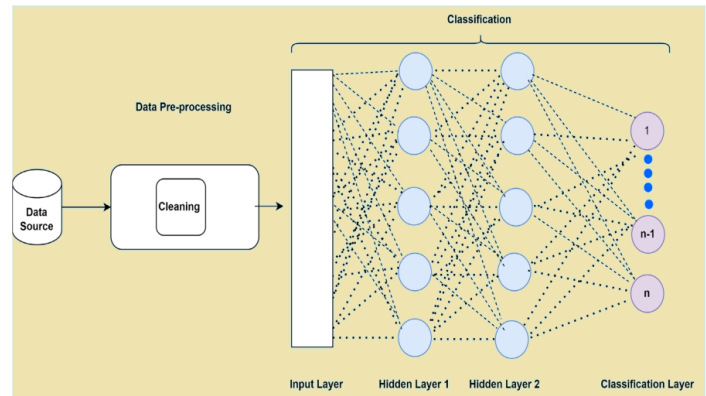


Fig. 7. Fundamental architectural design of a deep neural network applied to migraine classification.

The DNN model showed a fast improvement of performances within the first epochs, after which the accuracy curve was stable at above 95% up to the end of 100 iterations, as seen in Fig. 8. This behavior shows the ability of the model to form these important data features and learn them. The training accuracy matched close to the validation accuracy, which shows good generalization capability. The two curves diverge only mildly, indicating that the model avoids overfitting and can keep providing high performance on unseen data.

The loss curve, also shown in Fig. 8, drops swiftly in the first few epochs, indicating a decent learning process that reduces errors. The optimization is seen to be successful when the curve plateaus at a low value. The same pattern occurs in the validation loss, which indicates that the model was well-regularized and neither overfit nor underfit. Lastly, an additional line indicates the test loss, which tells the generalization of this model to independent data, lying very close to the training and validation losses.

This robustness of the model has also been corroborated by the recall and F1-score metrics, assessing its capacity to classify each class correctly and strike a balance between precision and recall. All average values for these metrics exceed 0.95.

The confusion matrix in Fig. 9 provides a detailed treatment of misclassifications. The great clustering of values along the diagonal illustrates that most samples have been correctly assigned. There are, however, slight misclassifications, including

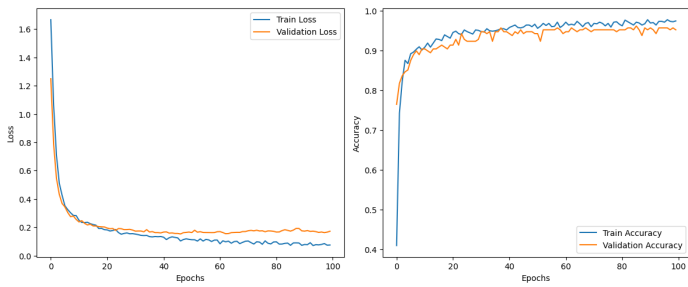


Fig. 8. Accuracy and loss graph of DNN model.

the assignment of four sporadic hemiplegic migraine samples to other categories and the wrong assignment of two Basilar-type Aura samples as Typical Aura along with migraine. These unintentional errors are common in any multi-class classification problem; being rare, they barely dent the overall efficacy of the model.

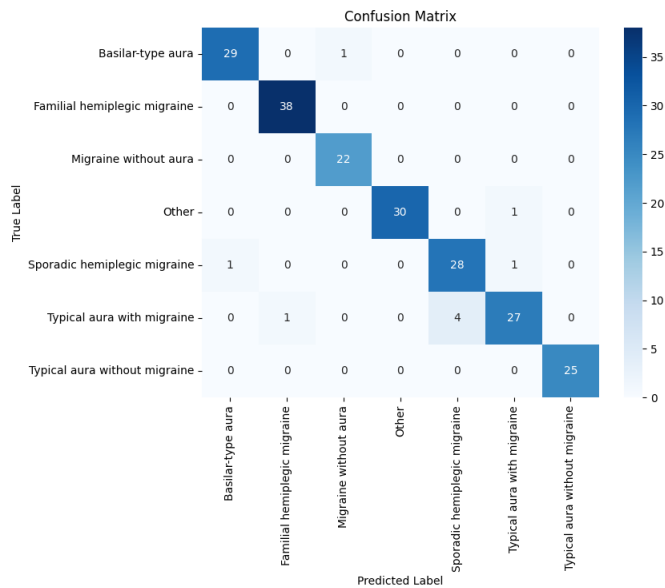


Fig. 9. Confusion matrix of DNN model.

The findings reveal the successful classification of migraines by the DNN model and its high generalization capability for fresh data. The resilience of the model is evidenced by the almost-perfect results in numerous categories and overall high accuracy. These encouraging results suggest that DNN can support the diagnoses and categorizations of different types of migraine while maintaining a strong balance between learning and generalization.

C. Multi-Layer Perception Model

The architecture of the Multi-Layer Perceptron (MLP) model, presented in Fig. 10, is essential for classification tasks, particularly in predicting migraine types. Taking factors such as age, migraine severity, or symptom frequency as inputs, the architecture comprises three main modules: the input layer, hidden layer(s), and the output layer. The hidden neurons nest in various buried layers, using activation functions and

weight computations to identify complex patterns in the data. The output layer gives predictions like Migraine with Aura or Migraine without Aura. Every neuron uniformly influences the neurons in the downstream layers, thus enabling the MLP to model complex and intricate non-linear relations while analyzing clinical migraine data.

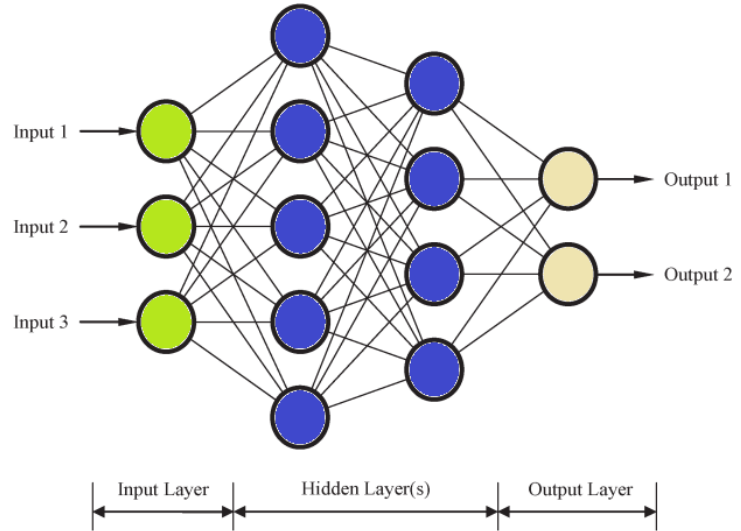


Fig. 10. MLP Model architecture.

How well the Multi-Layer Perceptron (MLP) model have performed in migraine prediction, detection, and classification is validated by the learning curves (Fig. 11). The loss curve which is an insight into learning, shows a very high decrease from 1.75 to around 0.2 during the initial 20 epochs, thereby it rapidly develops. It then stabilizes between 0.1 and 0.2, which indicates the model has been successfully trained and has effectively converged. Notably, when the training and validation loss curves are in close proximity it implies a lack of overfitting and high generalization to new data by the model which is another way of saying its accuracy is high hence.

The accuracy curves (Fig. 11) further underscore the model's success, showing a swift increase in performance, reaching 95 to 97% accuracy within the first 20 epochs and maintaining stability afterward. Although slight fluctuations in the validation curve occur likely due to mini-batch variations or sample differences they do not compromise the overall robustness of the results. The consistency observed between the training and validation curves for both accuracy and loss demonstrate that the model is well-regularized and capable of generalizing effectively.

These results support the MLP model's dependability in accurately diagnosing migraines while striking the ideal balance between generalization and learning.

The results gathered by the model are quite promising and reassuring in finding which kind of migraines types is dissimilar, even in the case of a multi-class scenario, when things are much more complicated. As the confusion matrix (Fig. 12) is quite detailed, we can witness the model's performance within different migraine categories.

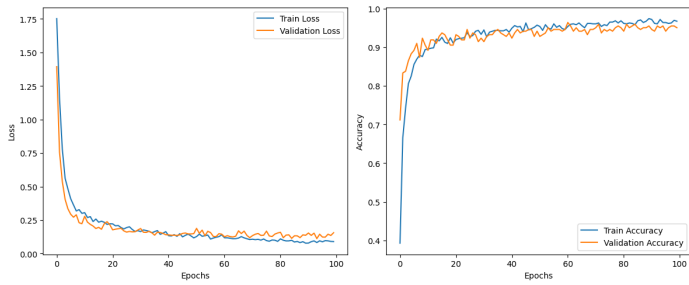


Fig. 11. Accuracy and loss graph of MLP model.

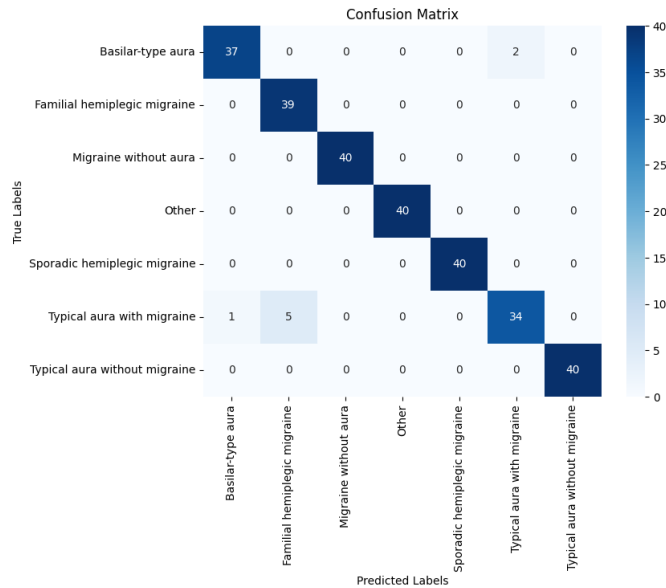


Fig. 12. Confusion matrix of MLP model.

The performance of classes demonstrates very high precision in a small number of categories. Notably, Migraine without aura, Other, Sporadic hemiplegic migraine, and Typical aura without migraine are the best ones with precision, recall, and F1-score equal to 1.00. These facts are evidence for the model's ability to practically learn to correctly classify different types of migraines according to specific characteristics among other features by the time the model is finished.

For Basilar-type aura, the model achieved a precision of 97% and a recall of 95%, resulting in an F1-score of 96%. Although these results are outstanding, they indicate that a small number of samples were misclassified. Similarly, for familial hemiplegic migraine, while the model achieved a perfect recall of 100% indicating all instances of this class were identified, its precision was 89%, suggesting some degree of misclassification with other classes.

Performance was somewhat worse in the event of a typical aura with migraine, with an F1-score of 89% (precision of 94% and recall of 85%). Overlapping characteristics with different migraine kinds are probably to blame for this, which could make classification difficult.

97% accuracy was achieved in the experiment of experimenting with errors with a particular reason by the model and

the training data, and the precision is 0.9915, recall is 0.5926, and F1-score is 0.7321. In multi-class tasks that complicate the problem well, such as difficulty in distinguishing the boundary between the two classes or the presence of overlapping behaviors, the ability of this model to maintain balance is more than necessary. Together, these results establish that the model is not only able to learn from a variety of types of data but also to classify properly into many classes in general.

V. DISCUSSION

The research compared the supervisory learning models: Gradient Boosting, Decision Tree, Random Forest, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Logistic Regression, Multi-Layer Perceptron (MLP), Artificial Neural Networks (ANN), and Deep Neural Networks (DNN) for migraine prediction and classification. Interesting information about the strengths and weaknesses of the models' related to the multi-class classification of migraine was exposed by the comparison.

The Gradient Boosting model exhibited impressive performance in classes such as Migraine Without Aura and Typical Aura Without Migraine, yielding an accuracy rate of 96.4%. But it struggled with complex types like Migraine with Typical Aura. The Decision Tree model exhibited stellar performance at 96.04% accuracy; yet, it faced challenges with Basilar-Type Aura, shown by an F1-score of 0.90.

In a number of classes, including Migraine Without Aura, Other, and Sporadic Hemiplegic Migraine, the Multi-Layer Perceptron (MLP) demonstrated the maximum accuracy of 97%, with flawless precision and recall. However, MLP performed worse for Typical Aura with Migraine (F1-score of 89%), most likely as a result of category overlap.

The Random Forest model had a very good accuracy at 95%, but exhibited signs of overfitting, particularly in certain categories. It performed well in less complex classes, such as Other and Typical Aura Without Migraine. With corresponding accuracies of 93.17% and 92.09%, KNN performed better than SVM. KNN proved effective in distinguishing Basilar-Type Aura and Migraine Without Aura but faced similar challenges in complex categories. Logistic Regression performed at a baseline with an accuracy of 89%, excelling in simpler classifications but struggling with nuanced categories, such as Typical Aura with Migraine, where its F1-score dropped to 70%.

The ANN model performance was robust in terms of accuracy with 95% on overlapping class problems as long as the scope of the clinical dataset was limited and well defined. It was a bit challenging for the ANN to manage overlapping classes. On the other hand, DNN performed slightly better at 95.19%, leveraging its deeper architecture to model more complex patterns, especially for challenging classes like Sporadic Hemiplegic Migraine. However, the DNN had high computational cost and also extensive regularization requirements which are its main downsides.

In conclusion, the best models overall were the Gradient Boosting and MLP as they gave consistently high accuracy. Both ANN and DNN also had their advantages, ANN was optimal for less complicated datasets which use One Dimensional

representations and DNN was optimal for multidimensional and complex datasets. All these findings stress the need to be selective on the choice of model to be developed, with respect to the complexity of the data and the nature of the task. For the future research, better performance may be obtained from harnessing class imbalance, better feature engineering, fine-tuning the regularization and adding ensemble methods.

In Table III, the results of existing work for migraine classification are equated with the accuracy of the classification produced by our proposed model.

TABLE III. COMPARATIVE RESULTS

Model	Accuracy (%)	F1-score (%)	Sensitivity (%)	precision (%)
Gradient Boosting	96.4	96	96	97
Random Forest	95	71.8	73.2	70.8
SVM	92.09	91.64	91.63	91.64
KNN	93.17	92.63	92.74	93.08
Decision Tree	96.04	95.87	95.84	96.03
Logistic Regression	89.21	89.2	89	89.8
MLP	97.12	94.14	94.25	94.32
ANN	95.86	95.4	95.7	95.7
DNN	95.19	95.15	95.08	95.18

VI. CONCLUSION

This study highlights the ability of machine learning to correctly define attacks of migraines through classification models. Gradient Boosting achieved an accuracy of 96.4%, excelling in classes like Migraine Without Aura, while MLP stood out as the best performer with 97% accuracy and perfect scores in several classes. Artificial Neural Network also performed well with ANN at 95% accuracy and DNN at 95.19%, although computational demands were notable. Other models such as Logistic Regression (89%) struggled with nuanced categories, while Random Forest (95%), KNN (93.17%), and SVM (92.09%) performed moderately. Finally, MLP and Gradient boosting were the outstanding models emphasizing the importance of model selection which depends on the complexity of the data set in improving clinical practice.

The implications of improving our understanding of how algorithm choice affects performance in classification and providing a way forward in performing more efficient migraine diagnosis are crucial for future research through feature engineering and model optimization. Future studies may incorporate ensemble methods, refine how complex models overfit and improve procedures for more detailed and specific types of migraines.

ACKNOWLEDGMENT

The authors extend their appreciation to LR-.Sys'Com-ENIT, Communications Systems LR-99-ES21 National Engineering School of Tunis, University of Tunis and MACS Laboratory: Modeling, Analysis and Control of Systems LR16ES22 National Engineering School Gabes, University of Gabes.

REFERENCES

[1] Hagen, K. et al. Te epidemiology of headache disorders: A face-to-face interview of participants in hunt4. *J. Headache Pain* 2018 , 19, 1–6.
[2] Yao, C. et al. Burden of headache disorders in china, 1990–2017: Findings from the global burden of disease study 2017. *J. Headache Pain* 2019, 20, 1–11 .

[3] Takeshima, T. et al. Prevalence, burden, and clinical management of migraine in china, japan, and south Korea: A comprehensive review of the literature. *J. Headache Pain* 2019, 20, 1–15 (2019).
[4] Zhang, L., et al. (2023). Multimodal Data Integration for Migraine Prediction: A Machine Learning Approach. *Neuroinformatics* 2023, 21, 1, 89-105.
[5] N. Riggins and L. Paris, “Legal Aspects of Migraine in the Workplace,” *Current Pain and Headache Reports*, Nov. 2022, doi: <https://doi.org/10.1007/s11916-022-01095-x>
[6] K. Pärli, “Presenteeism, Its Effects and Costs: A Discussion in a Labour Law Perspective,” *International Journal of Comparative Labour Law and Industrial Relations*, vol. 34, no. Issue 1, pp. 53–75, Mar. 2018, doi: <https://doi.org/10.54648/ijcl2018003>.
[7] Chen, W.-T. et al. Migraine classification by machine learning with functional near-infrared spectroscopy during the mental arithmetic task. *Sci. Rep* 2022, 12, 14590.
[8] O. Begasse de Dhaem and F. Sakai, “Migraine in the workplace,” *eNeurologicalSci*, vol. 27, p. 100408, Jun. 2022, doi: <https://doi.org/10.1016/j.ensci.2022.100408>.
[9] Wu, Q. et al. Determining the efficacy and safety of acupuncture for the preventive treatment of menstrual migraine: A protocol for a prisma-compliant systematic review and meta-analysis. *J. Pain Res* 2023,16, 101–109.
[10] Pacheco-Barrios, K. et al. Primary headache disorders in Latin America and the Aaribbean: A meta-analysis of population-based studies. *Cephalalgia* 2023, 43, 03331024221128265.
[11] Islam, J. et al. Modulation of trigeminal neuropathic pain by optogenetic inhibition of posterior hypothalamus in cci-ion rat. *Sci. Rep* 2023, 13, 489.
[12] Safri, S. et al. Te burden of Parkinson’s disease in the middle east and north Africa region, 1990–2019: Results from the global burden of disease study 2019. *BMC Public Health* 2023, 23, 107.
[13] Barral, E., Martins Silva, E., Garcia-Azorin, D., Viana, M. and Puledda, F. Diferential diagnosis of visual phenomena associated with migraine: Spotlight on aura and visual snow syndrome. *Diagnostics* 2023, 13, 252.
[14] Hansen, J. M. and Charles, A. Diferences in treatment response between migraine with aura and migraine without aura: Lessons from clinical practice and rets. *J. Headache Pain* 2023, 20, 1–10.
[15] Khanal, S. et al. A systematic review of economic evaluations of pharmacological treatments for adults with chronic migraine. *J. Headache Pain* 2022,23, 122.
[16] Smith, J., and Doe, A. Predicting Migraine Episodes Using Deep Neural Networks. *Journal of Neurological Disorders* 2022, 15, 3, 245-260.
[17] Garcia, M., et al. Serotonin and Glutamate Levels as Biomarkers for Migraine Detection. *Biomarkers in Medicine* 2021, 13, 4, 311-325.
[18] Müller, K., et al. Symptom-Based Subtyping of Migraine: Clinical Implications. *Headache Research* 2022, 18, 2, 134-150.
[19] Johnson, P., and Lee, S. Genetic Variants Associated with Migraine Susceptibility: A Meta-Analysis. *Genetic Medicine* 2023, 25, 5, 512-530.
[20] Lee, H., et al. (2023). Real-Time Detection of Migraine Predisposition Using Wearable Devices. *Journal of Wearable Technology* 2023, 9, 2, 77-95.
[21] Gulati, S., Guleria, K. and Goyal, N. Classification of migraine disease using supervised machine learning. In *'2022 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)* 2022, 1–7.
[22] Aslan, Z. Deep convolutional neural network-based framework in the automatic diagnosis of migraine. *Circuits Syst. Signal Process* 2022., 42, (5), 3054–3071.
[23] Göker, H. Automatic detection of migraine disease from EEG signals using bidirectional long-short term memory deep learning model. *Signal Image Video Process* 2022, 17 (4), 1255–1263.
[24] Sanchez-Sanchez, P. A., García-González, J. R. and Rúa Ascar, J. M. Automatic migraine classification using artificial neural networks. *F1000Research* 2020 9, 618.

- [25] Zhu, B., Coppola, G. and Shoaran, M. Migraine classification using somatosensory evoked potentials. *Cephalalgia* 2019, 39, 1143–1155.
- [26] Yang, H., Zhang, J., Liu, Q. and Wang, Y. Multimodal MRI-based classification of migraine: Using deep learning convolutional neural network. *Biomed. Eng. Online* 2018, 17, 1–14.
- [27] Garcia-Chimeno, Y., Garcia-Zapirain, B., Gomez-Beldarrain, M., Fernandez-Ruanova, B. and Garcia-Monco, J. C. Automatic migraine classification via feature selection committee and machine learning techniques over imaging and questionnaire data. *BMC Med. Inform. Decis. Mak* 2017, 17, 1–10.
- [28] Jindal, K. et al. Migraine disease diagnosis from eeg signals using non-linear feature extraction technique. In *'2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCI)* 2018, 1–4.
- [29] Sah, R. D., Sheetlani, J., Kumar, D. R. and Sahu, I. N. Migraine (headaches) disease data classification using data mining classifiers. *J. Res. Env. Earth Sci* 2017, 3, 10–16.
- [30] Pagán, J. et al. Robust and accurate modeling approaches for migraine per-patient prediction from ambulatory data. *Sensors* 2015, 15, 15419–15442.
- [31] Chong, C. D. et al. Migraine classification using magnetic resonance imaging resting-state functional connectivity data. *Cephalalgia* 2017, 37, 828–844.
- [32] Celik, U., Yurtay, N. and Pamuk, Z. Migraine diagnosis by using artificial neural networks and decision tree techniques. *AJIT-e Acad.J. Inform. Technol* 2014, 5, 79–90.
- [33] Ferroni, P. et al. Machine learning approach to predict medication overuse in migraine patients. *Comput. Struct. Biotechnol. J* 2020, 18, 1487–1496.
- [34] Krawczyk, B., Simić, D., Simić, S. and Woźniak, M. Automatic diagnosis of primary headaches by machine learning methods. *Open Med* 2013, 8, 157–165.
- [35] Chen, I. Y. et al. Ethical machine learning in healthcare. *Ann. Rev. Biomed. Data Sci* 2021, 4, 123–144.
- [36] Akben, S. B., Tuncel, D. and Alkan, A. Classification of multi-channel eeg signals for migraine detection. *Biomed. Res* 2016, 27, 743–748.
- [37] Akben, S. B., Subasi, A. and Tuncel, D. Analysis of repetitive flash stimulation frequencies and record periods to detect migraine using artificial neural network. *J. Med. Syst* 2012, 36, 925–931.
- [38] Subasi, A., Ahmed, A., Aličković, E. and Hassan, A. R. Effect of photic stimulation for migraine detection using random forest and discrete wavelet transform. *Biomed. Signal Process. Control* 2019, 49, 231–239.
- [39] Casas Pulido, A. F., Hernandez Cely, M. M. and Rodriguez, O. M. H. Análisis experimental de flujo líquido-líquido en un tubo horizontal usando redes neuronales artificiales. *Revista UIS Ingenierías* 2023, 22, 49–56.
- [40] Dumkrieger, G., Chong, C. D., Ross, K., Berisha, V. and Schwedt, T. J. The value of brain MRI functional connectivity data in a machine learning classifier for distinguishing migraine from persistent post-traumatic headache. *Front. Pain Res* 2023, 3, 1012831.
- [41] Nie, W., Zeng, W., Yang, J., Zhao, L. and Shi, Y. Classification of migraine using static functional connectivity strength and dynamic functional connectome patterns: A resting-state fmri study. *Brain Sci* 2023, 13, 596.
- [42] Marino, S. et al. Classifying migraine using pet compressive big data analytics of brain's μ -opioid and d2/d3 dopamine neurotransmission. *Front. Pharmacol* 2023, 14, 1173596.
- [43] Liu, F., Bao, G., Yan, M. and Lin, G. A decision support system for primary headache developed through machine learning. *PeerJ* 2022, 10, e12743.

A Comparative Study of Predictive Analysis Using Machine Learning Techniques: Performance Evaluation of Manual and AutoML Algorithms

Karim Mohammed Rezaul¹, Md. Jewel², Anjali Sudhan³, Mifta Uddin Khan⁴, Maharage Roshika Sathsarani Fernando⁵, Kazy Noor e Alam Siddiquee⁶, Tajnuva Jannat⁷, Muhammad Azizur Rahman⁸, Md Shabiul Islam⁹

Wrexham University, Faculty of Arts, Science and Technology, Wrexham LL11 2AW, UK¹

Centre for Applied Research in Software & IT (CARSIT), 80a Ashfield Street, London, England, E1 2BJ, UK^{2, 3, 4, 5, 7}

Multimedia University, Faculty of Engineering (FOE), Cyberjaya 63100, Malaysia^{6, 9}

Cardiff Metropolitan University, Department of Computer Science, Llandaff Campus, Western Avenue, Cardiff, CF5 2YB, UK⁸

Abstract—In this study, we have compared manual machine learning with automated machine learning (AutoML) to see which performs better in predictive analysis. Using data from past football matches, we tested a range of algorithms to forecast game outcomes. By exploring the data, we discovered patterns and team correlations, then cleaned and prepped the data to ensure the models had the best possible inputs. Our findings show that AutoML, especially when using logistic regression can outperform manual methods in prediction accuracy. The big advantage of AutoML is that it automates the tricky parts, like data cleaning, feature selection, and tuning model parameters, saving time and effort compared to manual approaches, which require more expertise to achieve similar results. This research highlights how AutoML can make predictive analysis easier and more accurate, providing useful insights for many fields. Future work could explore using different data types and applying these techniques to other areas to show how adaptable and powerful machine learning can be.

Keywords—Machine learning; predictive analytics; sports forecasting; automated machine learning (AutoML); feature engineering; model evaluation; data pre-processing; algorithm comparison; football analytics; sports betting; team performance metrics; exploratory data analysis (EDA); cross-validation techniques

I. INTRODUCTION

Millions of football fans from all around the world attend the UEFA European Championship, also referred to as the UEFA Euro. This esteemed competition, which is hosted by UEFA, features the top teams from throughout Europe, showcasing their talent, tenacity, and competitive spirit [1]. Analysts, enthusiasts, and commentators eagerly engage in predicting the outcomes of this highly anticipated and often unpredictable event. Recent advancements in machine learning (ML) algorithms, combined with the availability of extensive historical football data, have opened new avenues for predicting match results and identifying potential tournament winners. These sophisticated algorithms can detect patterns in complex datasets, providing valuable insights for predictive analysis and strategic decision-making.

Using ML algorithms to forecast the UEFA Euro winner involves a detailed analysis of historical match data, team

statistics, player performance metrics, and other factors that influence team success. By understanding the intricate interactions of these factors, ML models can predict future outcomes based on data from past tournaments. In this study, a variety of manual ML algorithms known for their efficiency in predictive modelling were used. These include Ada Boost, Random Forest, XGBoost, Decision Tree, Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbors (KNN), and Naive Bayes. Each algorithm has unique strengths, making them suitable for different predictive tasks in forecasting the winner.

Additionally, the study explores the realm of Automated Machine Learning (AutoML), utilizing advanced techniques to streamline and optimize the model development process. The AutoML framework incorporates a broad set of algorithms, such as Ridge, Quadratic Discriminant Analysis, Linear Discriminant Analysis, Extra Trees Classifier, Extreme Gradient Boosting, Light Gradient Boosting Machine, and Dummy Classifier, among others. This comprehensive approach allows for a thorough evaluation of predictive efficacy.

This study aims to demonstrate the efficacy of both manual and automated machine learning (AutoML) approaches in sports analytics, especially in forecasting the results of the UEFA Euro 2024. This study aims to demonstrate the potential for widespread acceptance and innovation in sports analytics by analysing the performance of several machine learning algorithms in projecting tournament results. The findings of this study are expected to enlighten stakeholders such as analysts, coaches, and investors, allowing them to make more informed judgements and strategic assumptions during the tournament.

The paper is organized as follows: Section I presents the introduction. In Section II, we define the problem and outline the objectives. Section III covers the literature review, addressing manual and automated machine learning separately. Section IV explains the research methodology. Section V details the preprocessing, cleaning, and data preparation steps. Section VI provides a comparative analysis of manual and automated machine learning across various classifiers. Section VII presents the results and evaluation, and finally, Section VIII concludes the paper.

II. PROBLEM DEFINITION AND OBJECTIVES

The purpose of this study is to show that both manual and automatic machine learning (AutoML) methodologies can accurately forecast UEFA Euro 2024 outcomes. The study creates prediction models for team performance by analysing historical data on international football matches and team characteristics. The findings, which highlight the potential of machine learning, particularly AutoML, in improving prediction accuracy, seek to enlighten analysts, coaches, and bettors, supporting greater use and innovation in sports analytics.

The main objective of this study is to demonstrate that both manual and automatic machine learning (AutoML) methods can effectively predict the outcomes of the UEFA Euro 2024 competition. The goal is to create prediction models that accurately assess each team's likelihood of success by thoroughly analysing historical data from international football matches and considering various team characteristics. This research aims to highlight the potential of machine learning (ML), particularly AutoML, in expediting predictive analytics processes and enhancing model accuracy.

The key objectives include the careful collection and pre-processing of historical international soccer match data and team attributes, followed by a detailed exploratory data analysis to identify relevant features. The study will then involve selecting informative features and applying rigorous feature engineering techniques to capture essential team and match characteristics. Various machine learning algorithms will be evaluated and compared to identify the most effective models based on performance metrics. The selected models will undergo thorough training using pre-processed data, with hyperparameter optimization to ensure optimal performance. Rigorous evaluation of predictive performance using appropriate metrics, along with the implementation of cross-validation techniques to ensure model generalizability and reduce overfitting, are crucial parts of this research.

Additionally, this project will test the performance of the models using benchmark datasets. The study aims to demonstrate the potential of these models in sports analytics by providing significant insights and encouraging wider adoption and innovation. The focus will be on evaluating the effectiveness of AutoML approaches.

III. LITERATURE REVIEW

A. Use of Manual Machine Learning

Forecasting tournament winners is an enduring challenge in sports analysis, extensively explored across research. Machine learning emerges as a prominent tool in this domain, offering predictive capabilities based on historical data. Leveraging past records, machine learning models discern patterns and variables correlated with auspicious outcomes, thereby enabling forecasts for future events. This methodology capitalizes on the wealth of information contained within historical datasets, facilitating the identification of key determinants of success. Through iterative learning processes, these algorithms refine their predictive accuracy, contributing to the advancement of sports analytics. The utilization of machine learning in predicting tournament winners emphasises the significance of

data-driven approaches in enhancing the understanding of sports dynamics and informing strategic decision-making processes within the realm of athletics. The challenge of predicting football match outcomes is addressed in the research by Hucaljuk & Rakipović, acknowledging the complexity stemming from numerous unquantifiable factors. A software solution is developed to tackle this challenge, undergoing testing to optimize feature and classifier combinations. Results demonstrate satisfactory predictive capabilities surpassing reference methods, with an accuracy exceeding the initial goal of 60%. However, the study suggests areas for enhancement, particularly in feature selection, proposing the inclusion of player form data for improved accuracy [2]. Additionally, increasing the size of the dataset for training could further enhance predictive performance. This project exemplifies successful advancement in football match prediction methodologies while highlighting avenues for future research and refinement in feature engineering.

Another research investigates the efficacy of utilizing machine learning techniques to predict football match outcomes by incorporating pre-game features instead of relying solely on post-game goal statistics. Custom-generated features are developed and compared against in-game data features using the XGBoost algorithm. Results indicate superior prediction accuracy with custom features, demonstrating higher precision, recall, f1 score, and accuracy compared to in-game features. The research suggests that leveraging comprehensive player and team statistics, such as dribbling and expected goals, could further enhance predictive performance. Additionally, considering factors like team formation and fan sentiment from social media could provide valuable insights into match outcomes. The study underscores the potential for enhanced predictive modelling in football matches through the incorporation of diverse pre-game features and data sources [3]. The studies by Hucaljuk & Rakipović and Rose et al. 2022 both address the challenge of predicting football match outcomes using machine learning techniques [2] [3]. Hucaljuk & Rakipović develop a software solution to optimize feature and classifier combinations, surpassing an initial accuracy goal of 60% [2]. They highlight the importance of feature selection and suggest incorporating player form data to enhance predictive accuracy. Conversely, focus on incorporating pre-game features, such as comprehensive player and team statistics, using custom-generated features [3]. This study demonstrates superior prediction accuracy compared to relying solely on post-game goal statistics, emphasizing the potential for enhanced predictive modelling through diverse pre-game features and data sources. Both studies contribute to advancing football match prediction methodologies and highlight avenues for future research in feature engineering and data analysis. The research by Groll et al. [4] introduces a hybrid modelling approach for predicting soccer match scores, combining random forests with two ranking methods: Poisson ranking and bookmakers' odds. By incorporating team covariate information and ability parameters derived from both ranking methods, the model accurately estimates team strengths. The approach is applied to FIFA Women's World Cups 2011, 2015, and 2019, with simulations favouring the USA as the top contender for the 2019 title, followed by France, England, and Germany. The study highlights the effectiveness of integrating

diverse methodologies for robust predictions in soccer tournaments, offering insights into team performance and tournament outcomes [4].

Another study focuses on employing machine learning techniques to predict the winner of the ICC Men's T20 World Cup 2020. Four algorithms, including Random Forest, Extra Trees, ID3, and C4.5, were compared, with Random Forest exhibiting the highest proficiency at 80.86% custom accuracy. Australia emerged as the predicted champion. Future directions include optimizing the predictive models and incorporating additional parameters like match venue and weather forecast to enhance accuracy. The study underscores the utility of machine learning in sports prediction and offers insights into potential improvements for future analyses, emphasizing the importance of considering various factors for more accurate forecasts in cricket tournaments [5]. The both studies focus on utilizing machine learning techniques for sports prediction, albeit in different contexts. Groll and the team introduce a hybrid modelling approach for predicting soccer match scores, incorporating random forests with Poisson ranking and bookmakers' odds. This study demonstrates the effectiveness of integrating diverse methodologies for robust predictions in soccer tournaments, providing insights into team performance and outcomes. In contrast, Basit and the team concentrate on predicting the winner of the ICC Men's T20 World Cup 2020 using machine learning algorithms such as Random Forest, Extra Trees, ID3, and C4.5 [4] [5]. This research underscores the utility of machine learning in sports prediction, emphasizing the importance of considering various factors for more accurate forecasts, particularly in cricket tournaments. Both studies contribute to advancing predictive modelling in sports and offer valuable insights into improving accuracy in tournament predictions. In examining cricket match prediction models, emphasize the development of a machine learning model specifically tailored for Indian Premier League (IPL) matches, achieving nearly 90% accuracy [6]. Conversely, Kumar, et al. [7] focus on Decision Trees and Multilayer Perceptron Network models, highlighting the superiority over traditional statistical methods. This CricAI system offers a user-friendly prediction tool, emphasising the flexibility of machine learning approaches. Vistro et al [8] similarly explore IPL match prediction using various machine learning algorithms, achieving high accuracies up to 94.87%. While all studies underscore the significance of data science in sports analytics, two studies emphasize the potential applicability of the methodologies beyond cricket, which could inform predictive analytics in UEFA Euro and other sports contexts [7][8].

The research by Elmiligi & Saad presents a novel hybrid approach combining machine learning and statistical methods to predict soccer match outcomes [9]. Analysing a dataset comprising over 200,000 match results from 2000/2001 to 2016/2017, the research explores various features including team and player statistics, home/away advantage, and data recency. Two hybrid models are developed, with the best achieving 46.6% prediction accuracy on a test set. The study also evaluates hypotheses regarding feature engineering, finding no significant improvement with recent match data or separate models for each league [9]. Additionally, the research plans to extend its analysis to other sports and conduct

comparative feature significance studies. This work contributes to advancing predictive modelling in sports and lays the groundwork for future research directions. Another study delves into predicting football match outcomes, focusing on the 2022 FIFA World Cup, leveraging Exploratory Data Analysis (EDA) and various machine learning algorithms. Notably, Random Forests, Decision Trees, K-Nearest Neighbours, XGBoost, and Gradient Boosting are tested, with XGBoost and Gradient Booster achieving the highest average accuracy of 98.34%. The study introduces a novel approach combining EDA and machine learning to address the challenges of sports match prediction, proposing Multi-output Regressor as a solution. It suggests that this method could accurately forecast sporting event outcomes and encourages further research into incorporating additional factors like current world ranking and new age metrics. The findings contribute to advancing predictive modelling in football and offer potential avenues for enhancing prediction accuracy in future studies [10]. The research by Athish et al. [11] explores the application of the Bayesian approach in predicting soccer match outcomes, leveraging authentic squad information and match results sourced from platforms like Kaggle and Sofifa.com. The Gaussian Naive Bayes model demonstrates 85.43% accuracy in match result prediction, surpassing the 79.81% accuracy achieved by the Decision Tree Classifier. The study offers a tool for users to assess team probabilities in tournaments, although it emphasizes individual discretion in betting due to uncertainties inherent in sports outcomes. The findings contribute to the understanding of machine learning techniques in soccer prediction and provide a basis for further research in the field. The studies discussed present diverse methodologies and approaches for predicting soccer match outcomes using machine learning and statistical techniques [9] [10][11]. Research by Elmiligi & Saad introduces a hybrid model that analyses team and player statistics, achieving a prediction accuracy of 46.6%. It emphasizes the importance of feature engineering and plans to extend its analysis to other sports [9]. In contrast, Majumdar and team focus on the 2022 FIFA World Cup, employing exploratory data analysis and various machine learning algorithms. Their approach yields high accuracy, with XGBoost and Gradient Boosting achieving 98.34% [10]. This study proposes a novel method combining EDA and machine learning, suggesting avenues for further research. Athish et al. explores the Bayesian approach for predicting soccer match outcomes, achieving an accuracy of 85.43% with the Gaussian Naive Bayes model. This study provides insights into machine learning techniques for soccer prediction, emphasizing individual discretion in betting [11]. Overall, these studies contribute to advancing predictive modelling in sports and offer valuable insights for future research directions.

The studies by Chin et al. and Daundkar & Kandhway both employ machine learning techniques to enhance predictive capabilities in sports, focusing on ice hockey and NBA match outcomes, respectively. Chin and the team analysed various machine learning techniques using NHL data from 2015-2021, with Logistic Regression achieving the highest accuracy at 77.82% [12][13]. This study highlights the significance of incorporating match-specific data for improved predictive accuracy [12]. Conversely, Daundkar & Kandhway predict NBA match outcomes based on past team performances,

achieving a prediction accuracy of approximately 66%. This research underscores the relevance of machine learning in sports betting and offers insights into the predictive capabilities of historical data in forecasting NBA match outcomes, aligning closely with human expert accuracy [13]. Both studies contribute to advancing predictive analytics in sports, offering valuable implications for future research and applications [12][13]. The research by Kumar, et al. [7] introduces an advanced approach to football analysis by utilizing predictive data like expected goals instead of descriptive data like shots taken and goals scored. By applying fixed parameters on machine learning algorithms, the method aims to evaluate teams and players based on performance rather than results, enhancing scouting and strategy formation. Results indicate that the light XGBoost machine learning model provides a better match of shot quality, as measured by McFadden's pseudo-R-squared score. Incorporating a "big chance" component further improves assessment criteria, although this capability is not available in the dataset used. Feature importance measurements highlight variables crucial to the model's outputs, offering valuable insights for performance analysis and talent identification. Another study focuses on predicting halftime results and league winners in the English Premier League using classification models. Leveraging ensemble techniques, the study achieves 80% accuracy in halftime result prediction and up to 95% accuracy in league winner prediction [14]. Building upon previous work, the research incorporates additional features such as team form and form points, enhancing prediction accuracy. By training models at a match week level, the study offers insights into predicting league winners throughout the season. The findings highlight the potential of utilizing dynamic datasets and simple features for accurate football match predictions and league analysis [14].

The studies by Tiwari et al. and Jaeyalakshmi et al. explore machine learning techniques for predicting football match outcomes, albeit with different emphases [15][16]. Tiwari and the research team focus on utilizing Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM) to leverage the abundance of statistical football data, aiming to enhance prediction accuracy for various match-related information. The conclusion of this study highlights the superiority of LSTM-based RNNs over traditional methods, suggesting further improvements by incorporating player statistics and larger datasets [15]. Conversely, Jaeyalakshmi and the team introduce a machine learning approach for forecasting football match results, emphasizing feature selection, data imbalance treatment, and model generalization. Achieving over 81% accuracy with the Random Forest Algorithm, this study emphasises the unpredictable nature of football and advocates responsible betting, suggesting future enhancements through additional statistics incorporation [16]. Another study introduces a machine-learning model employing a random forest algorithm for predicting football player performance and optimizing fantasy football team line-ups. Back-testing on historical data yielded a Mean Square Error (MSE) of 4.921 and a Root Mean Square Error (RMSE) of 2.2275, with the model presenting the potential for profitability in fantasy football betting [17]. The analysis involves web scraping, data segregation, and hyperparameter tuning. Results showcase the

model's capability in team formation for different formations, with future scope including subscription services and incorporation of additional data sources to enhance predictive accuracy while acknowledging inherent limitations in predicting future events and relying exclusively on past performance data [17].

Proposing a data-driven approach, the study Al-Asadi & Tasdemir [18] aims to estimate football players' market values using machine learning algorithms applied to FIFA 20 video game data. Comparing four regression models, the research finds that random forest outperforms others in accuracy and error ratio. The results suggest potential applications in streamlining negotiations between clubs and players' agents by providing objective market value estimations. Further research avenues include integrating the model into FIFA games for player valuation and developing a calculator to assist gamers in making informed decisions. The methodology of this study demonstrates superiority over traditional approaches, offering practical implications beyond gaming simulations. Further one research introduces a system leveraging real-time feedback and advanced technologies to enhance football technique learning. It utilizes pose estimation with Media pipe and classification algorithms like the Dollarpy, KNN, RFE, and SVM, achieving varying accuracies [19]. Another article tackles the challenge of predicting football match outcomes for sports betting, employing machine learning methods and historical match statistics [20]. The models in this research yield a 65.26% accuracy rate, offering potential profitability. The study by Gifford & Bayrak [21] focuses on NFL game outcome prediction using decision trees and logistic regression. With turnover statistics as key predictors, the models achieve up to 83% accuracy, contributing insights for strategic decision-making in sports analytics. These studies collectively highlight the diverse applications of machine learning in sports and the potential for technology to enhance performance and decision-making in athletics.

B. Use of Automated Machine Learning (AutoML)

An article explored Automated Machine Learning (AutoML) as an end-to-end process for streamlining model development without manual intervention. The paper provides insights into AutoML segments and approaches, emphasizing its practical applicability in industry [22]. Furthermore, it discusses recent trends and suggests future research directions, advocating for a generalized AutoML pipeline and a central meta-learning framework. The study highlights the importance of advancing AutoML to address evolving challenges in machine learning model development, both in academia and industry. The study discussed the significance of Automated Machine Learning (AutoML) in mitigating the challenges of ML adoption, especially for small and medium-sized organizations. This paper highlights its diverse applications across industries and advocates for its potential in democratizing machine learning [23]. It suggests various research opportunities in Information Systems (IS), including qualitative and quantitative studies on AutoML adoption, the development of AutoML adoption theories, and the exploration of fairness and explainability concerns. Additionally, the authors underscore the importance of human-in-the-loop research and address the limitations and boundaries of AutoML

applications, emphasizing the role of IS researchers in advancing AutoML adoption in organizations.

The articles Truong et al. [24] and Ferreira et al. [25] investigated Automated Machine Learning (AutoML) tools' effectiveness but with different emphases. Truong and the team compare commercialized and open-source AutoML tools, highlighting varying strengths and weaknesses across datasets. This research emphasizes the absence of a single superior tool, indicating ongoing AutoML evolution [24]. In contrast, Ferreira and the research team conducted a study exclusively on open-source AutoML tools, focusing on supervised learning scenarios. The results of this research reveal the competitive performance of General Machine Learning (GML) AutoML tools, particularly in binary and regression tasks [25]. Both studies underscore the need for further advancements, Truong, et al in addressing gaps in AutoML pipeline support, and Ferreira et al in expanding comparisons to encompass more technologies and datasets, especially in big data contexts [24][25]. Another study presents a survey on automating the process of building machine learning models, particularly focusing on Combined Algorithm Selection and Hyperparameter tuning (CASH). It discusses the challenges of efficiently constructing high-quality models due to the vast amounts of data produced daily. The paper comprehensively reviews state-of-the-art efforts in AutoML frameworks and highlights research directions and challenges. By addressing these issues, the aim is to automate the machine-learning pipeline and reduce human intervention, catering to both researchers and practitioners in advancing the field [26]. The article [27] commences with an analysis of existing research in AutoML, hyperparameter tuning, and meta-learning. It highlights the lack of clear documentation and consensus on evaluation criteria in this field. The paper discusses the strengths and weaknesses of various approaches, emphasizing the need for further research to develop a fully automated industrial standard system. Assembling and meta-learning are proposed as effective methods for automating hyperparameter tuning. The authors aim to bridge gaps in existing solutions and plan to devise an architectural style for an efficient AutoML system based on accumulated knowledge and identified drawbacks. The article by Tsiakmaki et al. [28] focuses on applying automated machine learning (AutoML) in Educational Data Mining (EDM) to predict students' learning outcomes. It emphasizes interpretability by restricting the search space to tree-based and rule-based models. The study highlights that AutoML tools surpass default parameter values, especially in classification and regression tasks, highlighting the significance of transparent tools for educators. The findings suggest AutoML has the potential to aid early performance estimation and intervention strategies, offering promising avenues for enhancing academic outcomes in educational

environments. In contrast, Shi, et al. [29] present a domain-specific AutoML framework tailored for risk prediction and behaviour assessment in autonomous vehicles (AVs). The system in this research integrates unsupervised risk identification, feature learning with XGBoost, and model auto-tuning using Bayesian optimization. Evaluation of Next Generation Simulation (NGSIM) data demonstrates the framework's efficacy in distinguishing safe from risky behaviours, thus enhancing risk decision-making in Autonomous Vehicle (AVs). Additionally, it provides insights into sensor configurations and data mining, contributing to AV safety and design improvements.

In a nutshell even while previous research shows a variety of approaches and developments in basketball, cricket, and football sports prediction, a large number of studies continue to mostly rely on manual machine learning techniques, neglecting the benefits of automated approaches. By automating feature engineering, model selection, and hyperparameter tuning, the combination of automated machine learning (AutoML) tools with conventional techniques offers promising potential to increase prediction efficiency and accuracy. Machine learning techniques, coupled with automated machine learning (AutoML) tools, showcase promising capabilities in predicting match outcomes and enhancing sports analytics. Our research addresses this gap.

IV. RESEARCH METHODOLOGY

This study examines a vast dataset comprising historical records from international football matches alongside various team performance metrics, intending to use state-of-the-art machine learning techniques to predict the winner of UEFA Euro 2024. The primary objective is to develop a prediction model capable of accurately assessing the likelihood of victory for each participating team, providing valuable insights to analysts, bookmakers, coaches, and athletes. Notably, the study employed Artificial Neural Networks (ANNs) to construct a model for forecasting tournament match outcomes. To ensure high prediction accuracy, this model underwent cross-validation to prevent overfitting, refinement to improve generalization, and extensive training on substantial datasets [30].

The methodology section evaluates methodological choices based on literature and previous studies, such as ensemble methods for improved predictive accuracy in sports analytics. It also discusses potential limitations like the unpredictability of events and potential biases in historical data. Obstacles and solutions include data quality issues, model over-fitting, and changes in team dynamics. Robust pre-processing techniques, regularization and pruning of decision trees, and ongoing data updates are employed. Fig. 1 depicts the research methods used in this study.

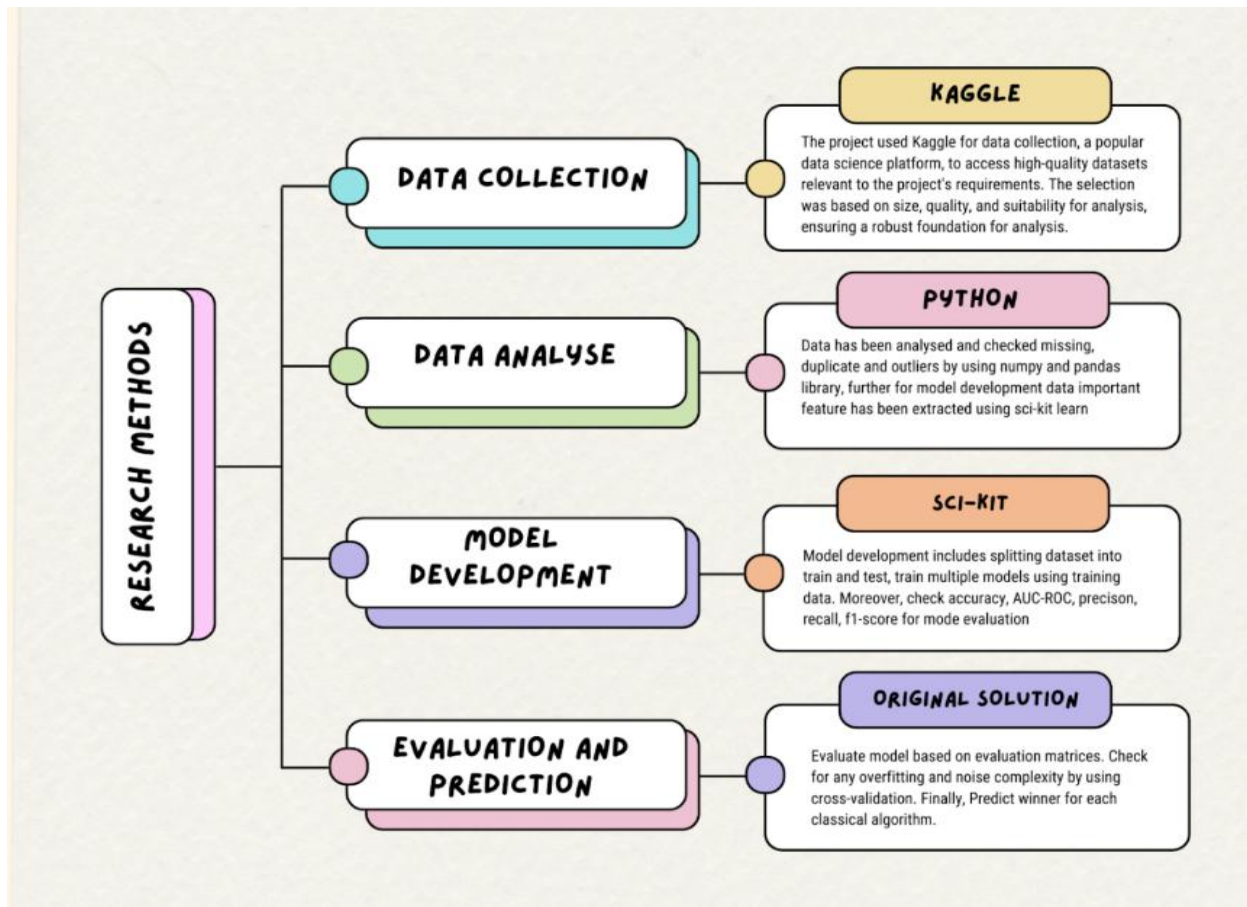


Fig. 1. Research methodology.

A. Chosen Approach

Our study adopts a quantitative research design, utilizing the numerical nature of data and modern computational methods to address the complexities inherent in predictive analytics. By leveraging a comprehensive dataset that includes player stats, team performance metrics, and past match results, we employ machine learning—a sophisticated branch of artificial intelligence—to manage and analyze this vast amount of information efficiently. Our approach integrates traditional statistical methods with advanced machine learning models, allowing us to uncover intricate patterns and interactions that conventional techniques might overlook. This dual strategy enhances the reliability and precision of our predictions, ensuring a robust analysis conducive to accurate forecasting.

In our investigation, we use well-established machine learning algorithms such as Random Forest and Support Vector Machines (SVM). Furthermore, we explore the potential of Automated Machine Learning (AutoML), a revolutionary development in predictive analysis. AutoML automates the selection, application, and refinement of various machine learning models, significantly streamlining the analytical process. This automation reduces the time and expertise required to implement complex models, as AutoML efficiently evaluates numerous algorithms and their configurations to identify the most suitable one for our data.

By integrating AutoML, our study enriches the toolkit available for predictive analytics and sets a standard for future research. The use of AutoML not only enhances predictive accuracy but also demonstrates the potential of advanced automation in pushing the boundaries of data forecasting. This methodology presents an exciting frontier for further exploration, promising significant advancements in the field of predictive analysis.

V. DATASET COLLECTION

For predictive modelling, this study utilized historical European football match data sourced from Kaggle (<https://www.kaggle.com/datasets/mahadinour/international-football-matches>) [31]. The dataset encompassed match outcomes, team metrics, and player performance statistics. To maintain relevance, the data underwent filtration to focus solely on European matches from the UEFA Euro 2024 tournament. Python's Pandas library facilitated efficient filtering based on competition type and team geographical locations. This dataset played a pivotal role in achieving the study's goals. To ensure research reproducibility, a detailed data dictionary will be made available alongside the study, outlining each collected variable and its origin.

A. Methods used to Analyse Collected Data

1) *Feature engineering*: The study will employ sophisticated feature engineering techniques to develop new

variables that could influence match outcomes, such as team morale indices or fatigue levels.

2) *Model training and testing*: Utilizing cross-validation techniques to partition the data and evaluate the model's performance across different subsets, ensuring the model's ability to generalize to new data.

3) *Hyperparameter optimization*: Advanced techniques like Bayesian Optimization will be used for tuning the models to find the optimal configuration of parameters.

B. Experimental Set-Up and Results

Exploratory Data Analysis (EDA) is a crucial stage in data analysis, utilizing statistical and visualisation methods to understand the dataset's structure, identify trends, and gain insights. It helps identify patterns, outliers, and correlations,

and aids in initial data processing, feature selection, and machine learning techniques. EDA is essential for data science and machine learning.

Fig. 2 shows the dataset, containing 2391 rows and 25 columns, which includes information about UEFA Euro qualification tournament football matches. It includes team details, match date and location, and final score. Analysing this data can reveal trends, patterns, and identify factors affecting national team success.

To find out a concise summary of the given dataframe, a set of Python commands like `info()`, `dtypes`, `select_dtypes`, etc were used. This helps to understand the descriptive statistics for "object" datatype, which represents a string or categorical data. Fig. 3 describes the same.

date	home_tea	away_tea	home_tea	away_tea	home_tea	away_tea	home_tea	away_tea	home_tea	away_tea	home_tea	away_tea	home_tea	away_tea	home_tea	away_tea	home_tea	away_tea	home_tea	away_tea	home_tea	away_tea	home_tea	away_tea	
20-04-1994	Northern	Liechtenst	Europe	Europe	35	160	0	0	4	1	UEFA Eurc	Belfast	Northern	FALSE	No	Win									
04-09-1994	Estonia	Croatia	Europe	Europe	111	90	0	0	0	2	UEFA Eurc	Tallinn	Estonia	FALSE	No	Lose									
04-09-1994	Israel	Poland	Europe	Europe	52	32	0	0	2	1	UEFA Eurc	Ramat-Ga	Israel	FALSE	No	Win									
06-09-1994	Czech Rep	Malta	Europe	Europe	44	69	0	0	6	1	UEFA Eurc	Ostrava	Czech Rep	FALSE	No	Win									
07-09-1994	Belgium	Armenia	Europe	Europe	20	159	0	0	2	0	UEFA Eurc	Brussels	Belgium	FALSE	No	Win									
07-09-1994	Cyprus	Spain	Europe	Europe	71	6	0	0	1	2	UEFA Eurc	Limassol	Cyprus	FALSE	No	Lose									
07-09-1994	Faroe Islai	Greece	Europe	Europe	127	34	0	0	1	5	UEFA Eurc	Toftir	Faroe Islai	FALSE	No	Lose									
07-09-1994	Finland	Scotland	Europe	Europe	45	33	0	0	0	2	UEFA Eurc	Helsinki	Finland	FALSE	No	Lose									
07-09-1994	Georgia	Moldova	Europe	Europe	124	149	0	0	0	1	UEFA Eurc	Tbilisi	Georgia	FALSE	No	Lose									
07-09-1994	Hungary	Turkey	Europe	Europe	55	57	0	0	2	2	UEFA Eurc	Budapest	Hungary	FALSE	No	Draw									
07-09-1994	Iceland	Sweden	Europe	Europe	42	3	0	0	0	1	UEFA Eurc	Reykjavá	Iceland	FALSE	No	Lose									
07-09-1994	Latvia	Republic c	Europe	Europe	86	13	0	0	0	3	UEFA Eurc	Riga	Latvia	FALSE	No	Lose									
07-09-1994	Liechtenst	Austria	Europe	Europe	151	37	0	0	0	4	UEFA Eurc	Eschen	Liechtenst	FALSE	No	Lose									
07-09-1994	Luxembol	Netherlan	Europe	Europe	116	5	0	0	0	4	UEFA Eurc	Luxembou	Luxembol	FALSE	No	Lose									
07-09-1994	North Mar	Denmark	Europe	Europe	117	12	0	0	1	1	UEFA Eurc	Skopje	North Mar	FALSE	No	Draw									
07-09-1994	Northern	Portugal	Europe	Europe	39	25	0	0	1	2	UEFA Eurc	Belfast	Northern	FALSE	No	Lose									
07-09-1994	Norway	Belarus	Europe	Europe	8	142	0	0	1	0	UEFA Eurc	Oslo	Norway	FALSE	No	Win									
07-09-1994	Romania	Azerbaijan	Europe	Europe	7	170	0	0	3	0	UEFA Eurc	Bucharest	Romania	FALSE	No	Win									
07-09-1994	Slovakia	France	Europe	Europe	43	16	0	0	0	0	UEFA Eurc	Bratislava	Slovakia	FALSE	No	Draw									
07-09-1994	Slovenia	Italy	Europe	Europe	75	2	0	0	1	1	UEFA Eurc	Maribor	Slovenia	FALSE	No	Draw									
07-09-1994	Ukraine	Lithuania	Europe	Europe	79	84	0	0	0	2	UEFA Eurc	Kyiv	Ukraine	FALSE	No	Lose									
07-09-1994	Wales	Albania	Europe	Europe	35	101	0	0	2	0	UEFA Eurc	Cardiff	Wales	FALSE	No	Win									

Fig. 2. Chosen dataset.

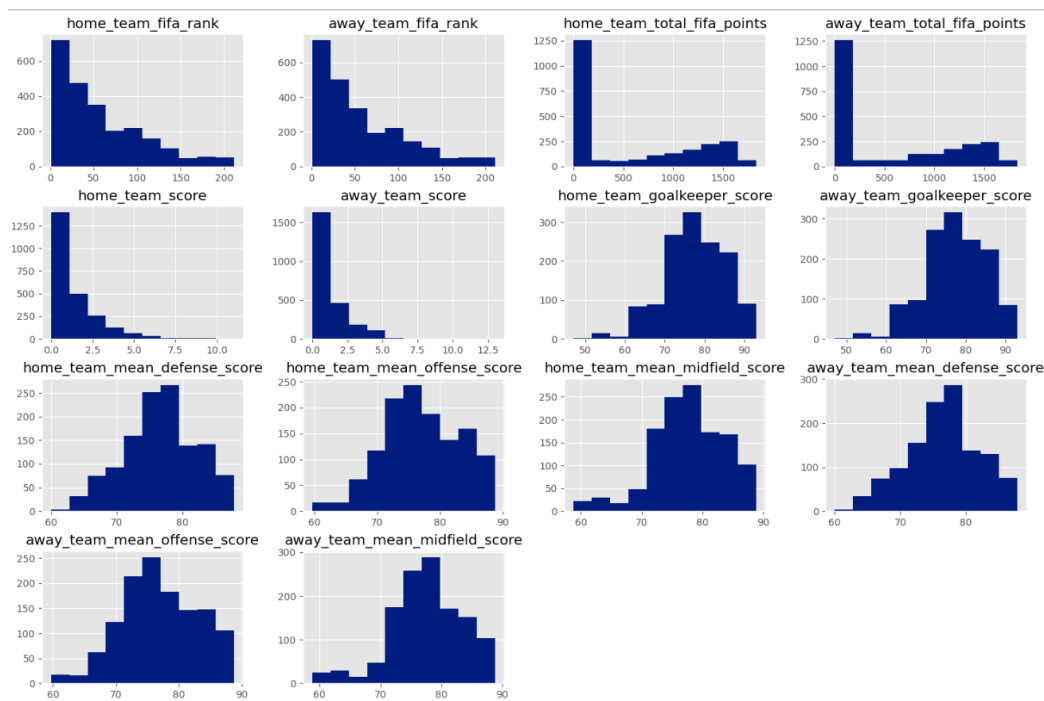


Fig. 3. Histogram plots.

Fig. 4 shows histograms plotted on various columns in a dataset, revealing potential patterns and frequency distributions within the variables.

C. Data Pre-Processing

Data pre-processing involves various procedures like transformation, cleansing, reduction, normalisation, and integration. It involves handling NaN data, managing noisy data, and error fixation. Data transformation converts data into a more intelligible format, while data integration combines data from multiple sources. Data normalisation scales data for similar distribution. Data pre-processing is crucial for machine learning (ML) algorithms to ensure suitable and high-quality data.

1) Visualizing NaN: Fig. 4 illustrate the findings done as part of a null check for the given dataset. Each heatmap corresponds to a data cell, and the colour of each heatmap cell indicates whether or not NaN values are present. The density of NaN values in a given column or row is larger when the colour

is darker. It is noted that the following columns in the dataset contain more NaN values than the other columns –

- a) home_team_goalkeeper_score
- b) away_team_goalkeeper_score
- c) home_team_mean_defense_score
- d) away_team_mean_defense_score
- e) home_team_mean_offense_score
- f) away_team_mean_offense_score
- g) home_team_mean_midfield_score
- h) away_team_mean_midfield_score

2) Visualizing missing data across different years: Fig. 5 displays the trend of missing data across different years which would help to identify years with higher proportion of missing data (data quality assessment). This is a scatter plot and each data point on the plot corresponds to the proportion of missing values for a specific year.

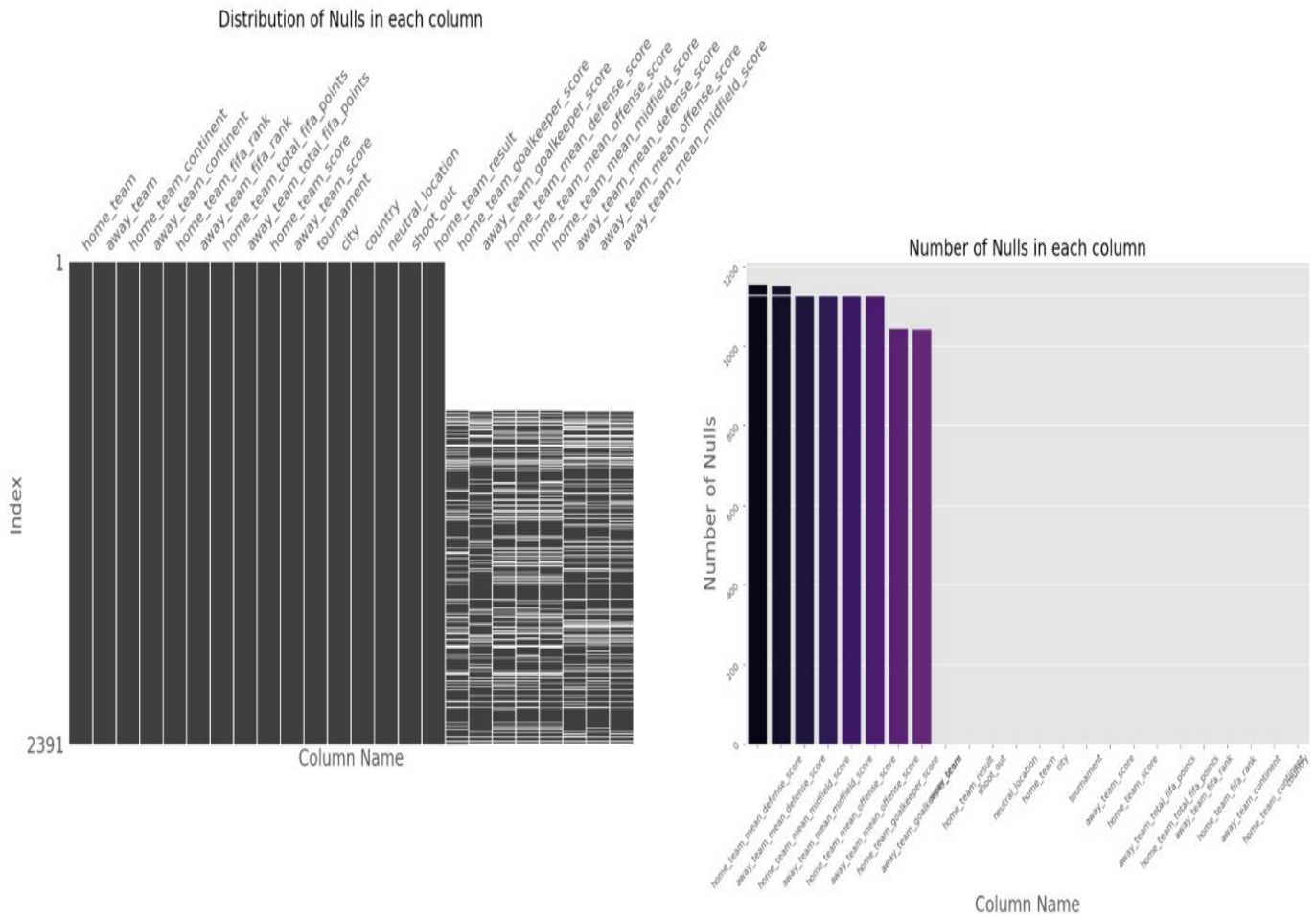


Fig. 4. NaN visualization using HeatMap.

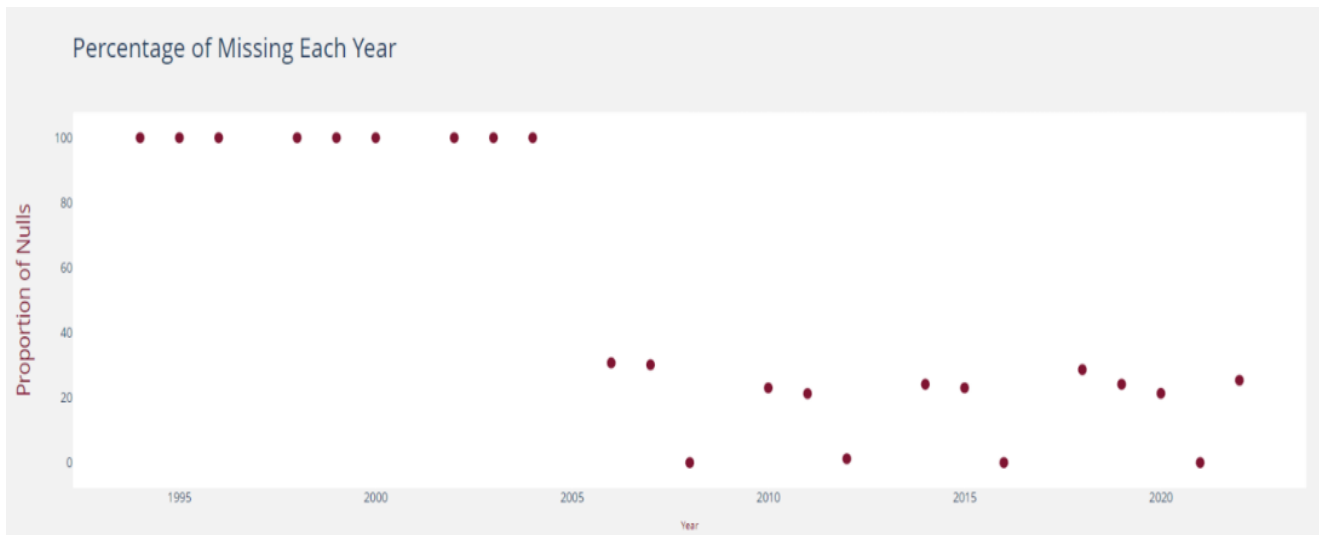


Fig. 5. Visualizing missing data across different years (Scatter plot).

3) *Checking for duplicates*: This is a check to identify the duplicates in the given dataset. As per the result, we did not find any duplicates in the dataframe.

4) *Outliers detection*: Fig. 6 represents a parallel coordinate plot. This plot helps us visualize multivariate data

by showing multiple variables or attributes as parallel vertical axes. An individual data point is depicted by each polyline by connecting its values across different variables. This technique helps to identify similarities, patterns or relationships between different variables in the given dataset.

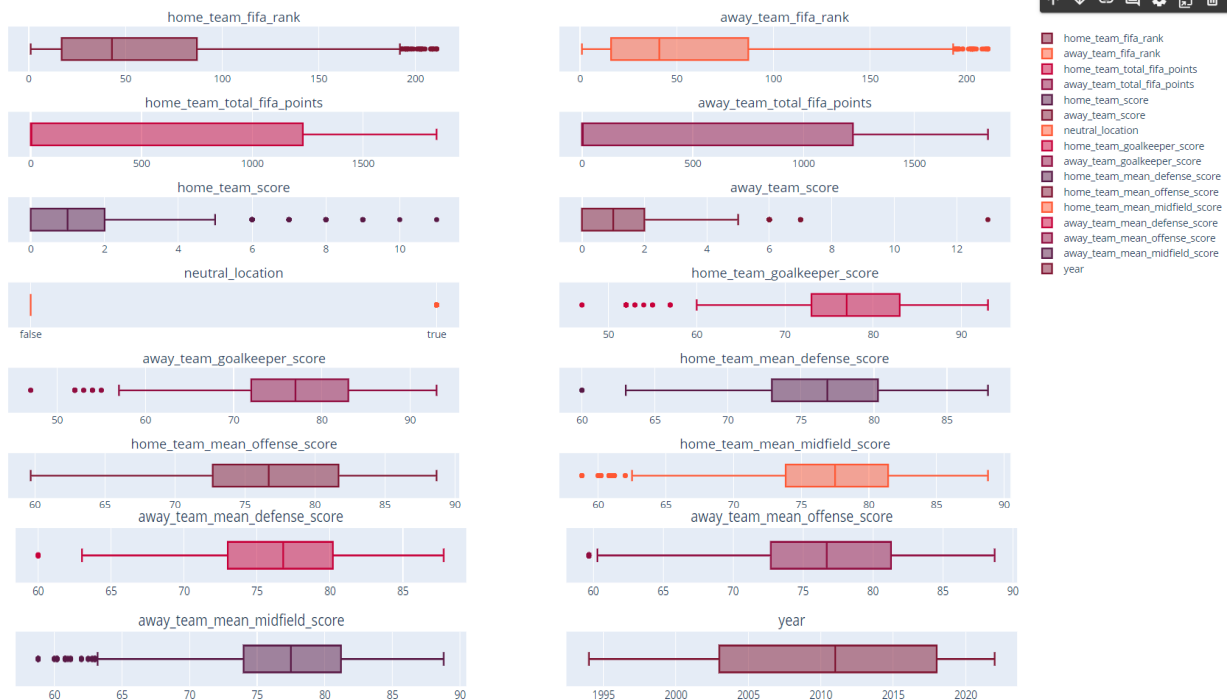


Fig. 6. Identifying similarities, patterns or relationships between different variables.



Fig. 7. Team points distribution with time.

Multiple visualizations between team points distribution with time is illustrated in Fig. 7. Overall point distribution for home and away teams are depicted in top histograms and the variation in average points per year for home and away teams are depicted in the bottom histograms.

5) *Data analysis and transformation*: To make the dataset more useful for analysis, we added a few new columns to summarize the performance of the home team. These columns—`home_win`, `home_draw`, and `home_lose`—were created based on the results of each match. This made it easier to see whether the home team won, drew, or lost a game, simplifying the dataset for comparison across different teams.

We also added several other columns to provide a richer, team-level view of the data, including:

a) *Total points for home and away teams*: This is to capture the sum of FIFA points for each team across all matches.

b) *Average FIFA points per team*: By averaging the FIFA points of both the home and away teams, we got a clearer idea of the overall strength of each team.

c) *Team rankings and performance metrics*: We introduced columns like the median FIFA rank, home and away goal scores, and goals conceded to offer more detailed insights into how each team performed.

Additionally, we included information on the continent each home team comes from, allowing us to identify any geographical trends in performance. A custom function helped us find the most frequent (mode) continent for each team.

To address missing data, we used the backward fill (bfill) method, which ensured that any gaps were filled with the most recent available data. This was especially useful in cases like carrying forward a goalkeeper's score for the next match if the data was incomplete.

VI. COMPARATIVE ANALYSIS OF MANUAL ML AND AUTO ML

A. Case 1 – Using Manual Machine Learning

1) *Model selection and training*: Google Colaboratory, a free cloud-based environment, is the platform used for writing and running the Python code, inclusive of machine learning models. The technique used here is One-Hot encoding. This technique or program is made more scalable by creating a generic function that will fit the data and forecast the result based on the chosen methods. The accuracy levels and models' correctness are also specified for each model instance.

Algorithm 1 – Random Forest Classifier

Supervised machine learning uses ensemble technique to improve model performance by combining multiple classifiers. This approach boosts forecasting accuracy by using multiple decision trees on different datasets, utilizing feature randomization and bagging [32]. For Random Forest, data pre-processed with One-Hot Encoding yields an accuracy of 70%, as illustrated in Fig. 8.



Fig. 8. Model evaluation for random forest.

Algorithm 2 – XG Boost Classifier

Sequential decision trees use XGBoost, assigning weights to independent variables. The second decision tree is used after adjusting the weight of miscalculated components, allowing faster training of large datasets through parallel processing [33]. As illustrated in Fig. 9, XGBoost reported 70% accuracy for data that has undergone One Hot Encoding pre-processing.

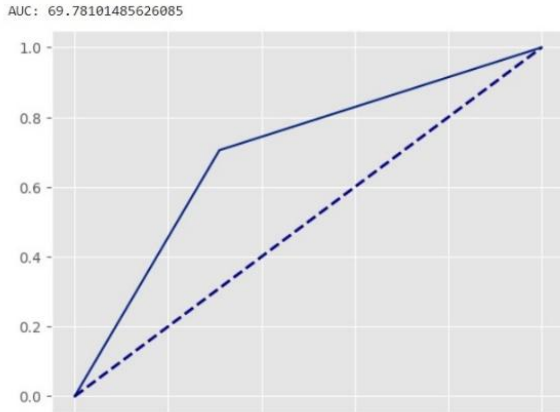


Fig. 9. Model evaluation for XG boost.

Algorithm 3 – Support Vector Machine

This model uses supervised learning algorithms to solve regression, detecting outliers and complex classification by executing optimal data transformations that set boundaries between data points on predefined classes or labels. This model has an accuracy of 71% as shown in Fig. 10.

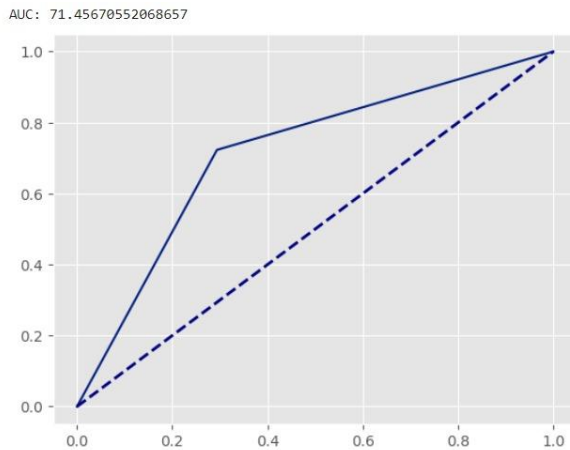


Fig. 10. Model evaluation for SVM.

Algorithm 4 – AdaBoost Classifier

AdaBoost is an iterative ensemble boosting classifier that combines inefficient classifiers to increase precision. It can be trained on a dataset, but its main drawback is hindering parallelization. It requires interactive training on various weighted instances and limiting training errors for perfect matches [34].

Fig. 11 illustrates the AdaBoost Classifier accuracy for data pre-processed with One-Hot Encoding, which was 71%.

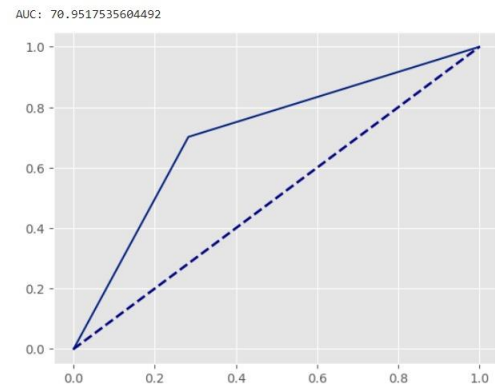


Fig. 11. Model evaluation for AdaBoost.

Algorithm 5 – Logistic Regression

This is a data analysis technique which is used to find out the dependency or relationship between two data factors. This relationship is then further used to determine or predict the value of the other factor. This results in a finite number of outcomes. Fig. 12 illustrates the accuracy check for this model, which is 70%.

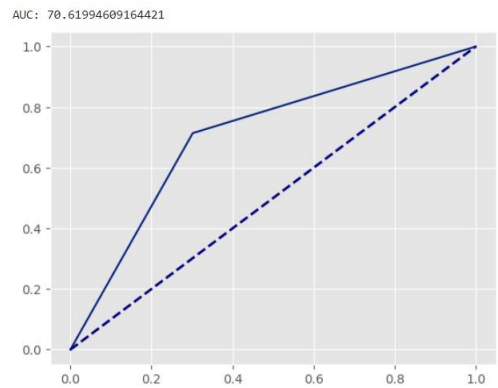


Fig. 12. Model evaluation for logistic regression.

Algorithm 6 – K-Nearest Neighbour Classifier

KNN is a supervised learning algorithm, which is non-parametric and is used in both classification as well as regression. Refer to Fig. 13.

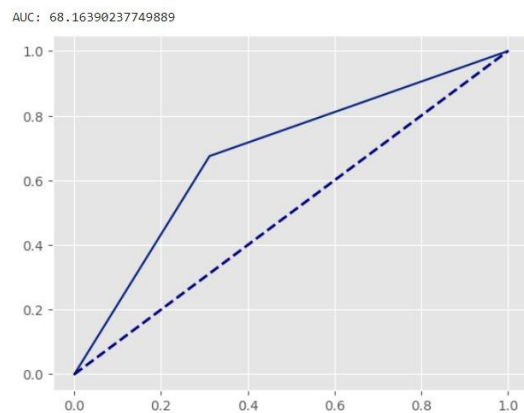


Fig. 13. Model evaluation for KNN.

Algorithm 7 – Gaussian Naive Bayes

This is a machine learning classification technique which is based on a probabilistic approach. Here, each class is assumed to follow a normal distribution. Refer to Fig. 14.

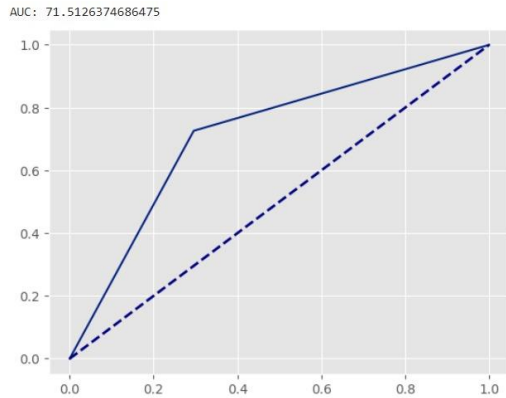


Fig. 14. Model evaluation for gaussian naive bayes.

1) Model Evaluation and Visualization

a) *Model simulation:* Table I, shows a machine learning model calculating the "home_team" and "away_team" means for a tournament. The simulation is run 1000 times, saving outcomes in variables for each round, quarter-finals, semi-finals, and finals. The model calculates the home team's victory probability and predicts match outcomes. The simulation continues until the tournament winner is determined.

Model 1 –Logistic Regression is the model used in this instance.

Out of the sixteen teams, eight teams were chosen for the quarter-finals, while the other 8 teams. From the eight teams, four were chosen to go to the semi-finals, while the other four teams failed. For the remaining rounds, the same marking procedure is used. Refer to Table II, where the green ones show the winning teams in each round and the pink ones are the failed ones.

TABLE I. ML MODEL SIMULATION OF TEAM MEANS AND MATCH OUTCOME PREDICTIONS

Away Team	Away Team FIFA Rank	Team Total FIFA Points
Albania	73.166667	492.977444
Austria	34.675	776.971667
Belgium	19.133333	922.177778
Croatia	18.576923	647.143086
Czech Republic	23.966102	542.766882
Denmark	16.48	580.456254
England	8.111111	760.278236
France	6.076923	675.307329
Georgia	86.560976	409.585366
Germany	6.351852	609.736055
Hungary	49.282609	580.877127
Italy	9.350877	641.229123
Netherlands	8.468085	684.159543
Poland	26.837209	634.290698
Portugal	11.12963	731.470994
Romania	23.313725	550.907308
Scotland	37.869565	556.342391
Serbia	29.655172	822.627949
Slovakia	34.666667	668.602163
Slovenia	57.065217	516.256522
Spain	6.067797	677.954132

TABLE II. USING LOGISTIC REGRESSION

Round 16	Quarter-Finals	Semi-Finals	Finals
Czech Republic	England	France	Portugal
Denmark			
Italy			
France	Italy		
Switzerland			
Portugal			
England	France	Portugal	Portugal
Netherlands			
Germany			
Spain	Portugal		
Belgium			
Romania			
Ukraine	Netherlands	Italy	France
Slovakia			
Croatia			
Scotland	Denmark		
Poland			
Serbia			
Turkey	Czech Republic	England	France
Hungary			
Austria			
Albania	Switzerland		
Slovenia			
Georgia			

The UEFA European Championship was won by Portugal among the two teams, while France finished in second place. The UEFA European Championship 2024 has been won by Portugal, according to the Logistic Regression model.

Model 2 – Random Forest is the model used in this instance.

Out of the sixteen teams, eight teams were chosen for the quarter-finals, while the other 8 teams failed. 4 of the 8 teams were qualified to the semi-finals, and the other four teams were unsuccessful. The remaining rounds are marked using the same criteria. Refer to Table III, where the green ones show the winning teams in each round and the pink ones are the failed ones.

TABLE III. USING A RANDOM FOREST ALGORITHM

Round 16	Quarter-Finals	Semi-Finals	Finals
Italy	France	Portugal	Portugal
Netherlands			
Germany			
France	Italy		
Ukraine			
Czech Republic			
Portugal	Portugal	Switzerland	Portugal
Denmark			
Switzerland			
Spain	Switzerland		
Belgium			
England			
Serbia	Germany	Italy	Switzerland
Slovakia			
Croatia			
Hungary	Netherlands		
Slovenia			
Turkey			
Georgia	Czech Republic	France	Switzerland
Romania			
Austria			
Scotland	Denmark		
Poland			
Albania			

The UEFA European Championship was won by Portugal amongst the two teams, while Switzerland finished in second

place. The UEFA European Championship 2024 has been won by Portugal, according to the Random Forest model.

Model 3 – Gaussian Naive Bayes is the model used in this instance.

8 out of 16 teams qualified for the quarter-finals, while the other 8 teams failed. Out of the eight teams, four were chosen

to go to the semi-finals, while the other four teams did not advance. For the remaining rounds, the same marking procedure is used. Refer to Table IV, where the green ones show the winning teams in each round and the pink ones are the failed ones

TABLE IV. USING GAUSSIAN NAÏVE BAYES

Round 16	Quarter-Finals	Semi-Finals	Finals	
Czech Republic	Italy	England	Portugal	
Italy				
Switzerland				
Portugal	Portugal			
Denmark				
France				
Netherlands	France	Portugal	England	
England				
Germany				
Belgium	England	Portugal		England
Ukraine				
Spain				
Croatia	Netherlands	Italy	England	
Romania				
Slovakia				
Scotland	Denmark	France		England
Serbia				
Poland				
Austria	Switzerland	France	England	
Slovenia				
Turkey				
Hungary	Czech Republic	France		England
Albania				
Georgia				

The UEFA European Championship was won by Portugal, and England finished as the runner-up. The UEFA European Championship 2024 has been won by Portugal, according to the Gaussian Naive Bayes model.

Model 4 – XG Boost is the model used in this instance.

8 out of 16 teams qualified for the quarter-finals, while the other 8 teams failed. Out of the eight teams, four were chosen to go to the semi-finals, while the other four teams did not advance. For the remaining rounds, the same marking procedure is used. Refer to Table V, where the green ones show the winning teams in each round and the pink ones are the failed ones.

TABLE V. USING XG BOOST

Round 16	Quarter-Finals	Semi-Finals	Finals	
Italy	Netherlands	France	Portugal	
Germany				
France				
Portugal	Portugal			
Czech Republic				
Netherlands				
Ukraine	Belgium	Portugal	France	
Belgium				
Spain				
Denmark	France	Netherlands		France
Switzerland				
England				
Serbia	Italy	Netherlands	France	
Hungary				
Slovenia				
Croatia	Ukraine	Belgium		France
Slovakia				
Romania				
Poland	Germany	Belgium	France	
Georgia				
Turkey				
Austria	Czech Republic	Belgium		France
Scotland				
Albania				

The UEFA European Championship was won by Portugal, and France finished as the runner-up. The UEFA European Championship 2024 has been won by Portugal, according to the XG Boost model.

Model 5 – SVM is the model used in this instance.

8 out of 16 teams qualified for the quarter-finals, while the other 8 teams failed. Out of the eight teams, four were chosen to go to the semi-finals, while the other four teams did not advance. For the remaining rounds, the same marking procedure is used. Refer to Table VI, where the green ones show the winning teams in each round and the pink ones are the failed ones.

TABLE VI. USING SVM

Round 16	Quarter-Finals	Semi-Finals	Finals
Czech Republic	Netherlands	Netherlands	Italy
Italy			
Denmark			
France			
Netherlands	Switzerland	Italy	
Portugal			
Switzerland			
Germany			
Belgium	Italy	Germany	
England			
Spain			
Ukraine			
Romania	Portugal	Germany	Netherlands
Croatia			
Serbia			
Slovakia			
Scotland	France	Switzerland	
Hungary			
Turkey			
Poland			
Slovenia	Denmark	Switzerland	
Austria			
Albania			
Georgia			
	Czech Republic		

The UEFA European Championship was won by Italy, and the Netherlands finished as the runner-up. The UEFA European Championship 2024 has been won by Italy, according to the SVM model.

Model 6 – KNN is the model used in this instance.

8 out of 16 teams qualified for the quarter-finals, while the other 8 teams failed. Out of the eight teams, four were chosen to go to the semi-finals, while the other four teams did not advance. For the remaining rounds, the same marking procedure is used. Refer to Table VII, where the green ones show the winning teams in each round and the pink ones are the failed ones

TABLE VII. USING KNN

Round 16	Quarter-Finals	Semi-Finals	Finals
Germany	Portugal	Netherlands	Netherlands
Italy			
France			
Portugal			
Czech Republic	Netherlands	Germany	
Netherlands			
Belgium			
Switzerland			
Croatia	Switzerland	Germany	
Serbia			
Ukraine			
Denmark			
Slovenia	Italy	Switzerland	Germany
Romania			
England			
Spain			
Hungary	Belgium	Portugal	
Slovakia			
Poland			
Turkey			
Scotland	Czech Republic		
Austria			
Albania			
Georgia			
	France		

The UEFA European Championship was won by the Netherlands, and Germany finished as the runner-up. The UEFA European Championship 2024 has been won by the Netherlands, according to the KNN model.

Model 7 – AdaBoost is the model used in this instance.

8 out of 16 teams were qualified to the quarter-finals, while the other 8 teams failed. Out of the eight teams, four were chosen to go to the semi-finals, while the other four teams did not advance. For the remaining rounds, the same marking procedure is used. Refer to Table VIII, where the green ones show the winning teams in each round and the pink ones are the failed ones.

TABLE VIII. USING ADABOOST

Round 16	Quarter-Finals	Semi-Finals	Finals
Czech Republic	Italy	France	France
England			
Switzerland			
Turkey	France		
France			
Portugal			
Slovenia	England	Portugal	
Italy			
Poland			
Croatia	Portugal		
Belgium			
Germany			
Scotland	Switzerland	England	Portugal
Slovakia			
Spain			
Albania	Turkey		
Netherlands			
Denmark			
Ukraine	Czech Republic	Italy	
Austria			
Romania			
Hungary	Slovenia		
Serbia			
Georgia			

The UEFA European Championship was won by France, and Portugal finished as the runner-up. The UEFA European Championship 2024 has been won by France, according to the AdaBoost model.

In order to forecast the winner of the UEFA European Championship 2024, 7 models were utilised. The total results are depicted in Table IX.

TABLE IX. WINNER PREDICTION USING MANUAL MACHINE LEARNING

Model	Winner	Runner-Up
Logistic Regression	Portugal	France
Random Forest	Portugal	Switzerland
Gaussian Naive Bayes	Portugal	England
XG Boost	Portugal	France
SVM	Italy	Netherlands
KNN	Netherlands	Germany
AdaBoost	France	Portugal

B. Case 2 – Using AutoML

For performing AutoML, the pycaret library is to be installed in Google Colab.

In automated machine learning, while setting up the environment, the module itself runs a series of pre-processing and data transformation steps. After the environment is set up, the performance metrics of various classification models are evaluated on the transformed data. All the pre-processing information regarding AutoML is given in Table X.

TABLE X. PRE-PROCESSING AND SETUP

Description	Value
Target	is_won
Target Type	Binary
Original Data Shape	(2391, 43)
Transformed Data Shape	(2391, 47)
Train Set Shape	(1673, 47)
Test Set Shape	(718, 47)
Numeric Features	26
Categorical Features	10
Preprocess	True
Imputation Type	Simple
Numeric Imputation	Mean
Categorical Imputation	Mode
Maximum One-Hot Encoding	25
Encoding Method	None
Fold Generator	StratifiedKfold
Number of Folds	10
CPU Jobs	-1 (All CPUs)
Use GPU	False
Log Experiment	False
Experiment Name	clf-default-name

AutoML itself identifies the best model for this particular dataset. In this case, Logistic Regression is chosen as the best model, as using the 'lbfgs' solver, the Logistic Regression model was optimized with typical L2 regularization (penalty='l2') to avoid overfitting. The intercept term (fit_intercept=True) was included in the model, and the regularization strength was adjusted to 1 (C=1.0). A maximum of 1000 iterations were performed, with a tolerance value of

0.0001 to guarantee convergence. Reproducibility was ensured by using random_state=6250, and the imbalance was handled without the use of class weights (class_weight=None). For the classification challenge, this setup produced a reliable and effective model.

Fig. 15 shows the plotted ROC curve and the confusion matrix for the logistic regression model.

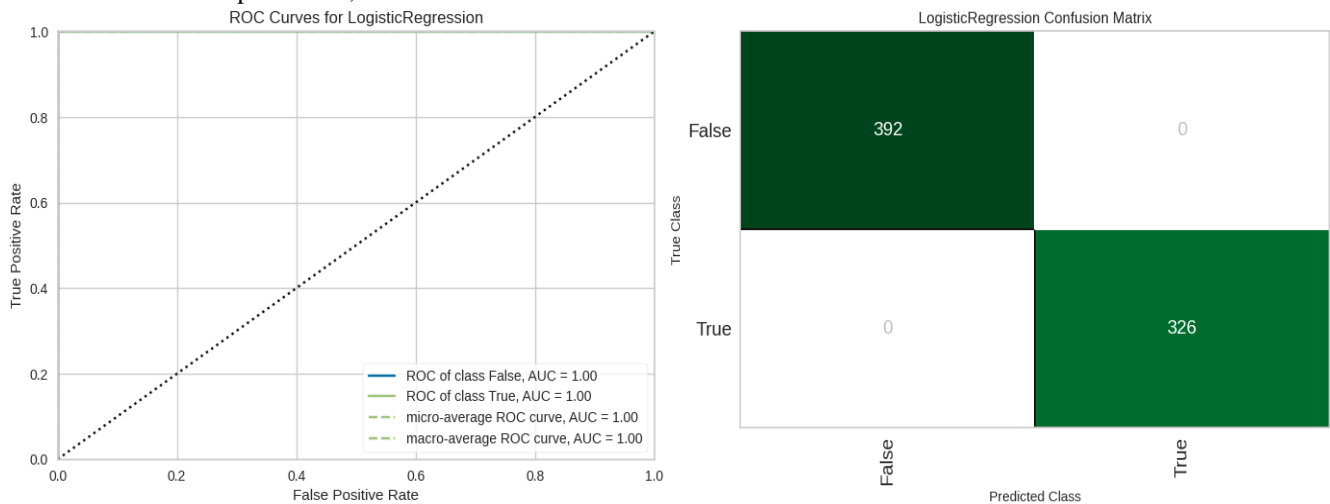


Fig. 15. ROC curve and confusion matrix.

Model Chosen by Auto ML – Logistic Regression

1) *Logistic regression model simulation*: Applying the same simulation logic for autoML as well (same as used in manual ML). 8 out of 16 teams qualified for the quarter-finals, while the other 8 teams failed. Out of the eight teams, four were

chosen to go to the semi-finals, while the other four teams did not advance. For the remaining rounds, the same marking procedure is used. Refer to Table XI, where the green ones show the winning teams in each round and the pink ones are the failed ones.

TABLE XI. USING AUTOML (LOGISTIC REGRESSION)

Round 16	Quarter-Finals	Semi-Finals	Finals
Czech Republic	Denmark	Italy	Portugal
Denmark			
Italy			
Switzerland	Portugal		
France			
England			
Netherlands	England		
Portugal			
Germany			
Belgium	Italy		
Croatia			
Ukraine			
Romania	Switzerland	England	
Spain			
Slovakia			
Scotland	France		
Turkey			
Poland			
Serbia	Czech Republic	Denmark	
Austria			
Hungary			
Albania	Netherlands		
Slovenia			
Georgia			

The UEFA European Championship was won by Portugal, and Italy finished as the runner-up. The UEFA European Championship 2024 has been won by Portugal, according to the Logistic Regression model via AutoML.

precision, recall, F1-score, accuracy, and AUC (Area under the curve) for the manual approach. It evaluates both models' ability to predict positive and negative classes and also identifies the more effective model. Here, 0 represents the "False" class and 1 represents the "True" class.

VII. COMPARATIVE RESULTS AND EVALUATION

Table XII given below compares the performance of various classification models based on the key metrics such as

TABLE XII. SUMMARY OF CLASSIFICATION PERFORMANCE METRICS (MANUAL)

Metric		Random Forest	XG Boost	SVM	AdaBoost	Logistic Regression	K-Nearest Neighbour	Gaussian Naive Bayes
Precision	0	0.77	0.77	0.78	0.75	0.78	0.73	0.79
	1	0.61	0.61	0.63	0.67	0.62	0.63	0.63
Recall	0	0.69	0.69	0.71	0.72	0.70	0.69	0.70
	1	0.71	0.71	0.72	0.70	0.71	0.68	0.73
F1-Score	0	0.73	0.73	0.74	0.73	0.74	0.71	0.74
	1	0.65	0.65	0.68	0.68	0.66	0.65	0.67
Support	0	213	213	211	198	212	202	213
	1	146	146	148	168	147	157	146
AUC	-	69.78	69.78	71.46	70.95	70.62	68.16	71.51

1) *Precision*: From the precision values it can be measured that how many of the predicted "True" (1) cases were correctly classified. For the "False" (0) class, most models have similar precision, with Gaussian Naive Bayes performing the best at 0.79, followed closely by K-Nearest Neighbor (KNN) and AdaBoost. In terms of classifying the "True" (1) class, AdaBoost achieves the highest precision at 0.67 where we can say that it more accurately identifies true positives compared to the other models.

2) *Recall*: Recall measures how well the model identifies all actual "True" cases. From the table we observe that for the "False" (0) class, recall values vary around 0.69 to 0.72 which shows that the models are consistent in recognizing the "False" cases. For the "True" (1) class, XGBoost, Random Forest, and

SVM exhibit similar recall values which is around 0.71, indicating that they are effective at correctly identifying positive cases.

3) *F1-Score*: Generally, F1-Score provides a balanced view of model performance. We can see that for the "False" (0) class, most models have similar F1-scores, with values ranging from 0.71 to 0.74. This means that the models perform well in identifying negative cases, particularly Gaussian Naive Bayes, AdaBoost, and SVM. For the "True" (1) class, F1-scores are slightly lower, with values ranging from 0.65 to 0.68. This suggests that predicting "True" cases is more challenging for these models. SVM and AdaBoost perform slightly better in this regard, with F1-scores of 0.68.

4) *Support*: Support refers to the number of actual instances in each class. There are more "False" (0) cases (213 instances) than "True" (1) cases (146 instances) in the dataset which may indicate an imbalance in the class distribution.

5) *AUC (Area under the curve)*: AUC represents the model's ability to distinguish between classes. So, higher values indicate better performance. Gaussian Naive Bayes achieves

the highest AUC at 71.51, indicating it is the best at distinguishing between "False" and "True" cases. XGBoost and Random Forest have identical AUC values of 69.78, suggesting similar performance in overall classification accuracy.

The overall summary of Manual ML and Auto ML is depicted in Table XIII.

TABLE XIII. OVERALL SUMMARY OF AUTO ML AND MANUAL ML

Manual ML			AutoML	
Algorithm Name	Accuracy	Prediction Result	List of Algorithms chosen	
AdaBoost	71%	Using AdaBoost, Winner – France 1st Runner Up - Portugal	Best Model chosen via AutoML	Logistic Regression
Random Forest	70%	Using Random Forest, Winner – Portugal 1st Runner Up - Switzerland	Prediction Result	The best Model chosen through AutoML was Logistic Regression. Using Logistic Regression, Winner – Portugal 1st Runner Up - Italy
XG Boost	70%	Using XG Boost, Winner – Portugal 1st Runner Up - France		
SVM	71%	Using SVM, Winner – Italy 1st Runner Up - Netherlands		
Logistic Regression	70%	Using Logistic Regression, Winner – Portugal 1st Runner Up - France		
KNN	68%	Using KNN, Winner – Netherlands 1st Runner Up - Germany		
Naive Bayes	71%	Using Gaussian Naive Bayes, Winner – Portugal 1st Runner Up - England		

VIII. CONCLUSIONS

This study shows how both manual and automated machine learning (AutoML) techniques can effectively predict football match outcomes. By using a comprehensive dataset of historical match data and applying various ML algorithms, we created models that significantly improve the accuracy and reliability of sports predictions. We found that AutoML models, especially logistic regression, offered better predictive accuracy than traditional manual methods. AutoML streamlined the model selection and tuning process, making predictive analysis more efficient and less reliant on manual intervention. AutoML proved it could optimize ML model performance by automating key steps like data pre-processing, feature selection, and hyperparameter tuning.

Manual ML techniques, while effective, required more effort and expertise to match the results achieved by AutoML. Manual methods like Random Forest, XGBoost, SVM, and AdaBoost performed well but were more time-consuming and needed more domain-specific knowledge. Our findings highlight the importance of thorough data preprocessing and feature engineering in boosting model performance. Using

cross-validation techniques and hyperparameter optimization further improved the models' accuracy and robustness, ensuring they are applicable to real-world scenarios.

Additionally, this research provided valuable insights into the factors that influence football match outcomes. This knowledge is invaluable for sports industry stakeholders, including analysts, coaches, and betting agencies, giving them a powerful tool for strategic decision-making.

In summary, this study demonstrated the effectiveness of both manual and AutoML techniques in sports analytics, paving the way for broader adoption and innovation. The results suggest that AutoML can greatly enhance the efficiency and effectiveness of predictive modelling in sports. Future research could incorporate diverse data sources and extend these methods to other sports, showcasing the versatility and scalability of machine learning.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used the QuillBot tool [<https://quillbot.com/grammar-check>] to check grammar as well as paraphrasing. After using this tool/service,

the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

REFERENCES

- [1] "European Championship | History, Winners, & Facts | Britannica," [www.britannica.com](https://www.britannica.com/sports/European-Championship). <https://www.britannica.com/sports/European-Championship>.
- [2] J. Hucaljuk and A. Rakipović, "Predicting football scores using machine learning techniques," 2011 Proceedings of the 34th International Convention MIPRO, Opatija, Croatia, 2011, pp. 1623-1627.
- [3] J. D. Rose, M. K. Vijaykumar, U. Sakthi, and P. Nithya, "Comparison of Football Results Using Machine Learning Algorithms," IEEE Xplore, Jul. 01, 2022. <https://ieeexplore.ieee.org/document/9914265>.
- [4] Groll, Andreas & Ley, Christophe & Schauburger, Gunther & Eetvelde, Hans & Zeileis, Achim. (2019). Hybrid Machine Learning Forecasts for the FIFA Women's World Cup 2019.
- [5] A. Basit, M. B. Alvi, F. H. Jaskani, M. Alvi, K. H. Memon, and R. A. Shah, "ICC T20 Cricket World Cup 2020 Winner Prediction Using Machine Learning Techniques," IEEE 23rd International Multitopic Conference (INMIC), Nov. 2020, doi: <https://doi.org/10.1109/inmic50486.2020.9318077>.
- [6] Tekade P, Markad K, Amage A, Natekar B. Cricket match outcome prediction using machine learning. International journal of Advance Scientific Research and Engineering Trend, 2020 July, 5(7).
- [7] J. Kumar, R. Kumar and P. Kumar, "Outcome Prediction of ODI Cricket Matches using Decision Trees and MLP Networks," 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC), Jalandhar, India, 2018, pp. 343-347, doi: 10.1109/ICSCCC.2018.8703301.
- [8] Daniel Mago Vistro, F. Rasheed, and Leo Gertrude David, "The Cricket Winner Prediction With Application Of Machine Learning And Data Analytics," International Journal of Scientific & Technology Research, vol. 8, no. 9, pp. 985-990, Sep. 2019.
- [9] H. Elmiligi and S. Saad, "Predicting the Outcome of Soccer Matches Using Machine Learning and Statistical Analysis," IEEE Xplore, Jan. 01, 2022. <https://ieeexplore.ieee.org/document/9720896>.
- [10] A. Majumdar, R. Kaur, T. Kulkarni, M. Jiruwala, S. Shah, and N. Pise, "Football Match Prediction using Exploratory Data Analysis & Multi-Output Regression," IEEE Xplore, Dec. 01, 2022. <https://ieeexplore.ieee.org/abstract/document/10119340>.
- [11] A. V. P, R. D, and S. N. S. S, "Football Prediction System using Gaussian Naïve Bayes Algorithm," IEEE Xplore, Mar. 01, 2023. <https://ieeexplore.ieee.org/document/10085510/authors#authors>.
- [12] Jeremiah Samson Chin, Filbert Hilman Juwono, Ing Ming Chew, S. Sivakumar, and W. K. Wong, "Predicting Ice Hockey Results Using Machine Learning Techniques," Jul. 2023, doi: <https://doi.org/10.1109/icdate58146.2023.10248726>.
- [13] Dhananjay Daundkar and Kundan Kandhway, "Predicting Winner of a Professional Basketball Match," Oct. 2023, doi: <https://doi.org/10.23919/iccacs59377.2023.10316903>.
- [14] M. Vashist, V. Bahl, N. Sengar, and A. Goel, "Machine Learning for Football Matches and Tournaments," IEEE Xplore, May 01, 2022. <https://ieeexplore.ieee.org/document/9850673>.
- [15] E. Tiwari, P. Sardar and S. Jain, "Football Match Result Prediction Using Neural Networks and Deep Learning," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2020, pp. 229-231, doi: 10.1109/ICRITO48877.2020.9197811.
- [16] M. Jaeyalakshmi & S. Indrajith & C. Hirthik & K. Kaushiik & S. Eaknath. (2023). Predicting the outcome of future football games using machine learning algorithms. 1-7. 10.1109/RMKMATE59243.2023.10370000.
- [17] Amitesh Peddii and R. Jain, "Random Forest-Based Fantasy Football Team Selection," Mar. 2023, doi: <https://doi.org/10.1109/icaccs57279.2023.10113019>.
- [18] M. A. AL-ASADI and S. Tasdemir, "Predict the Value of Football Players Using FIFA Video Game Data and Machine Learning Techniques," IEEE Access, vol. 10, pp. 1-1, 2022, doi: <https://doi.org/10.1109/access.2022.3154767>.
- [19] A. M. Emam, O. Tarek Ali and A. Atia, "Football activities classification," 2023, Intelligent Methods, Systems, and Applications (IMSA), Giza, Egypt, 2023, pp. 520-525, doi: 10.1109/IMSA58542.2023.10217464.
- [20] F. Rodrigues and Â. Pinto, "Prediction of football match results with Machine Learning," Procedia Computer Science, vol. 204, pp. 463-470, 2022, doi: <https://doi.org/10.1016/j.procs.2022.08.057>.
- [21] M. Gifford and Tuncay Bayrak, "A predictive analytics model for forecasting outcomes in the National Football League games using decision tree and logistic regression," Decision Analytics Journal, pp. 100296-100296, Aug. 2023, doi: <https://doi.org/10.1016/j.dajour.2023.100296>.
- [22] K. Chauhan et al., "Automated Machine Learning: The New Wave of Machine Learning," 2020, 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 205-212, doi: 10.1109/ICIMIA48430.2020.9074859.
- [23] Singh, V. K., & Joshi, K. (2022). Automated Machine Learning (AutoML): an overview of opportunities for application and research, Journal of Information Technology Case and Application Research, 24(2), 75-85. <https://doi.org/10.1080/15228053.2022.2074585>.
- [24] Truong, A., Walters, A., Goodsitt, J., Hines, K., Bruss, C. B., & Farivar, R. (2019, November). Towards automated machine learning: Evaluation and comparison of AutoML approaches and tools. In 2019 IEEE 31st international conference on tools with artificial intelligence (ICTAI), pp. 1471-1479. IEEE.
- [25] L. Ferreira, A. Pilastrri, C. M. Martins, P. M. Pires, and P. Cortez, "A Comparison of AutoML Tools for Machine Learning, Deep Learning and XGBoost," 2021 International Joint Conference on Neural Networks (IJCNN), Jul. 2021, doi: <https://doi.org/10.1109/ijcnn52387.2021.9534091>.
- [26] Elshawi, R., Maher, M., & Sakr, S. (2019). Automated machine learning: State-of-the-art and open challenges. arXiv preprint arXiv:1906.02287.
- [27] Nagarajah, Thiloshon, and Guhanathan Poravi. "A review on automated machine learning (AutoML) systems." In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), pp. 1-6. IEEE, 2019.
- [28] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Implementing AutoML in Educational Data Mining for Prediction Tasks," Applied Sciences, vol. 10, no. 1, p. 90, Dec. 2019, doi: <https://doi.org/10.3390/app10010090>.
- [29] X. Shi, Y. D. Wong, C. Chai, and M. Z.-F. Li, "An Automated Machine Learning (AutoML) Method of Risk Prediction for Decision-Making of Autonomous Vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 11, pp. 7145-7154, Nov. 2021, doi: <https://doi.org/10.1109/tits.2020.3002419>.
- [30] Göksu, Semih & Sezen, Bulent & Balcioglu, Yavuz. (2024). Predicting the Uefa Euro 2024 Winner: An Artificial Neural Network Approach.
- [31] Mahadinour48, "International football matches," Kaggle.com, 2023. <https://www.kaggle.com/datasets/mahadinour/international-football-matches> (Accessed Jul. 29, 2024).
- [32] Simplilearn, "Random Forest Algorithm," Simplilearn.com, Nov. 07, 2023. <https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm>. (Accessed Jul. 29, 2024).
- [33] GeeksforGeeks, "XGBoost," GeeksforGeeks, Sep. 18, 2021. <https://www.geeksforgeeks.org/xgboost/>.
- [34] Prashant11, "AdaBoost Classifier Tutorial," Kaggle.com, Apr. 30, 2020. <https://www.kaggle.com/code/prashant11/adaboost-classifier-tutorial/notebook> (Accessed Jul. 29, 2024).

Detection of DDoS Cyberattack Using a Hybrid Trust-Based Technique for Smart Home Networks

Oghenetejiri Okporokpo, Funminiyi Olajide, Nemitari Ajenka, Xiaoqi Ma

Department of Computer Science, Nottingham Trent University, Clifton Lane, Nottingham NG11 8NS, United Kingdom

Abstract—As Smart Home Internet of Things (SHIoT) continue to evolve, improving connectivity and security whilst offering convenience, ease, and efficiency is crucial. SHIoT networks are vulnerable to several cyberattacks, including Distributed Denial of Service (DDoS) attacks. The ever-changing landscape of Smart Home IoT threats presents many problems for current cybersecurity techniques. In response, we propose a hybrid Trust-based approach for DDoS attack detection and mitigation. Our proposed technique incorporates adaptive mechanisms and trust evaluation models to monitor device behaviour and identify malicious nodes dynamically. By leveraging real-time threat detection and secure routing protocols, the proposed trust-based mechanism ensures uninterrupted communication and minimizes the attack surface. Additionally, energy-efficient techniques are employed to safeguard communication without overburdening resource-constrained SHIoT devices. To evaluate the effectiveness of the proposed technique in efficiently detecting and mitigating DDoS attacks, we conducted several simulation experiments and compared the performance of the approach with other existing DDoS detection mechanisms. The results showed notable improvements in terms of energy efficiency, improved system resilience and enhanced computations. Our solution offers a targeted approach to securing Smart Home IoT environments against evolving cyber threats.

Keywords—Trust; smart home; IoT; DDoS; denial of service; DoS; cyber threats; techniques

I. INTRODUCTION

Over the past few years, the advancement of Internet of Things (IoT) technology has resulted in ease of integration, seamless functionality and increased user satisfaction [1]. Since its inception, we have witnessed an increase in the number of smart home Internet of Things (SHIoT) devices such as smart bulbs, smart TVs, smart alarms, smart refrigerators, and smart fans [2] These have in turn resulted in diverse applications such as smart cities [3], smart grid systems [4], and smart healthcare systems [5].

However, security remains a paramount concern, specifically in smart home network environments, which usually encompass, wireless and mobile ad hoc networks. These environments generally deviate from traditional wired networks, boasting distinctive attributes such as shared resources, node mobility, and limited transmission range [6]. As a result of the generally limited processing power of mobile nodes in smart home networks, security techniques that have proven successful in wired networks tend to fail in wireless networks [7]. Furthermore, because nodes in smart home networks are free to join or leave, their dynamic nature causes network topologies to change quickly, which makes maintaining network security

extremely difficult. The creation of complex yet effective security measures suited to these environments is necessary [8].

In our work, we explore the escalating cybersecurity threats faced by Smart Home Internet of Things (SHIoT) networks, particularly focusing on Distributed Denial of Service (DDoS) attacks. Traditional security measures have proven inadequate in safeguarding these networks, requiring innovative solutions.

The contribution of this research is a proposed novel Trust-based DDoS attack detection model tailored specifically for SHIoT environments. Through comprehensive analysis, the paper identifies prevalent DDoS attack types targeting these networks, delving into their unique characteristics and implications. It evaluates the effectiveness of current cybersecurity measures and introduces a trust-based mitigation technique designed to counter each identified attack vector. By emphasizing the significance of trust-based approaches, the research not only contributes to the enhancement of cybersecurity in smart home settings but also identifies key avenues for future exploration. This study lays the groundwork for more resilient and secure smart home networks, ensuring the confidentiality and integrity of IoT communications amidst the evolving landscape of cyber threats.

The results show that the proposed technique can effectively improve the security of smart homes and enhance the efficiency of smart home network environments. The key contributions of our work are summarized as follows:

- The proposed approach incorporates Knowledge-based trust computations, resulting in more efficient and effective trust aggregations in smart homes.
- Observational-based Trust optimization is used to update trust and reputation, allowing for the system to draw upon the shared encounters of its neighbouring nodes or devices on the network which allows the network parameters to be adjusted as needed.
- The proposed technique deploys a hybrid trust-based technique for trust propagation, trust updation and trust formation which classifies malicious nodes using knowledge, reputation, and observational experience, resulting in better identification and mitigation of security threats in smart homes.

The layout of the paper is as follows. In Section II, we discuss the related work. In Section III, we describe our methodology for trust in smart home network environments. Section IV describes in detail our proposed trust-based system while in Section V. We evaluate the system performance within

smart home networks. In Section VI future research directions are highlighted.

II. RELATED WORK

In recent years, the field of SHIoT security has gained significant attention from researchers, because of the peculiar vulnerabilities of these SHIoT devices [9]-[12]. A smart home is an essential component of intelligent computing, by easily integrating with home devices to control and monitor their operations. It often uses cloud computing for storage and scalable processing power. Smart home appliances can now be remotely controlled from any location thanks to cloud computing [13]. Smart homes improve convenience, security, and energy efficiency by allowing users to effectively manage gadgets. These gadgets offer a great deal of convenience in addition to time, money, and energy savings. The main control interface for the smart home system is usually a smartphone or tablet. In this section, we review the existing literature covering key SHIoT security challenges, the nature of DDoS attacks on SHIoT networks, and existing cybersecurity solutions.

A. Overview of SHIoT Security Challenges

The ubiquitous nature of SHIoT creates some unique weaknesses and challenges which are inherent in their design. Almost any device can be equipped with the necessary technology to facilitate data transmission between IoT devices and their connected networks. Each node in a SHIoT network generally operates under energy constraints, creating an incentive for nodes to selfishly conserve their resources [14]. This self-preserving behaviour can negatively impact the overall functionality and efficiency of the network. Another unique challenge is due to their typical deployment in unattended and often hostile environments meaning that these networks often rely on thousands of low-cost sensors to monitor even small areas, which necessitates producing sensors at minimal cost. This cost reduction compromises the tamper-resistant properties of the SHIoT devices. SHIoTs are typically vulnerable to physical capture by adversaries [15]. Ensuring secure and efficient operation is challenging due to these factors particularly when threats like Distributed Denial of Service (DDoS) attacks target these SHIoT networks. One of the main concerns in smart homes is unauthorized access, where sensitive user data, such as video feeds or personal preferences can be intercepted if devices do not have proper access control protocols in place [16]. Due to the computational limitations of SHIoT devices, there are limits on the implementation of advanced cryptographic algorithms, thereby leading to exposure to various types of cyberattacks.

Existing security models oftentimes focus on traditional IT systems, overlooking IoT's resource limitations and real-time processing needs [8]. The lack of standardized security practices across IoT device manufacturers exacerbates these issues, leaving devices vulnerable to exploitation and making it challenging to implement uniform security measures across diverse IoT ecosystems.

B. DDoS Attacks

DDoS attacks have become increasingly common in SHIoT networks, largely due to the massive deployment of SHIoT devices, which can be easily exploited due to weak security

configurations [17]. Common DDoS attacks within the SHIoT environment include HTTP floods, UDP floods, and TCP SYN floods.

1) *HTTP Flood attacks*: HTTP flood attacks are one of the most common DDoS cyberattacks. These attacks are carried out by inundating the victim with a massive number of HTTP connection requests. These attacks aim to overwhelm the target server's resources and prevent legitimate traffic from accessing the server. In the context of IoT, HTTP floods can target cloud-based services associated with smart home devices, causing network slowdowns and disruptions [18]. Researchers Marleau et al. proposed an HTTP flood detection and mitigation technique for Software-defined networks (SDN) using Network Ingress Filtering [19].

2) *UDP Flood attacks*: UDP flood attacks flood the victim network or device with many User Datagram Protocol (UDP) packets. The extensive volume of packets inundates the target server, aiming to overwhelm its processing and response capabilities. UDP floods are particularly disruptive in SHIoT environments, where devices rely on minimal bandwidth and have limited packet-processing capabilities [20]. An example is the DNS amplification attack, where the attacker spoofs the source IP address of the victim and sends a small request to the DNS server. The DNS server replies with large responses, affecting the victim's performance. Researchers Lee et al. [21] proposed the use of specific IPtables rules and Linux-based firewall utilities, to mitigate UDP flood attacks.

3) *TCP SYN Flood attacks*: This type of attack exploits the TCP handshake mechanism by sending repeated SYN requests, but failing to respond to SYN-ACK replies, leaving the connection half-open. This can consume server resources and result in denial of service. Smart home devices, which often operate on simple network architectures, are vulnerable to these types of connection-based floods [22]. Bensaid et al. proposed a fog computing-based SYN Flood DDoS attack mitigation technique which uses an adaptive neuro-fuzzy inference system (ANFIS) and SDN assistance [23].

The impact of these attacks on SHIoT networks is significant, leading to degraded performance, reduced availability, and even complete network outages. DDoS attacks also open pathways for further malicious activities, such as data breaches or malware infiltration, by exploiting compromised devices within the IoT network [24].

C. Existing DDoS Mitigation Solutions

Current DDoS mitigation techniques include solutions like rate limiting, firewalls, and anomaly detection. However, while these methods offer some level of protection, they are often insufficient or computationally demanding for IoT environments:

1) *Rate limiting*: This approach restricts the number of requests allowed per unit of time, which can mitigate DDoS attacks. However, IoT devices may still be overwhelmed by legitimate traffic, and rate limiting does not effectively distinguish between malicious and legitimate requests [25].

2) *Firewalls*: Traditional firewalls monitor all incoming traffic attempting to enter a network and can block unwanted traffic. However, they are often unsuitable for IoT devices due to their processing and memory limitations. Additionally, firewalls require frequent updates to stay effective, which may not be feasible for resource-constrained SHIoT devices [26].

3) *Anomaly detection*: Anomaly detection, also known as behavioural detection, involves identifying predefined signatures or events that deviate from normal system behaviour. These systems use methods such as machine learning and analysis to identify abnormal patterns of network traffic [27]. While effective, these systems are computationally intensive, requiring processing power that most SHIoT devices lack. Moreover, the high rate of false positives in anomaly detection can lead to unnecessary slow-down in network performance, impacting the reliability of IoT services [28].

These traditional solutions, while useful in general networking environments, fall short of providing scalable, efficient, and reliable security for SHIoT networks, particularly when faced with DDoS attacks in smart home environments.

D. Trust-Based Security Approaches

As a result of the limitations of other DDoS mitigation techniques, researchers have explored trust-based security models tailored to various technologies. Trust-based security mechanisms aim to establish a level of trust for each device or network node based on behaviour, interaction history, and reputation, allowing the network to isolate untrustworthy devices or nodes in real time.

Several studies have highlighted the benefits of trust-based approaches in distributed and resource-constrained environments like IoT [29]-[31]. Trust-based models can effectively mitigate insider threats by flagging devices that exhibit suspicious behaviour, such as attempting excessive communication or participating in botnet-like activities [21]. Trust-based systems are also adaptable, requiring less processing power than anomaly detection making them suitable for IoT devices with limited computational capacity [32].

Shuhaiber and Mashal [33] presented a multilayered trust-based technique within IoT ecosystems, offering a theoretical insight into the intricate relationships between Trust Stance, and their impact on trust dynamics within IoT networks. Khatereh et al. [34] introduced a trust management model for anomaly detection using sequence prediction and deep learning for data security in IoT networks. The proposed model provides a detection mechanism to address four RPL attacks.

Shashank et al. [35] apply a trust-based technique for reliable data packet routing in WSNs. In their approach, trust management is integrated into routing protocols, deploying the decision-making Dempster-Shafer Theory (DST) algorithm for trusted clustering and the Whale Optimization Algorithm (WOA) for routing. However, one drawback of this approach is the high energy use which is not suitable for SHIoT networks.

Adla and Ramaiah [36] propose a blockchain solution for IoT with trust management consensus. The proposed technique uses a Grey Wolf Optimization (GWO) algorithm in addition to a trust-based ensemble consensus. The trust-based ensemble consensus uses Proof of Work (PoW) and Proof of Stake (PoS) procedures to calculate trust within the network. However, one disadvantage of this approach is that the network throughput progressively drops as the number of nodes increases. Researcher Farag [37] proposed a behavioural trust-based solution to mitigate energy exhaustion attacks on the RPL protocol. The proposed protocol protects against rank attacks and Sybil attacks in IoT networks. However, the disadvantage of the technique is that the trust value is computed solely based on direct observations by each node within the network.

Researchers have proposed various methods to deal with the DDoS attacks common with IoT networks. The approaches deployed vary and authors have focused on different aspects of the security of IoT networks. It is also evident from our study on trust-based techniques and deployments that a comprehensive model incorporating all aspects of security quantification for smart home networks and services is imperative. Thus, the core focus of this research work is a proposed trust-based system as a means of securing SHIoT networks. Trust-based management techniques employ a systematic method for effectively managing and ensuring trust within the network. By incorporating trust as a core component, our model provides an adaptive, lightweight solution that enhances the security of SHIoT networks.

III. METHODOLOGY

In this section, we present the trust-based methodology that the proposed system uses to detect and mitigate Distributed Denial of Service (DDoS) attacks in SHIoT networks. The methodology is centred around a trust management system where each node in the network maintains a trust score for other nodes based on their behaviour [32]. The trust scores are dynamically updated as nodes interact with each other. When malicious behaviour is detected, such as in the case of a DDoS attack, trust scores decline, and the system can identify the compromised node and take necessary actions to mitigate the attack.

A. Network Architecture

Trust-based systems are primarily comprised of three distinct properties. Durable nodes/devices that cumulate a repository of protocols for future communication, compilation, and dissemination of information regarding ongoing communications and ensuring its availability for future reference and deployment of a propagation mechanism to aid the dissemination of trust information to peer nodes/devices on the network. Fig. 1 shows a high-level overview of the proposed Trust-based system.

Assessing the security of a smart home network is essential for any setup within the smart home environment. We've compiled a comprehensive set of security parameters crucial for gauging security within a smart home network environment.

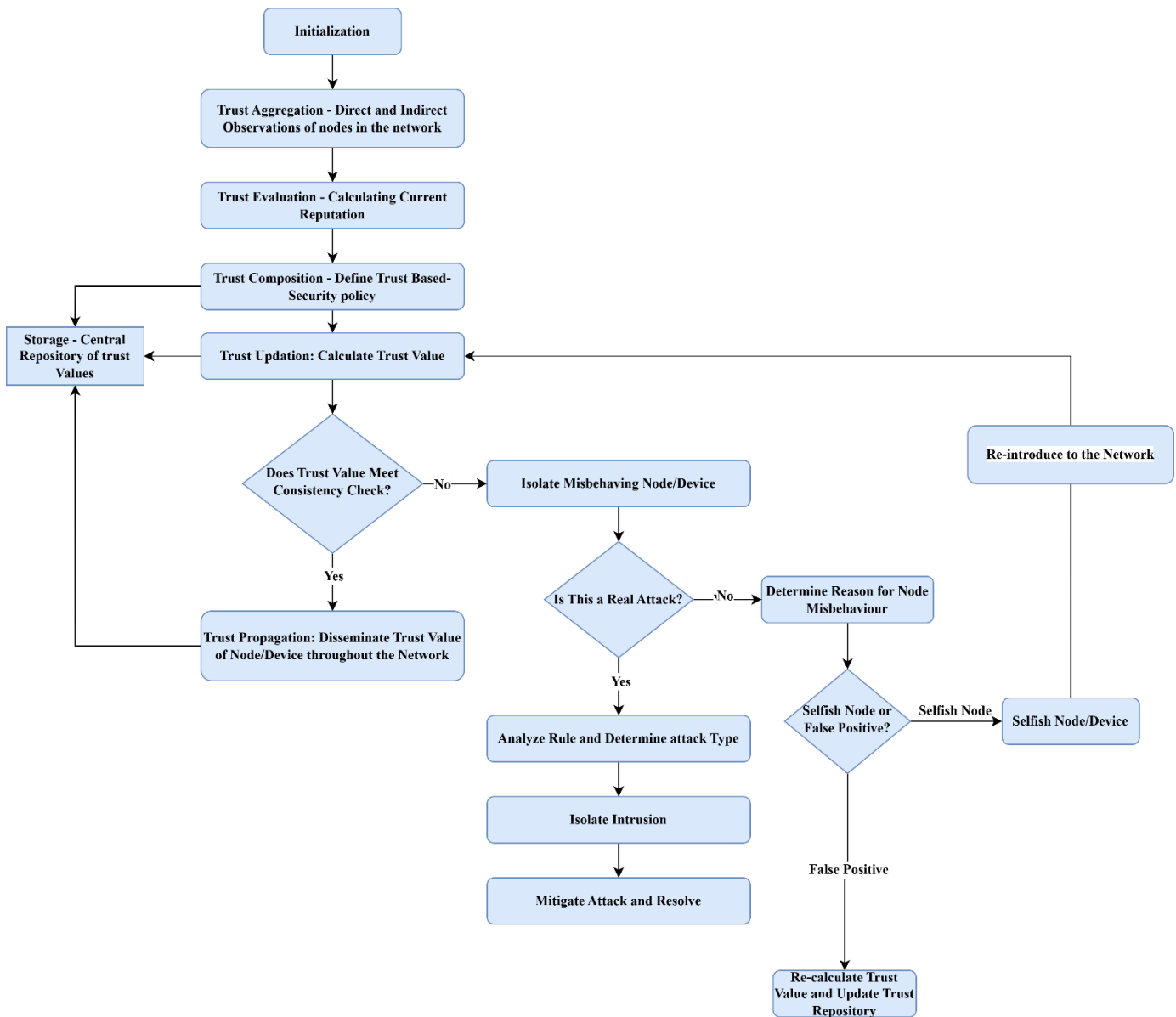


Fig. 1. High-level overview of proposed trust-based system.

These parameters form the basis of our trust model, yielding a trust value as an outcome. This trust value can either provide a holistic view of the overall security of the smart home network or can be dissected into various security aspects based on these parameters, represented as a vector.

B. Trust Definition

Trust is a measure of the reliability or reputation of a node in the network, quantifying how likely it is that a node will behave as expected, such as reliably forwarding packets without malicious intent. Trust in this context is represented as a numerical value, which is continuously evaluated and updated based on observed behaviour [30]. In the proposed system, trust is defined based on three main factors.

Packet Delivery Ratio (PDR): This measures the consistency and reliability of node y . The ratio of successfully delivered packets to the total number of packets transmitted. A high PDR

implies that the node can be trusted to forward packets efficiently, whereas a low PDR may indicate the dropping or mishandling of packets, which is indicative of malicious behaviour or a selfish node trying to conserve resources.

$$PDR_{x,y}(t) = \frac{\text{Packets Delivered by node } y}{\text{Total Packets Sent to node } y} \quad (1)$$

Anomaly Detection (AD): Anomalies such as sudden traffic spikes are common indicators of a node participating in a DDoS attack. The system continuously monitors the traffic patterns, and if it detects abnormal behaviour, the anomaly detection score decreases the trust score.

$$AD_{x,y}(t) = \begin{cases} 1 & \text{if no anomaly is detected,} \\ 0 & \text{if an anomaly is detected.} \end{cases} \quad (2)$$

Response Time (RT): The time a node takes to respond to communication requests from other nodes. If the response times are consistently high (i.e., the node is unresponsive or

overloaded), this could indicate that the node is under attack or has been compromised.

$$RT_{x,y}(t) = \frac{1}{\text{Observed Response Time of node } y} \quad (3)$$

C. Trust Calculation

Trust is a quantified measure based on the behaviour of nodes. Trust is evaluated based on parameters such as packet delivery ratio, response time, and anomaly detection. The trust $T_{(x,y)}(t)$ between two nodes x and y at time t is calculated as a weighted sum of the three factors mentioned above: Packet Delivery Ratio, Anomaly Detection, and Response Time. The trust calculation formula is:

$$T_{x,y}(t) = \delta \times PDR_{x,y}(t) + \theta \times AD_{x,y}(t) + \mu \times RT_{x,y}(t) \quad (4)$$

Where:

$T_{x,y}(t)$ is the trust value between nodes x and y at time t .

$PDR_{x,y}(t)$ is the packet delivery ratio between node x and node y . That is the ratio of successful packet deliveries.

$AD_{x,y}(t)$ is the anomaly detection score, indicating if node y 's behaviour is deemed suspicious.

$RT_{x,y}(t)$ is the response time of node y as observed by node x . That is the delay in responses from node y .

δ , θ , μ are the weights for each factor, with $\delta + \theta + \mu = 1$, determined based on the specific requirements of the network. For example, if packet delivery is prioritized, δ would be larger.

D. Trust Update Mechanism

Trust is not static and changes as nodes interact over time. The system continuously monitors the behaviour of each node, and the trust scores are updated periodically based on recent observations. This dynamic nature ensures that the system can adapt to evolving network conditions and malicious behaviours. The trust update mechanism works as follows:

- **Initial Trust Assignment:** Every node starts with an initial trust value. For example, the default trust value is set to 0.5 on a scale of 0 to 1, indicating neutral trust.

$$T_{x,y}(t) = 0.5 \quad (5)$$

- **Trust Evaluation:** After each interaction between two nodes, the trust score is recalculated based on the Packet Delivery Ratio, Anomaly Detection, and Response Time.
- **Trust Decay:** Trust decays over time if no recent interaction has occurred. This decay ensures that old interactions do not overly influence current trust evaluations.

$$T_{x,y}(t+1) = (1-\omega) \times T_{x,y}(t) + \omega \times \text{new interaction data} \quad (6)$$

Where ω is a decay constant that controls how quickly trust values diminish over time.

- **Threshold-Based Detection:** The system sets a threshold T_{thresh} below which a node is flagged as suspicious. If the trust value $T_{(x,y)}(t)$ falls below this threshold, the node is quarantined, i.e., its communication privileges are limited or monitored closely. The value of T_{thresh} is set

based on network performance requirements and the acceptable level of risk.

$$\text{If } T_{x,y} < T_{\text{thresh}} \text{ then node } y \text{ is flagged as suspicious.} \quad (7)$$

The value of T_{thresh} is set based on network performance requirements and the acceptable level of risk.

E. Trust Propagation in the Network

The trust scores are not only calculated on a one-to-one basis between nodes but also propagated through the network. For instance, if node x considers node y to be trustworthy, other nodes that trust x may adjust their trust values for y accordingly. This indirect trust propagation allows for faster identification of malicious nodes but also introduces a potential risk of trust manipulation. The propagation mechanism follows a weighted averaging approach.

$$T_{k,y}(t) = \frac{T_{k,x}(t) + T_{x,y}(t)}{2} \quad (8)$$

Where node k updates its trust score for node y based on its trust in node x and the trust score that node x has assigned to node y .

F. Trust-Based DDoS Attack Detection

The proposed trust-based system is used to detect DDoS attacks by identifying nodes whose trust scores consistently fall below the set threshold due to anomalies in their behaviour. DDoS attacks typically involve a sudden surge of requests from compromised nodes, resulting in dropped packets, increased response times, and detected anomalies, all of which contribute to a rapid decline in the trust score. The detection algorithm works as follows:

- 1) **Monitor trust values:** Continuously monitor the trust values for all nodes in the network.
- 2) **Detect malicious nodes:** If a node's trust score falls below the threshold T_{thresh} , flag the node as suspicious.
- 3) **Isolate suspicious nodes:** Once flagged, restrict the node's ability to communicate with other nodes until further investigation is carried out or the node is cleared.

IV. TRUST-BASED DDoS DETECTION SYSTEM

In this section, we delve deeper into the workings of our proposed trust-based system for detecting Distributed Denial of Service (DDoS) attacks in smart home networks. The system uses trust scores to detect anomalies in node behaviour that could indicate a malicious DDoS attack. By dynamically assessing the trustworthiness of each node, our system can identify compromised nodes that are part of a DDoS attack and take action to mitigate the attack in real time.

A. Trust Propagation and Decision Making

The trust-based DDoS detection system operates by continuously monitoring and updating trust values between nodes. Each IoT device in the network maintains a trust score for other devices it communicates with. Trust propagation ensures that trust information is shared across the network, allowing for more comprehensive decision-making.

- 1) **Trust evaluation:** Trust values are evaluated based on the behaviour of the nodes, as discussed in Section III(B). Nodes

regularly assess their neighbours based on metrics such as Packet Delivery Ratio (PDR), Response Time (RT), and Anomaly Detection (AD). A node that behaves consistently within normal parameters maintains a high trust score. Conversely, a node that shows erratic or malicious behaviour, such as failing to forward packets or exhibiting a high rate of traffic anomalies, will experience a drop in trust.

2) *Trust propagation and aggregation*: Trust propagation allows nodes to share their trust evaluations of other nodes, leading to a more informed decision-making process. If node x trusts node y but receives reports from other nodes indicating low trust in y , node x can adjust its trust value for y accordingly. This aggregation of trust values helps quickly isolate malicious nodes. Trust propagation is defined mathematically as follows:

$$T_{x,y}(t+1) = \frac{T_{x,y}(t) + \sum_{m \in N(x)} T_{m,y}(t)}{N(x)+1} \quad (9)$$

Where:

$T_{(x,y)}(t+1)$ is the trust value between nodes x and y at time t .

$N(x)$ is the set of neighbouring nodes if x ,

$T_{(m,y)}(t)$ is the trust value that node m assigns to node y .

This formula ensures that a node's trust score reflects not only its direct interactions but also the observations of other nodes in the network. This collective trust evaluation reduces the likelihood of isolated nodes manipulating their trust values to avoid detection.

B. DDoS Detection Algorithm

The detection of DDoS attacks in the trust-based system relies on identifying nodes with consistently low trust scores. These low scores indicate misbehaviour such as failing to forward packets, delaying responses, or generating abnormally high traffic. The following algorithm outlines the detection process:

1) *Step 1: Initialize Trust Values*: Each node x in the network initializes a trust score $T(x,y)(0)$ for every other node y . The initial trust value is set to a neutral level, such as 0.5.

2) *Step 2: Continuous Monitoring*: Nodes continuously monitor the behaviour of their neighbours based on the metrics discussed in Section III(B) (Packet Delivery Ratio, Response Time, and Anomaly Detection).

3) *Step 3: Trust Score Update*: Each node x updates its trust score for every other node y after each interaction. The updated trust score $T(x,y)(t)$ is calculated using the formula described in Section III(B).

4) *Step 4: Threshold Comparison*: At regular intervals, each node compares the trust score of its neighbours to a predefined threshold T_{thresh} . If the trust score $T_{(x,y)}(t)$ falls below T_{thresh} node y is flagged as suspicious.

If $T_{x,y}(t) < T_{\text{thresh}}$ then node j is flagged as suspicious. (10)

5) *Step 5: Quarantine Suspicious Nodes*: Once a node is flagged as suspicious, the system takes preventive action. The suspicious node is quarantined, meaning its communication

with other nodes is limited, and it is closely monitored. This limits the node's ability to participate in DDoS attacks. The algorithm can be represented as pseudo code as follows:

Algorithm 1: Quarantine Algorithm

```
for each node x in network:
  for each neighbour y of x:
    T[x,y] = CalculateTrust(x,y)
    if T[x,y] < T_thresh:
      FlagNode(y)
      QuarantineNode(y)
    End
```

C. Detection of Specific DDoS Attack Types

The proposed trust-based system can detect various types of DDoS attacks based on the specific behaviours they induce in the network. Below are three common types of DDoS attacks and how they are detected:

1) *TCP SYN Flood detection*: In a TCP SYN flood attack, a malicious node sends repeated SYN requests to overwhelm the victim node's resources. This attack results in.

a) Increased response times (since the victim node is overwhelmed).

b) Decreased Packet Delivery Ratio (as the victim node struggles to handle legitimate traffic).

The trust score of a node participating in a TCP SYN flood attack will drop due to poor Response Time (RT) and Packet Delivery Ratio (PDR). The system detects this as follows:

- **Response Time Monitoring**: If node x observes a consistent delay in receiving responses from node y , it will reduce the trust score $T_{(x,y)}(t)$ accordingly.

$$T_{x,y}(t) = T_{x,y}(t-1) - \Delta RT_{x,y}(t) \quad (11)$$

- **Packet Delivery Monitoring**: If node j is unable to deliver packets reliably, $PDR_{(x,y)}(t)$ will decrease, leading to a further reduction in trust.

$$T_{x,y}(t) = T_{x,y}(t-1) - \Delta PDR_{x,y}(t) \quad (12)$$

2) *HTTP Flood detection*: In an HTTP flood attack, a compromised node generates a high volume of HTTP requests to overload the victim's web services. This leads to:

- Abnormally high traffic generation.
- Anomalies detected in traffic patterns (AD).

The proposed trust-based system detects HTTP flood attacks by monitoring traffic volumes and identifying anomalies in the behaviour of nodes. Nodes that generate an unusually high number of HTTP requests will be flagged based on their Anomaly Detection (AD) score:

$$AD_{x,y}(t) = \begin{cases} 1 & \text{if no anomaly is detected,} \\ 0 & \text{if an anomaly is detected.} \end{cases} \quad (13)$$

A lower AD score leads to a drop in the overall trust value $T_{(x,y)}(t)$, eventually flagging the node as suspicious.

3) *UDP Flood detection*: A UDP flood attack involves sending large volumes of UDP packets to flood the victim's bandwidth. This results in: This leads to:

- High packet loss.
- Poor packet delivery ratio (PDR).

In this case, the system detects the attack by monitoring the Packet Delivery Ratio (PDR) of affected nodes. If node x observes that node y is consistently dropping packets, the trust score for node y is reduced:

$$PDR_{x,y}(t) = \frac{\text{Packets Delivered by node } y}{\text{Total Packets Sent to node } y} \quad (14)$$

A low PDR leads to a decline in trust:

$$T_{x,y}(t) = T_{x,y}(t-1) - \Delta PDR_{x,y}(t) \quad (15)$$

D. Mitigation Strategy

The detection of DDoS attacks in the trust-based system relies on identifying nodes with consistently low trust scores. These low scores indicate misbehaviour such as failing to forward packets, delaying responses, or generating abnormally high traffic. The following algorithm outlines the detection process:

1) *Node quarantine*: The system temporarily restricts the suspicious node's ability to communicate with other nodes in the network. This reduces the likelihood of the node participating in a DDoS attack. During quarantine, the system continues to monitor the node's behaviour.

2) *Traffic filtering*: Suspicious traffic from flagged nodes is filtered to prevent it from overwhelming legitimate network resources. The system prioritizes traffic from trusted nodes, ensuring that the network remains functional even during an ongoing attack.

3) *Reassessment of trust*: After a predefined period, the system re-evaluates the trust score of quarantined nodes. If the node's behaviour improves (e.g., it no longer generates anomalies or has improved packet delivery), the node can be re-integrated into the network. Otherwise, it remains quarantined or is permanently blacklisted.

V. RESULT AND ANALYSIS

To evaluate the performance of the proposed model, a simulation was carried out using OMNET++ simulator which was selected due to its platform independence and pre-defined function. To implement the trust-based detection system, we extend the IoT device modules with trust evaluation functionality. Each device calculates the trust score of its neighbours based on their behaviour (packet delivery, response time, and anomaly detection). We measured the following performance metrics:

- **Malicious Attack Detection Rate**: The percentage of malicious nodes correctly identified by the system.
- **False Positive Rate**: The percentage of benign nodes incorrectly flagged as malicious.

- **Latency**: The average time taken to detect and mitigate a DDoS attack.
- **Network Throughput**: The total amount of data successfully transmitted across the network, indicating the impact of DDoS attacks on network performance.

The complete simulation setup is illustrated in Table I.

TABLE I. COMMON SIMULATION PARAMETERS

Simulation environment	Values
Simulator	OMNET++ v 6.0.2
Platform	Windows 11
Number of Nodes	10-50
Time Interval	100-1000s
Topology	800m X 600m
Communication Range	50m
Default Trust Value	0.5
Trust Threshold Value	Data Link
Malicious Penalty	0.2
Decay Rate	0.99
Legitimate Reward	0.1

A. Malicious Attack Detection Rate

The percentage of malicious nodes correctly identified by the trust-based DDoS detection system is evaluated against TCP, UDP and HTTP flood attacks. The simulation results are captured and analysed based on the performance metrics. Table II show is a summary of the results:

TABLE II. DETECTION RATE

Attack Type	Detection Rate (%)
TCP SYN Flood	98
UDP Flood	95
HTTP Flood	92

Fig. 2 illustrates the comparison of our system with existing approaches and demonstrates that a higher detection rate is obtained by the system. The distributed denial-of-service detection mechanism (DiDDeM) system showed a 92% detection rate for TCP SYN flood attacks, 91% for UDP flood attacks and 88% for HTTP flood attacks. Whilst the Adaptive threshold algorithm (ATA) has a detection rate of 93.85%, 92% and 89% respectively for all three attack types. Our trust-based detection system successfully detects most DDoS attacks, with a high detection rate across all attack types tested.

B. False Positive Rate

This is a measure of the percentage of benign nodes incorrectly flagged as malicious. Fig. 3 shows the results of the simulation of our system in comparison to other known systems including the Hybrid Deep Learning CNN-GRU model and the Adaptive threshold algorithm (ATA).

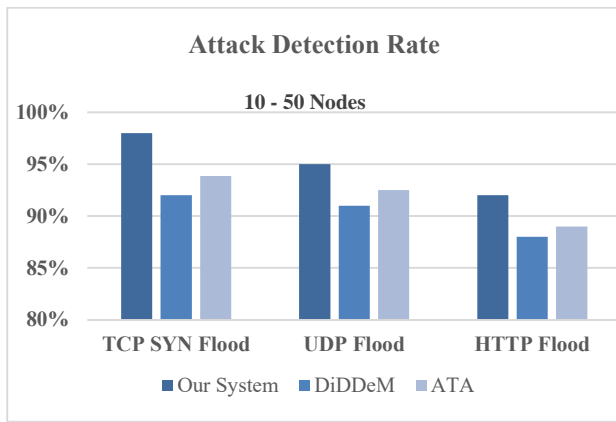


Fig. 2. Malicious attack detection rate.

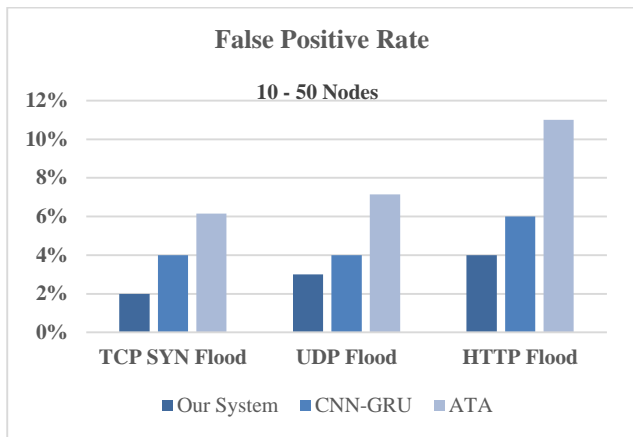


Fig. 3. False positive rate.

The system maintains a low false positive rate (Table III), ensuring that most benign nodes are not incorrectly flagged as malicious.

TABLE III. FALSE POSITIVE RATE

Attack Type	False Positive Rate (%)
TCP SYN Flood	2
UDP Flood	3
HTTP Flood	4

C. Latency

This refers to the delay introduced by the trust calculation and decision-making process. The trust-based system detects attacks with minimal latency (20–22ms), balancing trust evaluation overhead with efficient traffic forwarding and allowing for real-time mitigation. The simulation results are captured and analysed based on the performance metrics. Table IV show is a summary of the results:

TABLE IV. LATENCY

Attack Type	Detection Latency (ms)
TCP SYN Flood	20
UDP Flood	21
HTTP Flood	22

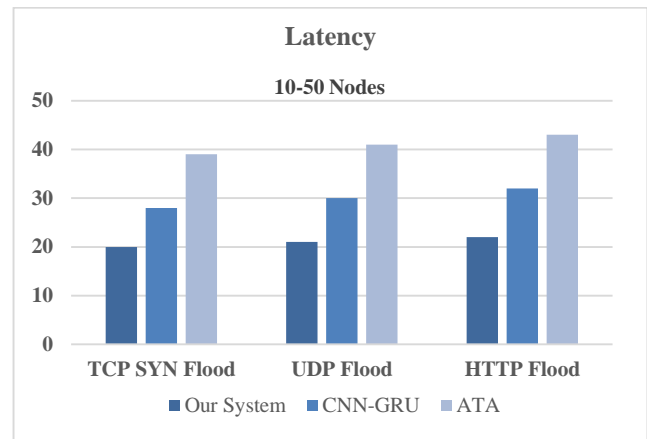


Fig. 4. Latency.

Fig. 4 illustrates the comparison of our system with other existing approaches. The Hybrid Deep Learning CNN-GRU model has the highest latency (28–32ms), due to computationally intensive traffic analysis and the Adaptive threshold algorithm (ATA) has a moderate latency (39–43ms) due to the additional analysis performed beyond threshold enforcement.

D. Network Throughput

This is a measure of the percentage of legitimate traffic successfully forwarded after isolating malicious nodes. Benign nodes that are incorrectly flagged as malicious. Table V. shows the results of the simulation of the system.

TABLE V. NETWORK THROUGHPUT

Attack Type	Throughput (Mbps) Before Attack	Throughput (Mbps) During Attack	Throughput (Mbps) After Detection
TCP SYN Flood	100	50	90
UDP Flood	100	40	85
HTTP Flood	100	45	88

The network throughput drops significantly during an attack but recovers after the trust-based system detects and mitigates the attack. The system maintained a high throughput even after detection (TCP - 95%, UDP - 85% and HTTP - 88%) by isolating only malicious nodes, ensuring minimal disruption to legitimate traffic. There were no instances of occasionally dropping legitimate traffic due to misclassification.

VI. CONCLUSION AND FUTURE WORK

Our proposed trust-based detection mechanism for Distributed Denial of Service (DDoS) attacks in SHIoT networks demonstrates significant potential to improve the security and resilience of SHIoT environments. By utilizing trust scores, the system efficiently identifies and isolates malicious nodes while ensuring minimal impact on legitimate traffic. This research highlights the critical need for adaptive, lightweight, and scalable security solutions tailored to resource-constrained IoT environments. The integration of trust-based mechanisms tailored to SHIoT environments enables the mechanism to detect and mitigate multiple types of DDoS attacks, including TCP SYN Flood, HTTP Flood, and UDP Flood. Our technique

prioritizes lightweight computation to accommodate the limited processing and energy capacities of SHIoT devices. We achieve high detection accuracy, correctly identifying malicious nodes within a short time frame while maintaining a low false positive rate. This ensures the reliability of the network and protects against unnecessary isolation of legitimate nodes. The use of trust decay, penalties for malicious behaviour, and rewards for legitimate traffic ensures that trust scores dynamically reflect the behaviour of each node. This adaptability makes our technique robust against evolving attack patterns and intermittent malicious activities.

This research addresses a critical gap in SHIoT security by providing a lightweight yet effective solution for DDoS detection. As SHIoT adoption continues to grow, securing these networks would continue to be imperative in a bid to prevent disruptions, enhance the resilience of smart home networks, and ensure the integrity, privacy, and availability of SHIoT communications. Our proposed trust-based detection technique not only lays a strong foundation for SHIoT security but also opens avenues for further innovation. The findings of this study reinforce the importance of trust-based approaches in combating cyber threats in IoT networks and paves the way for the development of more secure and reliable IoT systems, ensuring a safer and better-connected future. Future research could explore the design of a scalable, trust-based, easily adaptable cloud/edge computing infrastructure as a service solution for SHIoT networks.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of the Nottingham Trent University (NTU) for a fully funded studentship. The authors also declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

- [1] T. Magara and Y. Zhou, "Internet of Things (IoT) of Smart Homes: Privacy and Security," *Journal of Electrical and Computer Engineering*, vol. 2024, (1), pp. 7716956, 2024.
- [2] A. M. Ansari, M. Nazir and K. Mustafa, "Smart Homes App Vulnerabilities, Threats, and Solutions: A Systematic Literature Review," *Journal of Network and Systems Management*, vol. 32, (2), pp. 29, 2024.
- [3] M. K. Wyrwicka, E. Więcek-Janka and Ł. Brzeziński, "Transition to sustainable energy system for Smart Cities—Literature Review," *Energies*, vol. 16, (21), pp. 7224, 2023.
- [4] M. Khalid, "Smart grids and renewable energy systems: Perspectives and grid integration challenges," *Energy Strategy Reviews*, vol. 51, pp. 101299, 2024.
- [5] A. H. Mohammed and R. M. A. Hussein, "A security services for internet of thing smart health care solutions based blockchain technology," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 20, (4), pp. 772-779, 2022.
- [6] A. Y. Dawod, M. F. Abdulqader and Q. M. Zainel, "Enhancing Security and Sensors Emerging Internet of Things (IoT) Technology of Homophone-Based Encryption using MANET-IoT Networks Technique," *Journal of Electrical Systems*, vol. 20, (6s), pp. 1345-1351, 2024.
- [7] K. Murat et al, "Security Analysis of Low-Budget IoT Smart Home Appliances Embedded Software and Connectivity," *Electronics*, vol. 13, (12), pp. 2371, 2024.
- [8] N. Solangi et al, "IoT based home automation system: Security challenges and solutions," in *2024 5th International Conference on Advancements in Computational Sciences (ICACS)*, 2024.
- [9] I. Cvitić et al, "An overview of smart home iot trends and related cybersecurity challenges," *Mobile Networks and Applications*, vol. 28, (4), pp. 1334-1348, 2023.
- [10] A. Aldahmani et al, "Cyber-security of embedded IoTs in smart homes: challenges, requirements, countermeasures, and trends," *IEEE Open Journal of Vehicular Technology*, vol. 4, pp. 281-292, 2023.
- [11] A. M. Al-Ghaili et al, "A review on role of image processing techniques to enhancing security of IoT applications," *IEEE Access*, vol. 11, pp. 101924-101948, 2023.
- [12] D. Singla et al, "Blockchain-powered healthcare: Revolutionizing security and privacy in IoT-based systems," in *2024 International Conference on Computational Intelligence and Computing Applications (ICCICA)*, 2024.
- [13] H. Yang, Y. Guo and Y. Guo, "Blockchain-based cloud-fog collaborative smart home authentication scheme," *Computer Networks*, vol. 242, pp. 110240, 2024.
- [14] Z. Zheng and H. Nazif, "An energy-aware technique for resource allocation in mobile internet of thing (miot) using selfish node ranking and an optimization algorithm," *IETE Journal of Research*, vol. 70, (4), pp. 3546-3571, 2024.
- [15] A. Allen et al, "Smart homes under siege: Assessing the robustness of physical security against wireless network attacks," *Comput. Secur.*, vol. 139, pp. 103687, 2024.
- [16] M. R. Ahmed and M. O. Rahman, "An Enhanced Secure User Authentication and Authorized Scheme for Smart Home Management," *International Journal of Advanced Computer Science & Applications*, vol. 15, (6), 2024.
- [17] P. Shukla, C. R. Krishna and N. V. Patil, "Iot traffic-based DDoS attacks detection mechanisms: A comprehensive review," *The Journal of Supercomputing*, vol. 80, (7), pp. 9986-10043, 2024.
- [18] D. S. Gonçalves, R. S. Couto and M. G. Rubinstein, "A protection system against HTTP flood attacks using software defined networking," *Journal of Network and Systems Management*, vol. 31, (1), pp. 16, 2023.
- [19] S. Marleau, P. Rahman and C. Lung, "DDoS flood detection and mitigation using SDN and network ingress filtering-an experiment report," in *2024 IEEE 4th International Conference on Electronic Communications, Internet of Things and Big Data (ICEIB)*, 2024.
- [20] O. M. Almorabea et al, "IoT Network-Based Intrusion Detection Framework: A Solution to Process Ping Floods Originating From Embedded Devices," *IEEE Access*, vol. 11, pp. 119118-119145, 2023.
- [21] J. Lee et al, "Rescuing QUIC flows from countermeasures against UDP flooding attacks," in *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, 2024.
- [22] S. Evmorfos et al, "Neural network architectures for the detection of SYN flood attacks in IoT systems," in *Proceedings of the 13th ACM International Conference on Pervasive Technologies Related to Assistive Environments*, 2020.
- [23] R. Bensaid et al, "Toward a Real - Time TCP SYN Flood DDoS Mitigation Using Adaptive Neuro - Fuzzy Classifier and SDN Assistance in Fog Computing," *Security and Communication Networks*, vol. 2024, (1), pp. 6651584, 2024.
- [24] M. Azroul et al, "Internet of things security: challenges and key issues," *Security and Communication Networks*, vol. 2021, (1), pp. 5533843, 2021.
- [25] S. Karmani, N. Agrawal and R. Kumar, "A comprehensive survey on low-rate and high-rate DDoS defense approaches in SDN: taxonomy, research challenges, and opportunities," *Multimedia Tools Appl.*, vol. 83, (12), pp. 35253-35306, 2024.
- [26] M. Patel et al, "DDoS Attack Detection Model using Machine Learning Algorithm in Next Generation Firewall," *Procedia Computer Science*, vol. 233, pp. 175-183, 2024.
- [27] R. N. Bashir et al, "Smart reference evapotranspiration using Internet of Things and hybrid ensemble machine learning approach," *Internet of Things*, vol. 24, pp. 100962, 2023.
- [28] H. I. Mhaibes, M. H. Abood and A. K. Farhan, "Simple Lightweight Cryptographic Algorithm to Secure Imbedded IoT Devices," *International Journal of Interactive Mobile Technologies*, vol. 16, (20), 2022.

- [29] O. Okporokpo et al, "Trust-based Approaches Towards Enhancing IoT Security: A Systematic Literature Review," arXiv Preprint arXiv:2311.11705, 2023.
- [30] S. M. Muzammal, R. K. Murugesan and N. Z. Jhanjhi, "A comprehensive review on secure routing in internet of things: Mitigation methods and trust-based approaches," IEEE Internet of Things Journal, vol. 8, (6), pp. 4186-4210, 2020.
- [31] H. Tyagi, R. Kumar and S. K. Pandey, "A detailed study on trust management techniques for security and privacy in IoT: Challenges, trends, and research directions," High-Confidence Computing, pp. 100127, 2023.
- [32] M. Nikravan and M. Haghi Kashani, "A review on trust management in fog/edge computing: Techniques, trends, and challenges," Journal of Network and Computer Applications, vol. 204, pp. 103402, 2022. Available: <https://www.sciencedirect.com/science/article/pii/S1084804522000613>. DOI: 10.1016/j.jnca.2022.103402.
- [33] A. Shuhaiber and I. Mashal, "A multi-layered trust model in the internet of things smart home ecosystem," in 2024 11th International Conference on Wireless Networks and Mobile Communications (WINCOM), 2024.
- [34] K. Ahmadi, R. Javidan and H. Park, "A Trust Based Anomaly Detection Scheme Using a Hybrid Deep Learning Model for IoT Routing Attacks Mitigation." IET Information Security (Wiley-Blackwell), vol. 2024, 2024.
- [35] S. Singh, V. Anand and S. Yadav, "Trust-based clustering and routing in WSNs using DST-WOA," Peer-to-Peer Networking and Applications, pp. 1-13, 2024.
- [36] A. Padma and M. Ramaiah, "GLSBIoT: GWO-based enhancement for lightweight scalable blockchain for IoT with trust based consensus," Future Generation Comput. Syst., vol. 159, pp. 64-76, 2024.
- [37] F. Azzedin, "Mitigating denial of service attacks in RPL-based IoT environments: trust-based approach," IEEE Access, vol. 11, pp. 129077-129089, 2023.

Forecasting the Emergence of a Dominant Design by Classifying Product and Process Patents Using Machine Learning and Text Mining

Koji Masuda¹, Yoshinori Hayashi², Shigeyuki Haruyama^{3*}

Graduate School of Sciences and Technology for Innovation, Yamaguchi University, Ube, Japan^{1,2}
Graduate School of Innovation and Technology Management, Yamaguchi University, Ube, Japan³

Abstract—Forecasting the emergence of a dominant design in advance is important because the emergence of the dominant design can provide useful information about the external environment for the product launch. Although the emergence of the dominant design can only be determined as a result of the introduction of the product into the market, it may be possible to predict the emergence of the dominant design in advance by applying a solution based on patent analysis. In the newly proposed technique of separating patents, we can capture changes in the state of technological innovation and analyze the emergence of the dominant design, but there is a problem that it requires processing of large amounts of patent data, and that the processing involves subjective judgments by experts. This study focuses on analyzing technological innovation trends using an approach that separates product patents from process patents, investigates whether this approach can be applied to machine learning, and aims to develop a learning model that automatically classifies patents. We applied text mining to patent information to create structured data sets and compared nine different machine learning classification algorithms with and without dimensionality reduction. The approach was effectively applied to machine learning, and the Random Forest, AdaBoost and Support Vector Machine models achieved high classification performance of over 95%. By developing these learning models, it is possible to objectively forecast the emergence of a dominant design with high accuracy.

Keywords—Dominant design; patent analysis; technological innovation; machine learning; text mining; classification

I. INTRODUCTION

A company's introduction of a product into a market can significantly change its competitive environment [1], while the external environment affects market entry [2], [3]. Thus, the timing of market entry is strategically important for companies [4]. Dominant design is defined as a design that has achieved market dominance [5], and some previous studies have discussed market entry timing in relation to dominant design. These studies point out that companies that enter the market when a dominant design is likely to emerge while timing their entry will win the market [6], and that entering the market just before the emergence of the dominant design is particularly advantageous and tends to have a low probability of failure [7]. However, the emergence of the dominant design is recognized as a result of a product's entry into the market and thus can only be known in retrospect [5], [8]. If the timing of market entry can be accurately predicted in advance, the probability of success

can be increased by formulating and implementing a growth and technology strategies in accordance with that timing. Therefore, predicting the timing of the emergence of the dominant design is necessary.

Since the timing of the emergence of the dominant design is when the competitive advantage shifts from product innovation to process innovation [5], it is necessary to capture the change in the state of technological innovation in order to predict the emergence of the dominant design. Since patent information is important as an innovation indicator for companies [9] and is useful as an information source for predicting future products [1], patent analysis can be used to predict the state of technological innovation.

The problem of patent analysis, which is a complex and time-consuming process [10] and involves subjective and qualitative judgments of experts [11], [12], is well known. We propose a new technique for patent analysis that separates patents related to product innovation (product patents) from those related to process innovation (process patents) [13], and show that the timing of the emergence of the dominant design can be predicted using this technique by analyzing specific product case studies [14]. However, subjective processing by experts still remains, and there are concerns about the variability of the processing results. Patent analysis using automatic classification with machine learning allows for objective forecast of the emergence of a dominant design with high accuracy and stability. The increased efficiency provided by automatic classification contributes to reducing the activities and investments of companies for patent analysis.

In order to forecast the emergence of the dominant design, this study examines whether the idea can be applied to machine learning based on a technique for separating product patents from process patents and develops a learning model that automatically classifies patents into product patents and process patents. Specifically, we apply text mining to patent information, which is textual information, to extract features to be input to the modeling. We compare several classification algorithms for supervised learning and construct an appropriate learning model.

This paper is organized to provide a comprehensive understanding of analytical methods for automatically classifying patents into product and process patents using machine learning and text mining. Section I provides background and emphasizes the importance of predicting the

emergence of the dominant design. Section II presents a literature review. Section III describes the methodology of the study, and Section IV presents the results and discussions. Section V presents the conclusions.

II. LITERATURE REVIEW

A. Dominant Design

There are previous studies that have analyzed the emergence of a dominant design based on patent information. In an analysis focusing on the number of patents per technology category, the dominant design is composed of technology categories with a large number of patents [15]. In an analysis focusing on the citation rate of patents, the dominant design exists when the ratio of patents citing the same patent in a patent class is 50% or more [16]. These analyze whether or not a dominant design emerges, but do not provide any information on the timing of the emergence of the dominant design.

The timing of the emergence of the dominant design is said to be the boundary between the fluid phase and the transition phase in "the dynamics of innovation" model [5]. Capturing changes in the state of technological innovation means that it may be possible to predict the timing of emergence by estimating the profiles of product and process innovation in the aforementioned model.

B. Classifying Product and Process Patents

In Japanese patent law, inventions are categorized into inventions of a product and inventions of a process, and inventions of a process are further categorized into inventions of a process that produces a product and inventions of a process that does not produce a product [17]. Patent laws in Europe and the United States categorize inventions in almost the same way [18], [19]. Product inventions are inventions relating to the product itself. Process inventions, on the other hand, refer to inventions relating to a process for manufacturing or producing a product, inventions relating to a process for improving or enhancing the characteristics of a product, and inventions relating to a process for expressing the function of a product, based on the content of the invention.

The patents related to product innovation and process innovation in the "dynamics of innovation" model are called product patents and process patents, and the two types of patents are shown in Table I, which maps them to the various inventions mentioned above.

Previous studies on the classification of product and process patents propose methods for experts and specialists to judge their classification, and they focus on the description of the F-term, which is a Japanese patent classification code [20], or on the title of the invention [14]. In both cases, the large amount of patent data has to be processed subjectively by experts, and there are concerns about the stability and efficiency of the processing results.

C. Machine Learning for Patent Analysis

Previous studies point out that patent analysis requires very large data sets and expertise, and that manual, subjective processing is time-consuming and costly [21], [22], [23], thus automation using machine learning is eagerly awaited. The

focus of patent analysis is on extracting specific technology information and investigating technology trends [24], and the analysis of "technology" is the main objective. For example, the following are examples of patent analysis using machine learning. In terms of technology information extraction, there is the extraction of vacant technology [25], the identification of emerging technologies [26], and the extraction of differences in technologies of competing companies [27]. In addition, for technology trend studies, there are the future technology trends in a certain industry [28], the trajectory of technology development from the present to the future [29], the current and future technology impact in a certain technology field [30], and the prediction of technology convergence in a certain industry or technology field [31].

TABLE I. CLASSIFICATION OF PRODUCT AND PROCESS PATENTS

Categories of Invention		Contents of Invention	Classification of Patent
Inventions	Inventions of a process	Inventions of a production process	Process patents
		Inventions of a non-production process	
	Inventions of a product	Inventions relating to the product itself	

We study the use of patent information not to analyze technologies for R&D, but to analyze innovations as value creation for customers, markets, and society [14]. In conventional patent analysis using machine learning, the main target of analysis is the investigation of technology trends, while few research reports are known to focus on the analysis of innovation. We focus on technological innovation in the analysis of patents using machine learning, especially in the investigation of innovation trends as shown in "the dynamics of innovation" model.

In the next section, we describe a patent analysis method that focuses on "title of the invention" as patent information, and automatically classifies patents into product patents and process patents by using machine learning and text mining.

III. METHOD

This study followed the process model developed by the Cross Industry Standard Process for Data Mining (CRISP-DM) project [32], which was a de facto standard process model for data mining projects that can be applied independently of industries and research domains [33]. Table II shows an overview of the individual phases of CRISP-DM, which consists of six phases, as well as the general tasks [34].

The following subsections described the methodology of this study for each phase.

TABLE II. PROCESS MODEL OF CRISP-DM

Phase	Outline and Generic Task
Business understanding	The business understanding phase focuses on understanding the objectives and requirements of the project from a business perspective, then developing data mining objectives and creating a plan, including an initial evaluation of tools and techniques, to achieve the objectives.
Data understanding	The data understanding phase begins with collecting the data to be used in the analysis, organizing the characteristics of the data to become familiar with the data, and performing simple tabulations. Activities proceed to understanding the meaning of the data and checking the quality of the data.
Data preparation	The data preparation phase includes all activities to prepare the final data set (the data supplied to the modeling tool) from the initial data. These activities include data selection, data cleaning, data construction, data integration, and data transformation.
Modeling	The modeling phase involves selecting and applying different modeling techniques and adjusting their parameters to optimal values. In general, there are several techniques for the same type of data mining problem. In the case of supervised learning, the data sets are usually divided into training and test data set, the model is built on the training data set, and its quality is estimated on the test data set. Metrics to evaluate the quality and validity of the model are generated before the model is built.
Evaluation	During the evaluation phase, it is important to review the steps taken to ensure that the model adequately achieves the business objectives. A more detailed review of data mining is appropriate to determine if any tasks have been overlooked.
Deployment	During the deployment phase, the process of building the model is documented, the entire project is reviewed, and a final report is compiled for future use.

A. Business Understanding

The objectives of the data analysis project were understood, and the resources and constraints for implementation were identified. Next, the data mining objectives were determined from a technical perspective and an action plan was developed. As a source of patent information, registered patents on projector products that predict and validate the emergence of the dominant design were selected [14]. An initial evaluation of tools and techniques was performed during this phase.

B. Data Understanding

As shown in Table I, inventions are classified into inventions of a product and inventions of a process. This classification can be easily made by paying attention to the "title of the invention" in the patent specification. In other words, it can be determined whether the keyword "process" is included in the "title of the invention" or not. Based on this understanding of the data, the "title of the invention" data of each patent was collected as the patent information to be used in the analysis.

Inventions of a product and inventions of a process were simply tabulated. Data quality was checked for completeness and missing values.

All inventions of a product belong to product patents. On the other hand, inventions of a process belong either to product patents or to process patents. Therefore, it is necessary to determine from the content of the "title of the invention" whether

the patent containing the invention is a product patent or a process patent.

C. Data Preparation

For the inventions of a process, the following two structured data sets were constructed to prepare a final data set from the collected "title of the invention" data. They were combined into the final data set.

One was a high-dimensional structured data set created by tokenization, data cleaning and feature extraction with TF-IDF (Term Frequency – Inverted Document Frequency) using text mining techniques on the unstructured text data of the "title of the invention". TF-IDF is a feature that assigns a lower weight to words that appear in more documents relative to the frequency of occurrence of the word. TF_{ij} is denoted by tf_{ij} , the frequency of the word w_j in document d_i (1), and IDF_j is denoted by Eq. (2), where N is the total number of documents and df_j is the number of documents containing the word w_j [35]. $TF-IDF_{ij}$ is denoted by Eq. (3), where the document refers the "title of the invention".

$$TF_{ij} = tf_{ij} \quad (1)$$

$$IDF_j = \log(1 + N/df_j) \quad (2)$$

$$TF-IDF_{ij} = TF_{ij} * IDF_j = tf_{ij} * \log(1 + N/df_j) \quad (3)$$

The other was a single row of structured data set formed by labeling whether the patent containing the inventions of a process method was a product patent or a process patent. The labeling was performed by engineers and experts familiar with the technology.

D. Modeling

We conducted modeling to classify patents containing inventions of a process into product patents and process patents using a machine learning algorithm. We investigated the well-known supervised learning classifiers: Decision Trees (DT), Linear Discriminant (LD), Logistic Regression (LR), Naive Bayes (NB), Support Vector Machine (SVM), k-Nearest Neighbors (kNN), Random Forest (RF), AdaBoost (AB), and Neural Networks (NN) [36], [37], [38], [39]. The data set created in the previous subsection was used as input, and the hyperparameters were tuned for each classification model using Bayesian optimization.

Because dimensionality reduction has the potential to improve model performance, all models were run with and without dimensionality reduction using Principal Component Analysis (PCA) on the input data set.

The ratio of training and test data sets was set to 80% and 20%. To avoid overfitting, a five-fold cross-validation was used for training. That is, the 80% training data set is divided into 64% for training and 16% for validation. The Mean Accuracy of the cross-validation on the training data set was calculated as a metrics of the quality of the classification model. In addition, we calculated Accuracy using the test data set, Recall, which indicates how well the model reproduces actual results, Precision, which indicates how well the model corrects predicted results, and F1-Score, which is the harmonic mean of Precision and Recall with a trade-off relationship. The confusion

matrix shown in Table III and the following Eq. (4), (5), (6) and (7) were used in these calculations.

TABLE III. CONFUSION MATRIX

		Prediction	
		Positive	Negative
Actual	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Positive (FP)	True Negative (TN)

$$Accuracy = (TP + TN) / (TP + FN + FP + TN) \quad (4)$$

$$Precision = TP / (TP + FP) \quad (5)$$

$$Recall = TP / (TP + FN) \quad (6)$$

$$F1-Score = 2 * TP / (2 * TP + FP + FN) \quad (7)$$

The classification models with good values for these metrics were selected.

E. Evaluation

To confirm that the business objective of forecasting the emergence of a dominant design was feasible, each step was reviewed and confirmed.

F. Deployment

The entire project, including the procedures for data preparation by text mining and modeling by machine learning, was summarized in this paper.

IV. RESULTS AND DISCUSSIONS

The results were presented in the order of the phases outlined in the previous section, followed by some discussion.

A. Business Understanding

The objective of data mining is the automatic classification of product patents and process patents, and in particular the classification of "inventions of a process" into product patents and process patents. For the initial evaluation of the tools and techniques, we conducted a preliminary experiment on 1,000 registered patents for projectors, which is a simplified version of a planned main experiment and confirmed that the planned experiment was feasible. In the preliminary experiment, we went through the procedures of data preparation, modeling, and evaluation, and found that it was likely to provide the desired accuracy, thus we decided to proceed with the main experiment. We used MATLAB R2023b version, Statistics and Machine Learning Toolbox, and Text Analytics Toolbox from Mathworks as the tools to perform text mining and machine learning.

B. Data Understanding

Registered patents were extracted from the Japan Patent Office (JPO) database using the search conditions of patent classification code and period. For the patent classification

codes, we used theme codes that are unique to Japan. Theme codes are organized by technical groupings and can be represented almost equivalently by a bundle of multiple IPCs. The number of registered patents extracted under the conditions shown in Table IV was 11,318. Based on a simple aggregation by the presence or absence of the keyword "process" in the "title of the invention", 8,932 patents were classified as inventions of a product, and 2,386 patents were classified as inventions of a process.

TABLE IV. SEARCH CONDITIONS

Item	Query
Database	Japan Patent Office
Patent classification code (Theme code)	2K103 or 2K203
Period	1/1/1981 – 12/31/2020
Search date	10/30/2023

The "title of the invention," which includes both inventions of a product and inventions of a process, was positioned as inventions of a process. This is because inventions of a process are classified as product patents and process patents in the next phase of modeling.

C. Data Preparation

For the 2,386 unstructured text data of "title of the invention," we performed cleaning and computed TF-IDF to create a structured data matrix of numerical variables with 2,386*714 dimensions. After tokenization, the cleaning process included stemming, erasing punctuation, removing stop words, removing a single character, and standardizing synonyms.

Labeling was performed by experts to create a 2,386*1 dimensional structured data matrix with process patents as "A" and product patents as "B." By merging the two structured data sets, a 2,386*715-dimensional matrix was created as the final data set for the modeling.

D. Modeling

Table V shows the mean accuracy of each model on the training data set for each of the nine classifiers. To check the effect of dimensionality reduction, PCA was performed on the input data set to achieve a cumulative contribution rate of at least 95%, and the dimensionality was reduced from 714 dimensions to 343 dimensions. All models were run without dimensionality reduction (without PCA) and with dimensionality reduction (with PCA).

Only NB and kNN had mean accuracy below 90%, while the rest of the models exceeded 90%. In particular, the AB model achieved good mean accuracy of over 95%.

Table VI shows the calculation results for each metric on the test data set. For each classification algorithm, the model with the higher mean accuracy was selected with and without PCA. In the case of with PCA, "with PCA" was added to the name of the classifier.

TABLE V. COMPARISON BETWEEN WITHOUT PCA AND WITH PCA (TRAINING DATASET)

Classification Algorithm (Classifier)	Mean Accuracy on the Training Data Set	
	Without PCA	With PCA
DT	94.1%	85.8%
LD	85.7%	93.8%
LR	89.3%	91.0%
NB	70.7%	84.7%
SVM	94.9%	94.1%
kNN	89.1%	89.6%
RF	94.7%	89.9%
AB	95.1%	92.1%
NN	94.0%	94.1%

TABLE VI. COMPARISON BETWEEN WITHOUT PCA AND WITH PCA (TEST DATASET)

Classification Algorithm (Classifier)	On the Test Data Set			
	Accuracy	Precision	Recall	F1-score
DT	94.6%	96.3%	95.7%	96.0%
LD with PCA	93.1%	96.2%	93.5%	94.8%
LR with PCA	93.1%	95.3%	94.4%	94.9%
NB with PCA	81.6%	82.9%	91.6%	87.1%
SVM	95.6%	95.8%	97.8%	96.8%
kNN with PCA	91.0%	92.4%	94.4%	93.4%
RF	95.6%	97.8%	95.7%	96.7%
AB	95.2%	96.9%	96.0%	96.4%
NN with PCA	94.3%	94.6%	97.2%	95.9%

Accuracy was highest for SVM and RF, followed by AB at more than 95%. Precision was highest for RF, and AB, DT, LD with PCA, SVM, and LR with PCA exceeded 95%. Recall was highest for SVM, followed by NN with PCA, AB, RF, and DT over 95%. The F1-Score, the harmonic mean of Precision and Recall, was also highest for SVM, followed by RF, AB, DT, and NN with PCA exceeding 95%.

Tables VII, VIII, and IX show the confusion matrices on the test data set for the three models SVM, RF, and AB, which performed well above 95% on all four metrics.

The high performance of several models in patent classification in this study suggested that the "title of the invention" was appropriate as patent information data, that the data preprocessing was effective, and that the idea of separating product and process patents was applicable to machine learning.

In this experiment, which combined nine classification algorithms with and without PCA, the prediction model using SVM, RF and AB algorithms achieved higher performance.

TABLE VII. CONFUSION MATRIX OF SVM

		Prediction	
		A	B
Actual	A	316	7
	B	14	140

TABLE VIII. CONFUSION MATRIX OF RF

		Prediction	
		A	B
Actual	A	309	14
	B	7	147

TABLE IX. CONFUSION MATRIX OF AB

		Prediction	
		A	B
Actual	A	310	13
	B	10	144

E. Evaluation

In order to forecast the emergence of a dominant design, which is the objective of the business, it was important to capture changes in the state of innovation. The changes were indicated by the trends of product patents and process patents according to the "the dynamics of innovation" model. Based on Table I, we categorized the patents to be analyzed into product patents and process patents. Since inventions of a product can be easily identified from the "title of the invention," we focused on classifying inventions of a process into product patents and process patents. The data preparation and modeling resulted in several prediction models with high classification performance in terms of the overall model correctness rate and the F1-Score, which is a balance between actual and predicted results.

Since the trends of product innovation and process innovation are visualized according to the classification results of the prediction model, and the emergence of a dominant design is predicted, we considered precision to be particularly important among the four metrics for this business objective. Therefore, the prediction model with the highest precision performance was preferred. Table VI shows that the precision performance of the RF model is 97.8%, and the predicted trends of product and process patents are almost the same as their actual trends. The above review confirmed that no tasks were missed in the steps performed and that the business objective was properly achieved. It also demonstrated that the automatic classification by machine learning worked effectively.

F. Deployment

In this project, the business objective was to predict the emergence of a dominant design, and the data mining goal for this purpose was to automatically classify "inventions of a

process" into product patents and process patents. Through data understanding, data preparation, modeling, and evaluation, the validity of the data we focused on and the predictive models that achieved high performance were confirmed, and thus the project was completed. We summarized the data mining process and results according to the CRISP-DM process model in this paper.

G. Discussions

The effect of dimensionality reduction on the classification algorithm was discussed by comparing the models with and without PCA. Table V shows that the five models with a higher mean accuracy with PCA than without PCA were LD, LR, NB, kNN, and NN. According to the idea that machine learning models can be divided into three models: geometric, probabilistic, and logical models [40], these five models were included in the geometric and probabilistic models. The models with a difference of less than 1% between those with and without PCA were SVM, kNN, and NN, all of which were geometric models. On the other hand, four models, DT, SVM, RF, and AB, had a higher mean accuracy without PCA than with PCA. Since RF and AB are ensemble learning with tree models, these three models including DT are considered to be logical models. These results suggested that dimensionality reduction may be effective in improving the performance of geometric and probabilistic models in this experiment.

It is known that SVM and ensemble learning, such as RF and AB, tend to show relatively high performance compared to other algorithms, and this study was consistent with this finding, as well as previous studies comparing multiple algorithms [36],[37].

Although this study achieved good results in classification performance, some limitations need to be considered. Instead of classifying product inventions and process inventions directly from the "title of the invention," this study focused on separating "inventions of a product" and "inventions of a process" from the "title of the invention" by a simple procedure first, and then classifying product inventions and process inventions from "inventions of a process." We used TF-IDF and five-fold cross-validation for feature extraction in data preparation and data partitioning in modeling, respectively, but other techniques could be considered to further improve classification performance.

V. CONCLUSIONS

In order to forecast the emergence of dominant designs, this study investigated an automatic classification method for product and process patents according to the CRISP-DM process model applied to data mining projects. We focused on "title of the invention" as patent information, extracted TF-IDF features by text mining, and evaluated nine classification algorithms with and without PCA by machine learning. As a result, the prediction model using the RF, AB, and SVM algorithms achieved over 95% performance in all four metrics: accuracy, precision, recall, and F1-Score. In the classification of product patents and process patents, it was shown that the "title of the invention" was appropriate as patent information data, that data preprocessing was effective, and that the idea of a technique for separating product patents from process patents was applicable to machine learning.

By using patent analysis, which uses machine learning and text mining to capture changes in product innovation and process innovation, that is, changes in the state of technological innovation, it is possible to objectively forecast the emergence of a dominant design with high accuracy. Therefore, it can be a useful piece of information about the external environment for companies to formulate and implement growth and technology strategies.

Increased efficiency in analyzing trends in technological innovation can lead to a reduction in the activities and investments of companies. In addition, the resources generated by the reduction are expected to make a new contribution.

REFERENCES

- [1] J. M. Gerken, M. G. Moehrl, and L. Walter, "One year ahead! Investigating the time lag between patent publication and market launch: Insights from a longitudinal study in the automotive industry," *R&D Management*, vol. 45, no.3, pp. 287-303, 2015.
- [2] B. L. Bayus, "Speed-to-market and new product performance trade-offs," *Journal of Product Innovation Management*, vol. 14, no. 6, pp. 485-497, 1997.
- [3] D. Morschett, H Schramm-Klein, and B. Swoboda, "Decades of research on market entry modes: What do we really know about external antecedents of entry mode choice?," *Journal of International Management*, vol. 16, no. 1, pp. 60-77, 2010.
- [4] F. F. Suarez, S. Grodal, and A. Gotsopoulos, "Perfect timing? Dominant category, dominant design, and the window of opportunity for firm entry," *Strategic Management Journal*, vol. 36, no. 3, pp. 437-448, 2015.
- [5] J. M. Utterback, "Mastering the dynamics of innovation," Boston, MA, USA: Harvard Business Review Press, 1994.
- [6] C. C. Markides and P. A. Geroski, "Fast second: How smart companies bypass radical innovation to enter and dominate new markets," San Francisco, CA, USA: Jossey-Bass, 2004.
- [7] C. M. Christensen, F. F. Suarez, and J. M. Utterback, "Strategies for survival in fast-changing industries," *Management Science*, vol. 44, no. 12-part-2, pp. 207-220, Dec. 1998.
- [8] P. Anderson and M. L. Tushman, "Technological discontinuities and dominant designs: A cyclical model of technological change," *Administrative Science Quarterly*, vol. 35, no. 4, pp. 604-633, Dec. 1990.
- [9] S. Rocheska, D. Nikoloski, M. Angeleski, and G. Mancheski, "Factors affecting innovation and patent propensity of SMEs: Evidence from Macedonia," *TEM Journal*, vol. 6, no. 2, pp. 407-415, May 2017.
- [10] K. Masuda and S. Haruyama, "Forecasting technology trends based on separation of product inventions and process inventions: The technology S-curve," *IOP Conference Series: Materials Science and Engineering*, 1034, 012123, 2021.
- [11] K. Masuda and S. Haruyama, "Forecasting the timing of the emergence of a dominant design: The case of the projectors," *International Journal of Technology*, vol. 15, no. 1, pp. 138-153, 2024.
- [12] N. Clymer and S. Asaba, "A new approach for understanding dominant design: The case of the ink-jet printer," *Journal of Engineering and Technology Management*, vol. 25, no. 3, pp. 137-156, 2008.
- [13] A. Brem, P. A. Nylund, and G. Schuster, "Innovation and de facto standardization: The influence of dominant design on innovative performance, radical innovation, and process innovation," *Technovation*, vol. 50-51, pp. 79-88, 2016.
- [14] The Japan Patent Office, "Introduction to the intellectual property act," Accessed: Jul. 1, 2024. [Online]. Available: https://www.jpo.go.jp/news/kokusai/developing/training/textbook/document/index/Introduction_to_The_Intellectual_Property_Act.pdf
- [15] The European Patent Office, "Guidelines for examination in the European Patent Office," Accessed: Jul. 1, 2024. [Online]. Available: <https://www.epo.org/en/legal/guidelines-epc>
- [16] The United States Patent and Trademark Office, "Manual of patent examining procedure (MPEP)," Accessed: Jul. 1, 2024. [Online]. Available: <https://www.uspto.gov/web/offices/pac/mpep/index.html>

- [17] Y. Ishii, K. Kaminishi, and S. Haruyama, "A study of identifying trends in projector using F-term codes from Japanese patent applications," *International Journal of Integrated Engineering*, vol. 13, no. 7, pp. 324-332, 2021.
- [18] X. Zhang, "Interactive patent classification based on multi-classifier fusion and active learning," *Neurocomputing*, vol. 127, pp. 200-205, 2014.
- [19] J.-L. Wu, P.-C. Chang, C.-C. Tsao, and C.-Y. Fan, "A patent quality analysis and classification system using self-organizing maps with support vector machine," *Applied Soft Computing*, vol. 41, pp. 305-316, 2016.
- [20] A. J. C. Trappey, C. V. Trappey, J.-L. Wu, and J. W. C. Wang, "Intelligent compilation of patent summaries using machine learning and natural language processing techniques," *Advanced Engineering Informatics*, vol. 43, 101027, 2020.
- [21] J. Choi, D. Jang, S. Jun, and S. Park, "A predictive model of technology transfer using patent analysis," *Sustainability*, vol. 7, 16175, 2015.
- [22] S. Jun and S.-S. Park, "Examining technological innovation of Apple using patent analysis," *Industrial Management & Data Systems*, vol. 113, No. 6, pp. 890-907, 2013.
- [23] C. Lee, O. Kwon, M. Kim, and D. Kwon, "Early identification of emerging technologies: A machine learning approach using multiple patent indicators," *Technological Forecasting & Social Change*, vol. 127, pp. 291-303, 2018.
- [24] S. Jun and S.-S. Park, "Examining technological competition between BMW and Hyundai in the Korean car market," *Technology Analysis & Strategic Management*, vol. 28, no. 2, pp. 156-175, 2016.
- [25] A. suominen, H. Toivanen, and M. Seppanen, "Firms' knowledge profiles: Mapping patent data with unsupervised learning," *Technological Forecasting & Social Change*, vol. 115, pp. 131-142, 2017.
- [26] L. J. Aaldering and C. H. Song, "Tracing the technological development trajectory in post-lithium-ion battery technologies: A patent-based approach," *Journal of Cleaner Production*, vol. 241, 118343, 2019.
- [27] D. Thorleuchter, D. Van den Poel, and A. Prinzie, "A compared R&D-based and patent-based cross impact analysis for identifying relationships between technologies," *Technological Forecasting & Social Change*, vol. 77, pp. 1037-1050, 2010.
- [28] T. S. Kim and S. Y. Sohn, "Machine-learning-based deep semantic analysis approach for forecasting new technology convergence," *Technological Forecasting & Social Change*, vol. 157, 120095, 2020.
- [29] R. Wirth and J. Hipp, "CRISP-DM: Towards a standard process model for data mining," in *Proceedings of the Fourth International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, Manchester, UK, 2000, pp. 29-40.
- [30] C. Schroer, F. Kruse, and J. M. Gomez, "A systematic literature review on applying CRISP-DM process model," *Procedia Computer Science*, vol. 181, pp. 526-534, 2021.
- [31] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth. *CRISP-DM 1.0: Step-by-step data mining guide*. (2000).
- [32] MathWorks. Help Center. Accessed: Jul. 1, 2024. [Online]. Available: <https://www.mathworks.com/help/index.html>
- [33] Y. Mahmud, N. S. Shaeali, and S. Mutalib, "Comparison of machine learning algorithms for sentiment classification on fake news detection," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 10, pp. 658-665, 2021.
- [34] M. Sivamanikanta and N. Ravinder, "Machine learning-driven integration of genetic and textual data for enhanced genetic variation classification," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 1, pp. 252-262, 2024.
- [35] M. Chistol and M. Danubianu, "Automated detection of autism spectrum disorder symptoms using text mining and machine learning for early diagnosis," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 2, pp. 610-617, 2024.
- [36] S. Fadili, M. Ertel, A. Mengad, and S. Amali, "Predicting optimal learning approaches for nursing students in Morocco," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 4, pp. 94-102, 2024.
- [37] P. Flach, "Machine learning: The art and science of algorithms that make sense of data," New York, NY, USA: Cambridge University Press, 2012.
- [38] K. OuYang and C. S. Weng, "A new comprehensive patent analysis approach for new product design in mechanical engineering," *Technological Forecasting & Social Change*, vol. 78, pp. 1183-1199, 2011.
- [39] G. Kim and J. Bae, "A novel approach to forecast promising technology through patent analysis," *Technological Forecasting & Social Change*, vol. 117, pp. 228-237, 2017.
- [40] J. Lee, N. Ko, J. Yoon, and C. Son, "An approach for discovering firm-specific technology opportunities: Application of link prediction to F-term networks," *Technological Forecasting & Social Change*, vol. 168, 120746, 2021.

Control Interface for Multi-User Video Games with Hand or Head Gestures in Directional Key-Based Games

Oscar Ramirez-Valdez, César Baluarte-Araya, Rodrigo Castillo-Lazo, Italo Ccoscco-Alvis,
Alexander Valdiviezo-Tovar, Alexander Villafuerte-Quispe, Dylan Zuñiga-Huraca
Universidad Nacional de San Agustín de Arequipa, Arequipa, Perú

Abstract—This paper describes the development and implementation of a hand or head gesture-based control interface for video games, enhanced for games that use directional keys. The objective is to develop an adaptive control system for a multiplayer video game that allows users to choose between the use of traditional directional keys or a gesture-based interface. The methodology used follows the Cross-Industry Standard Process for Data Mining (CRISP-DM) development model, which allows a structured integration of analysis, design, implementation and evaluation steps. Technologies such as OpenCV, MediaPipe and deep learning algorithms are used, translating hand movements into directional commands in real time. In addition, the system integrates a client-server architecture based on Node.js that supports multiple users, enabling an immersive gaming experience on PC and mobile platforms. The results highlight the accuracy of the system and its potential to improve accessibility, especially for users with motor disabilities by using their hands or head movements to control the directional keys. Concluding that the control interface for multi-user video games provides the necessary support to gamers in performing the task, promoting accessibility in the entertainment environment.

Keywords—Control interface; video games; artificial vision; gesture-based interface; directional commands; human-computer interaction; deep learning algorithms; accessibility; real-time; pattern recognition

I. INTRODUCTION

Gesture interaction has revolutionised the gaming experience [13], eliminating the dependence on traditional keyboards and controls. This paper presents an interface that allows video games [36] to be controlled by hand movements, using computer vision [3] and real-time pattern recognition. The solution aims to improve immersion and accessibility, especially for users with motor disabilities, promoting their integration in recreational and therapeutic activities.

The interface employs gesture detection algorithms and a robust client-server system, compatible with PC and mobile devices, providing an inclusive experience. As a result, a gesture recognition system was designed and developed to replace the directional keys, allowing precise and fluid control in real time. In addition, it supports multiple users with individual network configurations and in-game character selection. Developed in Unity, the game integrates traditional and gesture controls, adapting to both PC and mobile.

It is concluded that this interface is viable, offering an accessible alternative for controlling video games through hand gestures or head movements, broadening access to entertainment and promoting inclusion.

II. THEORETICAL FRAMEWORK

A. Gesture-based Control Interfaces and Games

Gesture-based control interfaces have revolutionised the way users interact with devices and computer systems [9], opening the door to immersive and intuitive experiences. In the context of video games, these interfaces allow the user to control game elements through body gestures, specifically hand movements. The advantages of this approach include greater immersion and accessibility [23], as it allows play without the need for traditional controls such as keyboards or controllers. To achieve this, advanced gesture recognition [1] and computer vision technologies are used to interpret the user's movements and translate them into real-time actions.

B. Artificial Vision and Gesture Recognition

Artificial vision is a discipline that allows machines to process and interpret the visual world. This technology uses image processing algorithms and automatic learning techniques to identify objects, gestures and patterns in real time [10] [15]. In hand gesture recognition, a sub-area of machine vision, it allows the detection and analysis of specific hand movements to control interactive applications [12]. In the context of this project, libraries such as OpenCV and MediaPipe are fundamental to capture images of hands, identify key points (such as articulations and fingers) and translate this data into control actions for games based on directional keys.

C. Real Time Pattern Recognition

Real-time pattern recognition is key to achieving smooth [2] [11] and accurate interaction in gesture-based interfaces. Real-time recognition systems allow capturing and processing images in a fraction of a second, detecting movements instantaneously. The ability to process gestures in real time is especially relevant for video game applications [25], where any delay can affect the user experience [6] and decrease the effectiveness of the control. To implement this functionality, fast image processing techniques and movement detection and analysis algorithms, optimised to operate on common devices such as webcams, are employed.

D. Image Processing Technologies and Benchmark points Models

Reference point models are essential to accurately recognise the position of fingers and hands. These models identify key points on the hands and, using deep learning techniques, detect specific gestures, such as left, right, up or down movements. These points help determine the orientation and height of the hands, which are essential for translating gestures into commands in the videogame. The precision of these models depends on the quality of the camera and the capacity of the algorithms to quickly process the images.

E. Gesture Control of Video Games: Benefits and Challenges

Gesture control of video games [7] has significant benefits, including greater immersion and accessibility for users with motor limitations. It also allows for a more natural and direct gaming experience, eliminating the need for additional control devices. However, there are also challenges, such as the need for recognition algorithms that work accurately in varied environments and lighting conditions. Also, minimising the delay [18] between capturing the gesture and executing the command in the game is critical to ensure a satisfactory experience.

F. Implementation of the Control Interface

The implementation of hand gesture control interface for video games [7] based on directional keys requires effective integration [19] of various software and hardware elements. In this project, a combination of Python libraries is used for the creation of the graphical user interface [17], camera control and gesture detection. Tkinter is used to develop the user interface, allowing custom settings for sensitivity and direction control. In addition, OpenCV and MediaPipe are employed for processing the captured images and detecting gestures in real time, as shown in Fig. 1 and Fig. 2.



Fig. 1. Control interface.



Fig. 2. Control of racing game in unity.

G. Server Implementation

The development of an efficient server is essential to guarantee real-time communication between players and to maintain synchronisation during the multiplayer gaming experience, see Fig. 3. The server implementation was carried out using modern technologies such as Node.js, which offers a lightweight and scalable environment, together with libraries such as Express for HTTP route management and Socket.io for real-time communication [20].



Fig. 3. Multi-user racing game in Unity.

III. RELATED WORK

The use of computer vision and hand gesture recognition techniques in video games has been an active area of research. The research in [31] presented GestureFlow, a novel hand gesture control system for interactive games that leverages advanced tools such as OpenCV, Mediapipe and Numpy. The study in [32] presented a methodology for gesture-based contactless operations, combining his algorithm with Mediapipe and OpenCV.

Various studies have focused on the development of game applications based on hand gestures. Thus, the study in [33] employed object detection and an artificial neural network for hand gesture recognition in games, using Python and OpenCV; also the study in [34] developed a game that interacts with users through hand gesture movements, using Mediapipe and Pygame; and the study in [35] explored the use of Mediapipe for real-time online games, creating a gesture recognition-based control system using OpenCV and Python.

The integration of hand gestures and voice commands for immersive gaming has also been studied. In study [36] they developed a game control system based on hand gesture recognition using Mediapipe, OpenCV and Python; also study in [37] reviewed the opportunities of using hand gestures to play video games, highlighting the use of Mediapipe and OpenCV for hand tracking and gesture recognition.

In addition, some studies have explored the application of gesture recognition in rehabilitation and quality of life improvement. Thus, the study in [38] presented a human body gesture-controlled gaming application using OpenCV and Mediapipe; on the other hand, the study in [39] developed a camera-based real-time motion detection gaming tool for cervical rehabilitation, employing a convolutional neural network for hand gesture recognition; also the study [40] used OpenCV functions and Mediapipe modelling technology for real-time human movement recognition and interaction in virtual fitness applications.

IV. METHODOLOGY

The methodology used in this work is based on the Cross-Industry Standard Process for Data Mining (CRISP-DM) development model, adapted for the context of gesture recognition and real-time video game control; it includes the following phases:

Phase 1. Understanding the Business and Defining the Objective

The principal objective of the project is to design and implement a control interface for video games that allows users to control the game using hand gestures or head movements, as an alternative to traditional directional keys. It is not only to improve immersion in the game, but also to promote accessibility for players with motor disabilities.

Phase 2. Data Collection and Preparation

Advanced artificial vision technologies, such as OpenCV and MediaPipe, are used to capture real-time images of the user's hand or head gestures via a webcam. This data is then processed and labelled for gesture detection. The model captures key points, such as finger articulations, and this data is translated into commands for directional key control.

Phase 3. Development of the Gesture Recognition Model

Deep learning algorithms are developed and trained to detect and recognise gestures. Key point reference models are implemented to map hand movements and translate them into directions. The algorithms were optimised to achieve a high degree of accuracy and low latency in real time.

Phase 4. Control Interface System Development and Implementation

Gesture detection is integrated with a user interface developed using Tkinter to configure the sensitivity and controls of the system; the interface also allows for actors selection and network settings. A client-server system is implemented using Node.js and Socket.io to ensure real-time communication between users, and enable a fluently gaming experience.

Phase 5. Integration and Evaluation of the System in Multiplayer Games

The system is integrated into a multiplayer video game developed in Unity, which supports the interaction of multiple players simultaneously. The gesture-based control is evaluated in terms of accuracy, latency and user experience compared to traditional control. Tests are conducted in different lighting conditions and types of movement to ensure the robustness of the system.

Phase 6. Optimisation and Results

Once the basic system is implemented, the algorithms are optimised to improve accuracy and minimise latency. The results of the system are analysed using metrics such as response time and gesture accuracy, and compared to traditional control to determine the effectiveness of the interface in multiplayer games.

This approach ensures a robust solution that not only responds to user requirements, but also contributes to achieving accessibility and usability across PC and mobile platforms.

A. Explanation of Control Interface Code

1) *Interface configuration and resources*: The graphical user interface (GUI) is created using Python's Tkinter for interface creation; the necessary libraries are imported to manage the interface, threads of execution, running local files and external URLs for Unity-based games [8].

2) *Camera and game control*: The functions implemented as `start_camera_thread` and `start_camera_tk_thread` are in charge of starting the camera capture using different control methods, such as `run_virtual_steering` and `run_virtual_tk`, which are executed in separate threads so as not to block the principal interface. Both functions take as arguments the control methods for the directional keys (such as hand or head) defined in `get_control_methods`; the `stop_camera_thread` and `stop_camera_tk_thread` functions stop these threads of execution using stop events (`stop_event` and `stop_event_tk`).

3) *Interface design*: For the principal interface, multiple frames are created representing sections such as the camera control and the Unity game, configured by grid to be arranged according to the selected layout.

The functions `show_columns`, `show_quadrants` and `show_rows` allow different GUI layouts to be changed according to user preference, see Fig. 4 `make_draggable` is used to make each frame draggable, offering flexibility in the layout of the interface.



Fig. 4. Design of the interface control.

4) *Personalisation of configurations*: The configurations frame includes sensitivity and distance threshold setting controls, implemented as sliders (`ttk.Scale`) that allow the user to customise the threshold and sensitivity of the gesture detection system. Additionally, the user can choose between

direction controls with different methods (hand or head gestures), defined in the settings_frame Radiobuttons.

5) *Menu bar*: The menu bar provides easy access to camera and configuration options. The Checkbuttons allow you to activate or deactivate the main sections of the interface, such as camera control or the Unity game, the views submenu offers different settings for the layout of the frames (show_columns, show_quadrants and show_rows), shown in Fig. 5.



Fig. 5. Menu options.

6) *Execution of the main window*: The main window is started with `root.mainloop()`, a loop that keeps the interface active until the user decides to close it, by calling the `on_closing` function, which stops all active camera threads.

7) *Camera processing and visualisation with OpenCV: With and Without Graphical User Interface (GUI)*

The `run_virtual_steering` function starts initialising the modules needed for gesture control [28].

- Image processing without GUI

The mediapipe libraries are used for recognition of hands [16] [26] and face, and OpenCV's `cv2` is used for processing the captured image. The webcam is initialised with `cv2.VideoCapture(0)`, capturing the video in real time; a virtual keyboard controller is configured using `pynput.keyboard.Controller`, used to send simulated commands to the game in Unity, see Fig. 6.



Fig. 6. Camera processing without GUI.

- Image processing with GUI

The real-time processed video is integrated into a graphical interface using Tkinter. A subwindow displaying the processed frames in a continuously updated Label component is shown in Fig. 7.



Fig. 7. Camera processing without GUI.

8) *Image processing and hand detection*: The software processes each camera frame in real time [24] [29], the captured image is flipped horizontally with `cv2.flip` for a more intuitive orientation, then converted to RGB before being passed to the mediapipe model to process the `hand_results` and `face_results` [22] [27]. These data allow the position and movement of the hands and face to be determined.

9) *Direction control based on vertical position of the hands*: Are used the reference points obtained to detect the position of the hands on the Y-axis, allowing to differentiate whether the right hand is more up than the left hand or vice versa. When it detects that the right hand is more raised, the program simulates a movement to the right by pressing the `Key.right` key. Similarly, if the left hand is higher, `Key.left` is pressed; shown in Fig. 8.



Fig. 8. Ownership of frames.

10) *Face distance based control*: The program calculates the distance of the face from the area of the face detection box; it is used to detect approaching or moving away movements, activating the `Key.up` key to accelerate as the face approaches and `Key.down` to slow down as it moves away. This control allows to adapt the speed of the vehicle within the game, simulating acceleration and deceleration depending on the proximity of the face, as shown in Fig. 9.



Fig. 9. Head movement conditionals to accelerate or decelerate the vehicle.

11) *Additional hand gesture control*: The program also includes an additional control [14] based on specific hand gestures [3]. If the index finger is fully extended downwards, it is interpreted as an acceleration gesture, activating the Key.up key, when the index finger is raised, it is assumed as a braking gesture, activating Key.down. This logic, according to the initial configuration in the interface, allows using both face movements and hand gestures [3] [30] to provide a more intuitive control experience [13].

12) *Error handling and camera shutdown*: If the OpenCV view window is closed or ESC is pressed, the program stops the image processing loop [5]. The close_camera function ensures that all OpenCV camera and window resources are properly released.

13) *Initialisation of interface styles*: Using the apply_styles function you can customise the visual styles of widgets; thus: a) the TFrame style sets a light blue background, a 5 pixel border and a ridge relief that gives depth to the frame, b) the TLabel style sets a 12 point Helvetica font with the same background, the text uses a darker blue to stand out against the background, c) the TButton style uses a 10 point Helvetica font in bold, with a dark blue background and white text to ensure contrast, d) with style map, the button is adjusted so that, on hover, the background changes to an even darker blue, while keeping the text white to ensure legibility.

14) *Initialisation of the camera module with tkinter*: OpenCV libraries are imported for video handling, MediaPipe for gesture and face detection, and pynput for keyboard input simulation. Global variables, camera_running (check if camera is active), last_action (store the last action performed) and press_duration (measure how long a key is pressed) are initialized to ensure continuous tracking of the system state during execution.

15) *Video capture and processing*: The video_stream function uses OpenCV to capture video in real time. Each frame is processed with MediaPipe to detect hand gestures and faces. The key points of the hands are used to calculate their position in space, allowing to determine actions such as moving left or right. In addition, the points and connections of the hands are visualized in the video to facilitate the interpretation of the system.

16) *Hand gesture detection*: Within video_stream, hand gestures are analyzed using the positions of specific points. This analysis includes logic to avoid simultaneous keystrokes and ensure smooth transitions between actions.

17) *Face distance based control*: The face distance is implemented by comparing the relative size of the detected bounding box around the face [21]. If it increases or decreases

beyond a configured threshold, the “up” or “down” keys are triggered, simulating actions such as accelerating or braking.

B. Explanation of the Server Code

1) *Server initialisation and dependency configuration*: The express, http, and socket.io libraries are used to create a server in Node.js; an HTTP server is configured with http.createServer and its functionality is extended with socket.io to handle real-time connections.

2) *HTTP path to check server status*: The path app.get('/') returns a simple message to confirm that the server is running. The /getPlayers path uses a JSON format returning information about the connected players.

3) *HTTP path to check server status*: The io.on('connection', callback) function handles each client connection. Each connected player is assigned a unique identifier (PlayerID) and is stored in the players object.

4) *Player synchronisation and start of the race*: When the number of connected players reaches the maximum allowed (maxPlayers), the server sends a startRace event to all clients, indicating the start of the game.

5) *Real time position update*: The updatePosition event receives data from the clients, such as position and rotation in the X, Y and Z axes; it is updated in real time in the players object; the server emits the updated list of players through io.emit('updatePlayers').

6) *Handling disconnections*: When a player disconnects, the server deletes his information from the players object and issues an event to update the list of players in the clients.

7) *Server and listening port configuration*: The server starts on port 3020 with address 0.0.0.0, which allows accepting connections from any IP address, ideal for multiplayer environments.

C. Explanation of the Multi-User Game Code in Unity

1) *Automatic object rotation (AutoRotation.cs)*: The AutoRotation script implements a simple functionality to make an object in Unity rotate continuously around its Y-axis; it is controlled by a public variable rotationSpeed.

2) *Dynamic player tracking (FollowPlayer.cs)*: The FollowPlayer script implements functionality for a camera to follow the player in Unity, dynamically adjusting to the player's position and rotation; it includes support for virtual reality (VR) scenarios, see Fig. 10.

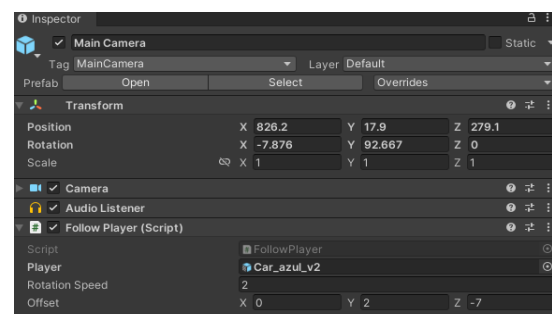


Fig. 10. Player camera configuration.

3) *Interactive In-game console (InGameConsole.cs)*: The InGameConsole script implements an in-game console in Unity, useful for real-time debugging; it allows to display system and user messages in a panel, see Fig. 11.



Fig. 11. View of the start of the game.

4) *Panel management and connection in the main menu (Panel.cs)*: The Panel script manages user interaction with a main menu, providing options to configure a network connection, select colors for a car and manage other related panels.

5) *Player behaviour (player.cs)*: The player.cs script controls the player's behavior in the game, including movement, interaction with the environment, updating the user interface and communication with the server; it is essential to manage the game logic.

6) *Player movement management (player.cs)*: The Update method is called once per frame and handles the main logic of the player's movement. Depending on the platform and the input mode (keyboard or gestures), the input values for horizontal and vertical movement are obtained.

7) *Straighten the overturned car (player.cs)*: The RightCar method is responsible for straightening the player's car by applying an upward force and continuing the game.

8) *User interface update (player.cs)*: The UpdateUI method updates the points and lives texts in the user interface; it is called whenever the player's points or lives change.

9) *Collision management (player.cs)*: The OnCollisionEnter method is called when the player's car collides with another object, if it has the ObjectCollision tag, the player's points are reduced, if the points reach zero, a life is reduced and the points are reset. If the lives reach zero, a "You lost" message is displayed.

10) *Restart car position (player.cs)*: The ResetCarPosition method resets the position and rotation of the player's car to its initial state. This is used when the car falls off the stage or when resetting the player's points and lives.

11) *Management and connection to the server (SocketManager.cs)*: The SocketManager.cs script in Unity handles the connection and communication with a server via WebSockets; it is crucial for the multiplayer functionality of the game, allowing data synchronization between players and the server.

12) *Connection to server (SocketManager.cs)*: The ConnectToServer method establishes the connection to the server using the IP address and the color of the car. It configures connection, disconnection and error, and defines handlers for various game events.

13) *Player ID assignment (SocketManager.cs)*: The OnAssignPlayerID method handles the player ID assignment event; it gets the ID from the server's response and stores it in a local variable.

14) *Initialisation and player update (SocketManager.cs)*: The OnInitializePlayers method creates or updates players at the start of the game, and OnUpdatePlayers updates player positions and rotations during the game.

15) *Position and rotation sending (SocketManager.cs)*: The UpdatePosition method sends the player's position and rotation to the server. It converts the data to a JSON object and outputs it to the server via the socket.

16) *Point and life adjustment in the interface (size.cs)*: The script tamano.cs in Unity adjusts the position and size of dot and life texts in the user interface.

D. Explanation of the Multi-User Game Code in Unity

The main figures of the video game scenario are shown below.

- Principal Camera, shown in Fig. 12.

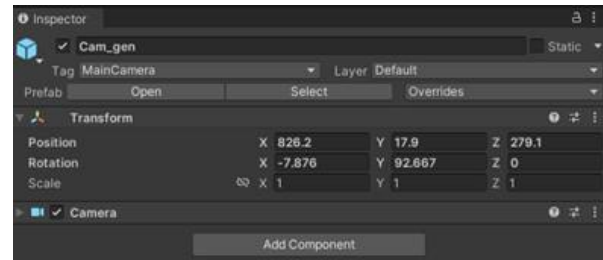


Fig. 12. Game camera position configuration.

- Vehicle models, we organized the designs of the vehicles to compete in the prefabs folder; this is shown in Fig. 13.



Fig. 13. 3D model of vehicles.

- The design of the race track, as would be the scenario, is shown in Fig. 14.

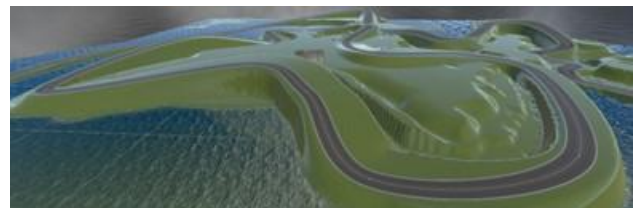


Fig. 14. Running track along the terrain.

- The configuration of the race track is shown in Fig. 15.

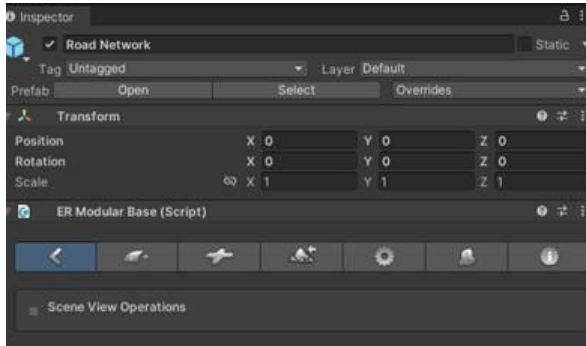


Fig. 15. Race track configuration.

- Principal player design, as shown in Fig. 16.



Fig. 16. Local player configuration.

- Principal menu configuration, three sub-panels are organized within a canvas where the user can configure the ip and connection port, as well as choose the color of the vehicle. If the entered data is validated, the system displays a confirmation message to start the multiplayer racing game, as shown in Fig. 17.

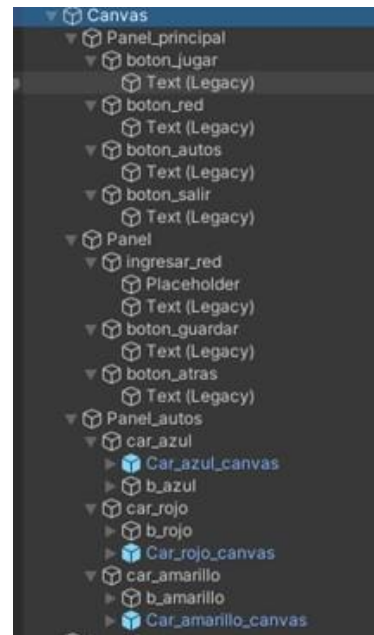


Fig. 17. Components view in unity.

- The principal menu of the game is displayed in the interface shown in Fig. 18.



Fig. 18. Main menu in the game.

- Eleccion of the player's cart, the options referred to the colours, is shown in Fig. 19.



Fig. 19. Selection of the player's car.

- Successful connection view of the player, ready to complete the players, is shown in Fig. 20.



Fig. 20. View of successful player connection.

V. RESULTS AND DISCUSSION

A robust client-server system was built using Node.js, capable of supporting a configurable number of players. Each client, developed in Unity, incorporates a script called SocketManager that facilitates bidirectional communication with the server. This allows to synchronise the start of the game when all players have connected and configured.

In the client, a start panel was designed that offers each player the possibility of configuring the network parameters, selecting a character and waiting for the minimum number of participants configured in the server to be reached before starting the game.

The developed videogame was adapted to run on both PC and mobile devices Tablet, Smartphone, see Fig. 21; providing a multiplatform experience. In addition, two control options were incorporated for players using PCs:

- Traditional Control: Use of directional keys to move left, right, forward or backward.
- Gesture Control Interface: A system that uses the device's camera to detect hand and face gestures, which are mapped to the game's directional key actions.



Fig. 21. Video game running on Multiplatform, PC, Tablet, smartphone.

In the following figures the execution of the multiplatform videogame is shown, in Fig. 22 the cars in full race can be seen on two platforms, in Fig. 23 the control of the red car with hand gestures, in Fig. 24 the car at a different speed leaves the circuit, in Fig. 25 with hand gestures the car returns to the circuit and continues the race.



Fig. 22. Cars in full race on two platforms, PC and tablet.



Fig. 23. Controlling the red car with hand gestures.



Fig. 24. Car at a different speed wanders off track.



Fig. 25. With a hands gesture the car is steered back to the circuit and the race continues.

The system executes key presses using two types of control: hand gestures and facial movements, both detected by MediaPipe [14]. For hand control, the relative position of the hands is evaluated: if the right hand is higher than the left hand, the system simulates the action of pressing the right arrow key; if the left hand is higher, it simulates the action of pressing the left arrow key. The green node, which marks the centre between the two hands, indicates that both hands are centred. In addition, the facial movement adjusts the distance control, triggering

zoom in or zoom out actions, depending on the position of the face in front of the camera. For acceleration and deceleration control [31], the system uses the position of the index finger relative to the wrist: if the index finger is higher than the wrist, it simulates the action of pressing the ‘up’ key to accelerate; if it is lower, it simulates the action of pressing the ‘down’ key to decelerate, thus allowing the speed to be dynamically adjusted by these movements.

As seen in Fig. 26, Fig. 27, Fig. 28, the system provides visual messages on screen, such as ‘Move left (Right hand higher)’, which orients the user on the movements required to interact as discussed by [11] with the virtual steering wheel for the game developed in Unity. This information is useful, that in some cases the message shows unrecognisable (“higher”) characters, thus improving the clarity of the instructions.

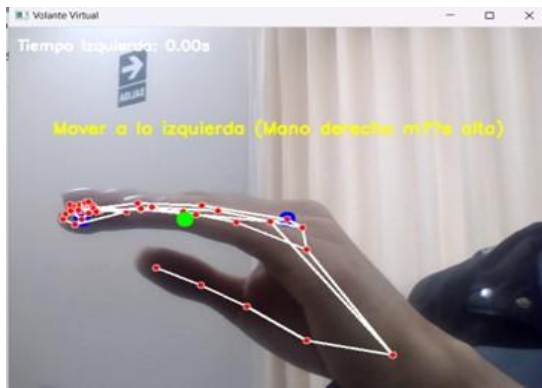


Fig. 26. Hand movement to the left (a).

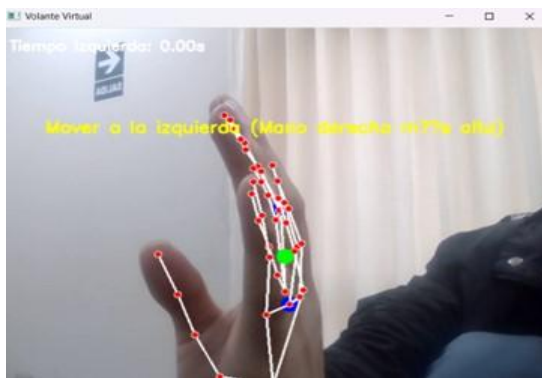


Fig. 27. Hand movement to the left (b).



Fig. 28. Hand movement to the left (c).

The system was able to detect the hand movements [20] in real time [25] and quickly reflect the changes in position by the variations in the positions of the points between the two images, where the hand is seen to rise and change orientation, resulting in an immediate adjustment in the structure detected by the system.

In the case of the multi-user server developed with Node, the tests showed that for a range of 2 to 4 players connected from PC or mobile it is feasible to have a competition with few moments of instability in the connection, which suggests that the optimisation of the secondary threads of each client should be deepened so that the system supports a greater number of users. Table I shows the technical aspects of the multi-user server.

TABLE I. TECHNICAL ASPECTS OF THE MULTI-USER SERVER

Aspect	Technical Details	Benefits	Limitations	Performance Metrics
Server Hardware	Lenovo with Core i5/i7 processor, 12GB RAM, integrated card or NVIDIA graphics.	Good performance on standard test hardware.	Performance may not be representative for lower hardware.	50-60% CPU usage during 4-player testing.
Server Operating System	Windows 11 in both cases (Core i5/i7).	Compatibility with modern systems.	Linux servers may offer better performance.	Response time: ~80-120 ms under ideal conditions.
Player Devices	Android devices (version 12 or higher) and PCs (Linux or Windows).	Stable connections on Android 12 or higher devices.	Variability in connection quality depending on device.	Connection success rate: 98% on mobile devices.
Control interface (PC)	Python control interface for PC connection.	Efficient connection from Linux or Windows PCs.	Interface could be more complex for non-technical users.	Synchronisation time: 100-150 ms
Gesture recognition	Gesture recognition performed with specific cameras and algorithms.	High accuracy in Gesture Recognition with suitable lighting conditions.	Success rate decreases with poor lighting or fast movement.	Success rate: 85-90% with ideal conditions. Failure rate: 15% in low light.
Latency and Response	Tests conducted in controlled environment with low latency local network.	Low latency under controlled conditions.	Latency increases with more players connected simultaneously.	Average latency: ~120 ms in local network. Average latency with 4 players: ~200 ms.

Finally, [7] uses gestural interaction techniques to control a video, as from [4] in hand gesture recognition, in the present work the developed gesture control interface has proved to be an inclusive tool, allowing players with disabilities to use their hands or head movements to control the directional keys in any multiplayer game in Unity. This solution facilitates the participation of all players, regardless of their physical abilities, promoting accessibility in digital entertainment. A sample of the code worked on the system can be seen in Fig. 29.

```
if face_results.detections:
    for detection in face_results.detections:
        bbox = detection.location_data.relative_bounding_box
        left = int(bbox.xmin * width)
        top = int(bbox.ymin * height)
        right = left + int(bbox.width * width)
        bottom = top + int(bbox.height * height)
        cv2.rectangle(frame, (left, top), (right, bottom), (0, 255, 0), 2)

    current_distance = bbox.width * bbox.height
    if last_distance is not None:
        if current_distance > last_distance * (1 + distance_threshold):
            if up_method == "Cabeza":
                keyboard.press(Key.up)
                keyboard.release(Key.down)
                last_action = "Acercando"
                action_text = "Acercando (Presionando tecla arriba)"
            elif current_distance < last_distance * (1 - distance_threshold):
                if down_method == "Cabeza":
                    keyboard.press(Key.down)
                    keyboard.release(Key.up)
                    last_action = "Alejando"
                    action_text = "Alejando (Presionando tecla abajo)"
            else:
                keyboard.release(Key.up)
                keyboard.release(Key.down)
        last_distance = current_distance
```

Fig. 29. Face detection.

VI. CONCLUSIONS

A hand gesture recognition system based on artificial vision [40] and real-time pattern detection was designed and implemented, achieving an innovative alternative to the use of directional keys in video games. This approach improves player immersion and makes the gaming experience more accessible, especially for users with motor disabilities.

The results demonstrated a high accuracy and speed of response to gestures ensuring a smooth and ideal interaction. Furthermore, customisation options were implemented in the interface, such as sensitivity, distance threshold and widget layout, contributing to its usability and versatility.

An efficient client-server architecture was developed that supports multiple simultaneous users, ensuring real-time communication and offering flexibility for individual network configurations and character selection within the game. The design uses threads and resources such as the camera responsibly, improving stability and minimising conflicts during interface use, contributing to increased performance and synchronisation.

The video game developed in Unity was adapted to run on PC platforms and mobile devices, extending its reach and allowing players to enjoy a consistent and immersive experience regardless of the device used.

The integration of traditional controls together with the gesture-based interface ensures greater inclusion of different play styles and user preferences.

In the comparative evaluation of the gesture control system against traditional methods, advantages in innovation and immersive experience were evident. While traditional controls maintain an advantage in accuracy and reliability under less optimal technical conditions, the gesture interface proved to be an intuitive and accessible solution, capable of improving user satisfaction by adapting to their preferences and needs.

FUTURE WORKS

As a result of the present work, a prospective vision for future work can be gained:

Optimisation of gesture recognition algorithms in variable lighting environments: It is essential to improve the accuracy of gesture recognition systems in changing lighting conditions, ensuring a consistent and reliable user experience.

Application of self-supervised learning in human-computer interaction systems: Implementing self-supervised learning techniques can improve the adaptability and efficiency of gesture interfaces, allowing systems to learn and adjust to individual user preferences and behaviours.

Integration of gesture recognition into augmented and virtual reality interfaces: Combining gesture recognition technologies with augmented and virtual reality environments can offer more immersive and natural gaming experiences, improving user interaction with digital content.

Development of deep learning models for the identification of complex gestures: The use of deep neural networks can facilitate the detection and classification of more sophisticated gestures, expanding the repertoire of available commands and improving interaction in multiplayer games.

Implementing gesture recognition using radar technology for mobile applications: The use of radar sensors in mobile devices can enable accurate gesture recognition without relying on cameras, offering an efficient and less invasive alternative for video game control.

ACKNOWLEDGMENT

Thanks to the Universidad Nacional de San Agustín de Arequipa for the support it provides for the development and implementation of proposals that benefit the continuous improvement of student performance through proposals that help to solve society's problems.

REFERENCES

- [1] M. Kassim, Y. San and R. Norlis, "Hand Gesture Recognition System using Image Processing," 2021 IEEE 17th International Colloquium on Signal Processing & Its Applications (CSPA), Selangor, Malaysia, 2021, pp. 57-62, doi: 10.1109/CSPA52141.2021.9377292.
- [2] T. Häckel, C. Eppner, and R. Stolkin, "Self-Supervised Learning of Hand-Eye Coordination for Robotic Grasping," IEEE Robotics and Automation Letters, vol. 6, no. 2, pp. 2648-2655, April 2021, doi: 10.1109/LRA.2021.3055843.
- [3] L. H. Chen, J. H. Wang, and M. T. Hsieh, "Real-Time Hand Gesture Recognition for Human-Computer Interaction," IEEE Transactions on Multimedia, vol. 23, pp. 226-235, Jan. 2021, doi: 10.1109/TMM.2020.2994535.
- [4] C. Li and M. Fu, "Real-Time Hand Gesture Detection and Recognition Based on Deep Learning," 2020 IEEE International Conference on Visual

- Communications and Image Processing (VCIP), Macau, China, 2020, pp. 1-4, doi: 10.1109/VCIP49819.2020.9301843.
- [5] H. R. Lee, J. Park, and Y.-J. Suh, "Improving Classification Accuracy of Hand Gesture Recognition Based on 60 GHz FMCW Radar with Deep Learning Domain Adaptation," *Electronics*, vol. 9, no. 12, p. 2140, Dec. 2020. DOI: 10.3390/electronics9122140.
- [6] B. P. S. Ahluwalia y R. Wason, "Gestural Interface Interaction: A Methodical Review," *International Journal of Computer Applications*, vol. 60, no. 1, pp. 21, Dec. 2012.
- [7] C. Peng, L. Cao, J. T. Hansberger, and V. A. Shanthakumar, "Hand gesture controls for image categorization in immersive virtual environments," 2017 IEEE Virtual Reality (VR), Los Angeles, CA, USA, 2017, pp. 18-22, doi: 10.1109/VR.2017.7892237.
- [8] F. W. Simor, M. R. Brum, J. D. E. Schmidt, R. Rieder, and A. C. B. De Marchi, "Usability evaluation methods for gesture-based games: A systematic review," *JMIR Serious Games*, vol. 4, no. 2, p. e17, 2016, doi: 10.2196/games.5860.
- [9] L. Chen, F. Wang, H. Deng and K. Ji, "A Survey on Hand Gesture Recognition," 2013 International Conference on Computer Sciences and Applications, Wuhan, China, 2013, pp. 313-316, doi: 10.1109/CSA.2013.79. keywords: {Gesture recognition;Computers;Thumb;Cameras;Human computer interaction;Robots;Human-Computer Interaction (HCI);Hand Gesture Recognition;Kinect},
- [10] A. S. Mohamed, N. F. Hassan, and A. S. Jamil, "Real-Time Hand Gesture Recognition: A Comprehensive Review of Techniques, Applications, and Challenges," *Cybern. Inf. Technol.*, vol. 24, no. 3, pp. 163–181, Sep. 2024, doi: 10.2478/cait-2024-0031
- [11] B. J. Jo, S.-K. Kim, and S. Kim, "Enhancing Virtual and Augmented Reality Interactions with a MediaPipe-Based Hand Gesture Recognition User Interface," *Ingénierie des Systèmes d'Information*, vol. 28, no. 3, pp. 311–318, 2023, doi: 10.18280/isi.280311.
- [12] C. Li, Q. Wu, and S. Gao, "Deep learning models for real-time gesture recognition in interactive applications," *Pattern Recognition*, vol. 131, pp. 108–114, 2023
- [13] J. Pirker, M. Pojer, A. Holzinger, and C. Gütl, "Gesture-Based Interactions in Video Games with the Leap Motion Controller," in *Human-Computer Interaction. User Interface Design, Development and Multimodality (HCI 2017)*, Lecture Notes in Computer Science, vol. 10271, Springer, pp. 620–633, May 2017.
- [14] M. L. Amit, A. C. Fajardo and R. P. Medina, "Recognition of Real-Time Hand Gestures using MediaPipe Holistic Model and LSTM with MLP Architecture," 2022 IEEE 10th Conference on Systems, Process & Control (ICSPC), Malacca, Malaysia, 2022, pp. 292-295, doi: 10.1109/ICSPC55597.2022.10001800.
- [15] Y. Zhu and B. Yuan, "Real-time hand gesture recognition with Kinect for playing racing video games," 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, China, 2014, pp. 3240-3246, doi: 10.1109/IJCNN.2014.6889481.
- [16] J. Xu, H. Wang, J. Zhang, and L. Cai, "Robust Hand Gesture Recognition Based on RGB-D Data for Natural Human-Computer Interaction," *IEEE Access*, vol. 10, pp. 46123-46133, 2022, doi: 10.1109/ACCESS.2022.3176717.
- [17] S. S. Rautaray and A. Agrawal, "Real-Time Hand Gesture Recognition System for Dynamic Applications," *International Journal of UbiComp (IJU)*, vol. 3, no. 1, pp. 21–31, Jan. 2012. doi: 10.5121/iju.2012.3103.
- [18] A. S. Khalaf, S. A. Alharthi, I. Dolgov, and P. O. Toups Dugas, "A Comparative Study of Hand Gesture Recognition Devices in the Context of Game Design," *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces, Daejeon, Republic of Korea*, 2019, pp. 397-402, doi: 10.1145/3343055.3360758.
- [19] J. Liu and M. Kavakli, "A survey of speech-hand gesture recognition for the development of multimodal interfaces in computer games," 2010 IEEE International Conference on Multimedia and Expo, Singapore, 2010, pp. 1564-1569, doi: 10.1109/ICME.2010.5583252.
- [20] J. Warchocki, M. Vlasenko, and Y. B. Eisma, "GRLib: An Open-Source Hand Gesture Detection and Recognition Python Library," *arXiv preprint arXiv:2310.12476*, 2023, [Online]. Available: <https://arxiv.org/abs/2310.12476>.
- [21] Y. Li, J. Huang, F. Tian, H.-A. Wang, and G.-Z. Dai, "Gesture interaction in virtual reality," *Virtual Reality & Intelligent Hardware*, vol. 1, no. 1, pp. 84-112, 2019, doi: 10.3724/SP.J.2096-5796.2018.0006.
- [22] K. Kondo, G. Mizuno, and Y. Nakamura, "Feedback Control Model of a Gesture-Based Pointing Interface for a Large Display," *IEICE Transactions on Information and Systems*, vol. E101-D, no. 7, pp. 1894-1905, Jul. 2018, doi: 10.1587/transinf.2017EDP7298.
- [23] S. Spanogianopoulos, K. Sirlantzis, M. Mentzelopoulos and A. Protopsaltis, "Human computer interaction using gestures for mobile devices and serious games: A review," 2014 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL2014), Thessaloniki, Greece, 2014, pp. 310-314, doi: 10.1109/IMCTL.2014.7011154.
- [24] D. Avola, L. Cinque, A. Fagioli, G. L. Foresti, A. Fragomeni, and D. Pannone, "3D hand pose and shape estimation from RGB images for keypoint-based hand gesture recognition," *Pattern Recognition*, vol. 129, p. 108762, 2022, doi: 10.1016/j.patcog.2022.108762.
- [25] O. Köpüklü, A. Gunduz, N. Kose and G. Rigoll, "Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks," 2019 14th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2019), Lille, France, 2019, pp. 1-8, doi: 10.1109/FG.2019.8756576.
- [26] Y. Yaseen, O.-J. Kwon, J. Kim, F. Ullah, J. Lee, and S. Jamil, "Comparative Analysis of Hand Gesture Datasets for Drone Control Using MediaPipe," *SSRN Electronic Journal*, pp. 1–24, Jun. 2024. DOI: 10.2139/ssrn.12345678.
- [27] J.-O. Kim, M. Kim, and K.-H. Yoo, "Real-Time Hand Gesture-Based Interaction with Objects in 3D Virtual Environments," in *Proceedings of the Digital Informatics and Convergence Symposium, Chungbuk National University, South Korea*, pp. 1–7, 2024.
- [28] D. Bachmann, F. Weichert, and G. Rinkenauer, "Review of Three-Dimensional Human-Computer Interaction with Focus on the Leap Motion Controller," *Sensors*, vol. 18, no. 7, p. 2194, Jul. 2018. DOI: 10.3390/s18072194.
- [29] S. S. Rautaray and A. Agrawal, "Interaction with virtual game through hand gesture recognition," 2011 International Conference on Multimedia, Signal Processing and Communication Technologies, Aligarh, India, 2011, pp. 244-247, doi: 10.1109/MSPCT.2011.6150485.
- [30] A. Safa et al., "Improving the Accuracy of Spiking Neural Networks for Radar Gesture Recognition Through Preprocessing," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 6, pp. 2869-2881, June 2023, doi: 10.1109/TNNLS.2021.3109958.
- [31] S. D. Bharatula, U. R. Vadhegar and M. Maiti, "GestureFlow: A Novel Hand Gesture Control System for Interactive Gaming," 2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT), Kamand, India, 2024, pp. 1-6, doi: 10.1109/ICCCNT61001.2024.10724912.
- [32] A. Gupta, N. Chawla, R. Jain, N. Thakur, and A. Devi, "Gesture-Based Touchless Operations: Leveraging MediaPipe and OpenCV," *NEU Journal for Artificial Intelligence and Internet of Things*, vol. 1, no. 2, pp. 1-10, Oct. 2023.
- [33] P. S. G. Deena, H. D. A. K. B. and H. S., "Gaming using different hand gestures using artificial neural network", *EAI Endorsed Trans IoT*, vol. 10, Feb. 2024.
- [34] M. R. Islam, R. Rahman, A. Ahmed, and R. Jany, "NFS: A Hand Gesture Recognition Based Game Using MediaPipe and PyGame," *Islamic University of Technology, Gazipur, Dhaka, Bangladesh*, 2022.
- [35] U. Patel, S. Rupani, V. Saini and X. Tan, "Gesture Recognition Using MediaPipe for Online Realtime Gameplay," 2022 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Niagara Falls, ON, Canada, 2022, pp. 223-229, doi: 10.1109/WI-IAT55865.2022.00039.
- [36] A. Sharma, Simran, L. Verma, H. Kaur, A. Modgil and A. Soni, "Hand Gesture Recognition Gaming Control System: Harnessing Hand Gestures and Voice Commands for Immersive Gameplay," 2024 International Conference on Emerging Innovations and Advanced Computing (INNOCOMP), Sonipat, India, 2024, pp. 101-107, doi: 10.1109/INNOCOMP63224.2024.00026.

- [37] E. Sophiya and S. S. Reddy, "Hand Gesture-Driven Gaming for Effective Rehabilitation and Improved Quality of Life - A Review," 2024 5th International Conference on Innovative Trends in Information Technology (ICITIT), Kottayam, India, 2024, pp. 1-6, doi: 10.1109/ICITIT61487.2024.10580667.
- [38] S. Metkar, J. Mahajan, J. Adsul and B. Chavan, "Human body gesture-controlled gaming application," 2022 Second International Conference on Next Generation Intelligent Systems (ICNGIS), Kottayam, India, 2022, pp. 1-6, doi: 10.1109/ICNGIS54955.2022.10079850.
- [39] D. Jatain, S. Singh, N. Jatana, G. Sharma, V. Garg and M. Niranjanamurthy, "A Real-Time Camera-based Motion Sensing Game Tool for Cervical Rehabilitation," 2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS), Chikkaballapur, India, 2024, pp. 1-8, doi: 10.1109/ICKECS61492.2024.10617271.
- [40] C. Yeh, W. -C. Shen, C. -W. Ma, Q. -T. Yeh, C. -W. Kuo and J. -S. Chen, "Real-time Human Movement Recognition and Interaction in Virtual Fitness using Image Recognition and Motion Analysis," 2023 12th International Conference on Awareness Science and Technology (iCAST), Taichung, Taiwan, 2023, pp. 242-246, doi: 10.1109/iCAST57874.2023.10359266.

Teaching Programming in Higher Education: Analyzing Trends, Technologies, and Pedagogical Approaches Through a Bibliometric Lens

Mariuxi Vinueza-Morales¹, Jorge Rodas-Silva², Cristian Vidal-Silva^{*3}

Faculty of Sciences and Engineering, Universidad Estatal de Milagro, Milagro, Ecuador¹

Director of Innovation in Academic Processes, SofTech Research Group, Universidad Estatal de Milagro, Milagro, Ecuador²
Facultad de Ingeniería y Negocios, Universidad de Las Américas, Manuel Montt 948, Providencia, 7500975, Santiago, Chile³

Abstract—In today’s information society, developing programming competencies is essential in higher education. Numerous studies have been conducted on effective strategies for fostering these skills. This study performs a bibliometric analysis of research on teaching strategies for programming in higher education, using data from the SCOPUS and Web of Science (WOS) databases between 2014 and 2023. The analysis identifies key trends, influential authors, and collaboration networks in this field. The most effective teaching strategies include project-based learning, flipped classrooms, and collaborative programming. Emerging technologies such as augmented reality and virtual reality are gaining prominence in programming education. Despite the growth of research in this area, challenges remain, such as the lack of longitudinal studies exploring the long-term impact of these methodologies and the need for greater geographic diversity in studies. This paper emphasizes the importance of exploring new technologies and interdisciplinary approaches and fostering international collaborations to enhance programming education. The findings guide researchers and educators on how to optimize programming learning in a global context.

Keywords—Programming; higher education; teaching strategies; bibliometrics

I. INTRODUCTION

In the digital era, programming has become an essential competency in the higher education curriculum [1], especially in Science, Technology, Engineering, and Mathematics (STEM) disciplines [2]. While programming experience in higher education was traditionally confined to technical or engineering fields, today, programming permeates diverse disciplines, recognizing its potential as both a technical tool and a critical thinking skill [3], [4]. Effective acquisition of programming skills requires not only familiarity with syntax and data structures but also deep logical understanding and problem-solving abilities [5].

Teaching and learning strategies for programming in higher education extend beyond simple knowledge transmission techniques. They instill a computational mindset, promote logical thinking, and help students address complex problems systematically [6]. As technology advances, these strategies must evolve to keep pace with the changing demands of the programming field and students’ needs, ensuring long-term educational outcomes [7]. For example, Bloom’s taxonomy classifies and describes different levels of learning achievement

that students can reach [8], [9]. Project-based learning (PBL) and pair programming also encourage collaboration and critical thinking [10], [11]. These strategies emphasize the practical application of knowledge and solving real-world problems, allowing students to consolidate and contextualize what they have learned in realistic settings [12].

As highlighted by Sun et al. [13], adaptability is essential in programming. Tools, languages, and methodologies continuously change and evolve [14]. Therefore, teaching strategies must not only transmit technical knowledge, but also teach students to become autonomous and adaptable learners [15]. This implies fostering skills such as self-learning, curiosity for new technologies, and resilience to overcome challenges [16]. Adaptive learning would cultivate logical and creative thinkers who can innovate and adapt in a constantly changing field like programming education [17]. The extensive literature on programming teaching and learning strategies reflects the growing importance and recognition of these strategies in both educational and professional realms [18]. Researchers, educators, and professionals worldwide have contributed numerous studies, theories, and methodologies, enriching the body of available knowledge [19], [20], [21], [22]. However, the vast amount of information on diverse programming education approaches, along with the fast pace of new developments, presents challenges in staying up to date and identifying the most impactful trends and practices.

The rapid evolution of programming education and the increasing volume of research publications pose a challenge in identifying effective teaching and learning strategies [23]. This research explores publication patterns, the most cited sources, academic collaboration networks, and other relevant aspects to shed light on the current state of research in the programming education at the higher education level. Thus, this document seeks to address the need for a comprehensive understanding of current trends, significant contributors, and prominent research in higher education programming education. Furthermore, our work aims to recognize the most influential authors, leading research centers, and areas that require more attention in programming education.

This study predominantly analyzes data from specific geographic regions, which can limit its generalizability to other educational contexts, particularly in underrepresented or developing countries. Future research should incorporate data from a broader range of regions to provide more globally

*Corresponding authors.

representative insights. This analysis relies exclusively on data from SCOPUS and Web of Science. Although these databases provide extensive coverage of high-quality research, including additional sources, such as local publications or databases not indexed on these platforms, could enhance the study's comprehensiveness.

A. Research Questions

This project is a quasi-experimental quantitative study designed to answer the following research questions:

- RQ1: What are the predominant trends in programming teaching and learning strategies in higher education? This article addresses this question by analyzing research from SCOPUS and WOS databases published between 2014 and 2023, highlighting frequently cited approaches such as project-based learning, flipped classrooms, and collaborative programming. The analysis focuses on the evolution of these strategies and their impact on programming competency development, acknowledging that relevant works in non-indexed conferences and journals may further complement the identified trend.
- RQ2: Who are the most influential authors, and what are the seminal publications shaping the field of programming education? This article responds by identifying the most cited authors and works from SCOPUS and WOS between 2014 and 2023. The analysis considers author networks, citation impact, and recurring themes in key publications, recognizing that other influential contributors may also emerge from non-indexed sources.
- RQ3: What journals and institutions make the most significant contributions to programming education research? This article explores this question by examining the journals and institutions with the highest SCOPUS and WOS production and citation metrics between 2014 and 2023. The review highlights institutions consistently contributing to shaping discourse on programming education, acknowledging valuable works published in non-indexed platforms.

II. BACKGROUND

Computational thinking plays a critical role in the modern digital era, providing individuals with essential problem-solving skills that transcend the boundaries of computer science and programming, as noted by Lu et al. [24]. Based on principles from computer science, mathematics, and logic [25], this approach enables individuals to break down complex problems, recognize patterns, and design algorithmic solutions [26]. As highlighted by Shen et al. [27], computational thinking significantly influences daily life by helping individuals make informed decisions and solve problems efficiently. Whether optimizing daily routines or critically evaluating online information, this skill set empowers individuals to navigate the complexities of the digital world [28]. Furthermore, when incorporated into educational curricula, it fosters essential competencies such as logical reasoning and creativity, preparing students for future challenges [29]. Algorithm 1 presents

an algorithm describing the steps to master new topics through computational thinking [30].

Algorithm 1 Procedure for Mastering a New Topic

Require: Topic is identified

Ensure: Relevant materials are gathered

Obtain an initial understanding

Define the boundaries of the topic

Search for appropriate resources

Develop a structured learning approach

Set criteria for successful learning

while *learning not achieved* **do**

Refine the selected resources

Reevaluate and explore the materials

Experiment with the information

Implement newly acquired knowledge

▷ If feasible

Share or explain learned concepts

▷ If feasible

end while

Beyond everyday applications, the benefits of computational thinking extend to a wide range of disciplines [25], [31]. Shin et al. [32] demonstrate how this competency enhances scientific research by helping scientists analyze complex data, simulate experiments, and develop better models to understand the natural world. Computational thinking supports research in fields such as biology, physics, and social sciences, improving decision-making by providing powerful tools for data-driven insights [33]. With the growing demand for digital literacy in the workforce, computational thinking equips individuals with the necessary tools to thrive in an evolving job market [34]. This competency empowers individuals not only as students or professionals but also as active participants in a technology-driven society.

A. Programming Competencies in Higher Education

Programming competencies hold critical importance in higher education, particularly within computer science and across Science, Technology, Engineering, Arts, and Mathematics (STEAM) disciplines [35]. The ability to think algorithmically, solve complex problems systematically, and develop automated solutions through coding has become an indispensable skill set for students, regardless of their field of study [36]. Developing programming competencies in higher education equips students with fundamental skills that go beyond coding:

1) *Algorithmic thinking*: This allows students to break down intricate problems into smaller, manageable tasks while constructing logical sequences of steps to address them [24].

2) *Problem-solving through programming*: Programming fosters creativity and resilience, pushing students to iteratively refine their solutions until achieving the most effective outcome [37].

3) *Abstract thinking*: Students can conceptualize real-world problems as abstract models, facilitating a deeper understanding of complex phenomena across various disciplines [38].

4) *Automation and efficiency*: Programming enables students to streamline repetitive tasks, enhancing both productivity and efficiency [39].

The impact of programming competencies transcends computer science, offering substantial benefits in STEAM disciplines [35], [40]. These skills contribute to both the technical and creative dimensions of each field.

- **Science:** Programming is a powerful tool for analyzing large datasets, modeling complex systems, and simulating experiments. In fields like biology, physics, and chemistry, the ability to automate data processing and perform statistical analyses enhances the precision and speed of scientific discoveries.
- **Technology:** Programming drives a deeper understanding of technological systems, empowering students to innovate and develop new software tools.
- **Engineering:** Programming is indispensable in engineering, whether used to model and simulate physical systems or optimize design processes. Coding equips students with tools that improve accuracy and efficiency in the mechanical, civil and electrical engineering disciplines.
- **Art:** In the arts, programming opens new avenues for digital creativity.
- **Mathematics:** Programming enhances mathematical problem-solving by allowing the simulation of mathematical models and solving large-scale calculations that would otherwise be impossible through manual methods.

III. METHODOLOGY AND RESEARCH DESIGN

The increasing volume of academic publications and the proliferation of research streams can make it challenging for researchers to stay updated within a specific field. Systematic literature reviews synthesize available scientific information and help identify areas of uncertainty where further investigation is required [41]. Typically, literature review research addresses a single scientific database, which limits the ability to gain a comprehensive view of knowledge and trends within a specific domain. Therefore, some authors argue for the necessity of using multiple databases [42]. In this context, the data for the present study were extracted from the Web of Science (WOS) and SCOPUS databases.

Using both WOS and SCOPUS for the bibliometric analysis of teaching and learning strategies for higher education programming ensures a comprehensive and high quality coverage of the relevant academic literature. Both databases are renowned for their extensive indexes of peer-reviewed journals, conference proceedings, and other scholarly outputs across disciplines, including computer science and education [43]. The international scope of these databases [44] ensures a global perspective on teaching methodologies, crucial for understanding the varied approaches to programming education in higher education. Both WOS and SCOPUS have significant prestige and are recognized for improving the robustness of analysis by cross-referencing data, ensuring completeness, and minimizing potential bias in the literature review [45]. Thus, the choice of SCOPUS and WOS allows for a comprehensive examination of publication patterns in the field of programming education. To achieve this, records from both sources were merged into a single dataset. Table I summarizes the search criteria, while

Fig. 1 shows the publication trends over the period of time (2014-2023).

In terms of coverage, WOS encompasses a broader range than SCOPUS, with 1,464 records compared to 361. After comparing and removing duplicates, the process identified 1,697 documents related to teaching and learning strategies in programming education, revealing that approximately 11% of the documents overlap or share similarities.

A. Evaluation Instruments

As highlighted by Chen et al. [46] and Trinidad et al. [47], analyzing publication trends provides researchers with a deeper understanding of the scientific landscape, allowing them to identify emerging knowledge areas and contribute to the advancement of their fields. To support this type of analysis, the bibliometric study used the Bibliometrix platform using the R programming language. Bibliometrix is an open-source toolset that offers flexibility by integrating with various statistical packages. This adaptability makes Bibliometrix particularly valuable for exploring evolving research areas and uncovering new trends in programming education strategies.

One of the key strengths of Bibliometrix is its free access and ease of use through a web-based interface, which encapsulates its core capabilities and establishes a framework for real-time data analysis [48]. For instance, Biblioshiny, a module of Bibliometrix, enables users to perform relevant bibliometric and visual analyses through an interactive web interface [49]. This tool facilitates identifying connections between changes in scientific production, citations, author collaborations, and other essential bibliometric indicators, especially in programming education. Consequently, Bibliometrix is a robust and accessible tool for the scientific community, providing an effective means to explore and evaluate academic literature.

IV. RESULTS

This section presents the results of the bibliometric analysis, highlighting the most relevant authors, important institutions, influential documents, and the keywords with the highest relevance in the analyzed articles.

A. Most Relevant Authors

The bibliometric analysis identifies the most prominent authors in the field. Table II presents a list of authors with notable contributions and the number of citations they have received. Among them, Li Yong, Liu Yonggang, Liu Yan, Liu Yuan, and Yong Wang stand out. These researchers were identified by comparing SCOPUS and WOS records using metrics such as the H-index and citation count, which measure scientific performance in the field [50]. The authors in Table II have made significant advancements in programming education through detailed studies on effective teaching and learning strategies. Their work includes innovative methodologies and practices with long-term impacts on students' learning outcomes.

By examining the H-index and citation count, we can evaluate the influence and impact of these researchers in promoting programming education. This analysis provides valuable insights into the key contributors driving the field and underscores the importance of developing effective teaching and learning strategies.

TABLE I. NON-PHARMACOLOGICAL INTERVENTION STRATEGIES BASED ON CONTROLLED EXERCISE AND NUTRITIONAL EDUCATION

Criterion	Values
Time Span	2014 – 2023
Date of Query	June 2023
Document Types	Journal articles and Conference proceedings
Journal and Conference Types	Any type
Search Fields	Title, Abstract, and Keywords
Search Terms	Programming AND Universities AND Strategies AND (Learning AND Teaching) – in English
Records	WOS: 1464; SCOPUS: 361
Total Records	1,697

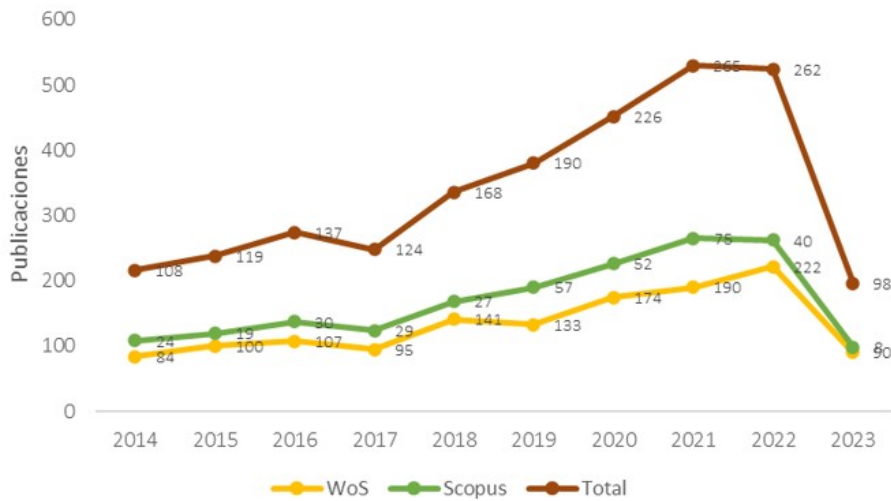


Fig. 1. Total number of publications in WOS and SCOPUS, 2015–2023.

TABLE II. TOP AUTHORS IN PROGRAMMING EDUCATION RESEARCH IN HIGHER EDUCATION

Authors	WOS Citations	WOS H-index	SCOPUS Citations	SCOPUS H-index	Publications
Li Yong	2058	23	7591	42	11
Liu Yonggang	29	3	682	15	11
Liu Yan	2072	22	9997	50	11
Liu Yuan	1151	18	938	16	11
Yong Wang	2124	26	8192	50	11

B. Key Institutions

The bibliometric review identified the most prominent institutions in the field based on their significant publication output. Table III lists the top ten universities, their country of origin, and the number of related publications. These results corroborate the findings of Apiola et al. [51] and Perez and Garcia [52], highlighting that universities play a crucial role in publishing research and disseminating experiences related to programming education in higher education. Their contributions emphasize the importance of integrating strategies that help students develop critical thinking and problem-solving skills, ultimately improving their academic and professional development.

C. Influential Documents

Identifying key documents begins by determining how each document connects to internal research references (endogenous references) and external sources (exogenous references) found in academic databases such as SCOPUS and WOS. Duque and Duque Oliva [53] explain that the average number of citations

TABLE III. TOP INSTITUTIONS CONTRIBUTING TO PROGRAMMING EDUCATIONAL RESEARCH

Institution	Publications	Country
University of California, San Francisco	74	USA
University of Toronto	69	Canada
University of Colorado Boulder	59	USA
University of Michigan	59	USA
University of Sydney	43	Australia
University of Calgary	36	Canada
McGill University	35	Canada
University of Pennsylvania	35	USA
National University of Singapore	34	Singapore
University of Minnesota	34	USA

is calculated by dividing internal references by the time elapsed since the document's initial publication. This approach offers a method to assess the influence and acceptance of research over time.

When analyzing influential documents, our findings show that the study by Sung et al. [54] has the highest number of citations, with a total of 620 and an average of 77.5 per year. McLaughlin et al. [55] follows closely with 603

citations and an average of 60.3 per year, while Bers et al. [56] has a total of 378 citations and an annual average of 37.8. These three authors stand out due to the high citation counts of their respective articles, which focus on topics related to programming education strategies. Table IV lists top cited documents in programming education research.

D. Keywords with the Highest Relevance

During the analysis, we extracted the terminologies that had the most significant impact across the reviewed documents. Using the Biblioshiny tool, we created a graphical representation or “word cloud” (Fig. 2) to visualize the most prominent terms identified in this study. As explained by Alsalem et al. [57], this method allows scholars to compare different sections and visually identify the most significant terms, highlighted in bold. The word cloud emphasizes key terms such as “learning”, “students”, “education”, “teaching”, and “programming”, which are closely aligned with the study’s focus on programming education for undergraduate students. This approach explores the interaction between pedagogical approaches, technology, and programming in educational settings while analyzing how these elements influence students’ ability to acquire programming skills.

Using a TreeMap further illustrates the clustering of potential keywords in the research articles, as discussed by Secinaro et al. [58]. Table V presents the top ten keywords by frequency and percentage of appearance in the analyzed documents. The analysis shows that 41% of the most relevant terms include words such as “education”, “students”, “teaching”, “learning”, and “programming”, which underscores the importance of this research topic within the field of scientific publications.

V. DISCUSSION

This study provides a comprehensive overview of current trends in teaching and learning programming through a bibliometric analysis of the SCOPUS and WOS databases. The results highlight several key areas that require further exploration and themes that have been consistently addressed in existing research.

A. Current Trends

One of the main trends identified in the analysis is the emphasis on project-based learning and flipped classrooms. These strategies have proven to be effective in enhancing problem-solving skills and fostering greater engagement by allowing students to apply learned concepts to real-world problems [54]. In addition, there has been a growing adoption of collaborative programming techniques such as pair programming, which improve code quality and overall student performance [55]. Regarding emerging topics, there is increasing interest in using immersive and simulation technologies such as augmented reality and virtual reality to teach programming [56].

B. Limitations of Current Research

Despite advances in programming education, several challenges and limitations persist. One of the key limitations identified is the lack of longitudinal studies that measure the long-term impact of different programming teaching strategies.

Most of the analyzed studies focus on short-term outcomes, such as grades or students’ performance in individual courses, but more research is needed to explore how these strategies influence long-term professional development and career outcomes [59]. Another limitation is the lack of diversity in the educational contexts studied. Most of the research has been conducted in developed countries, particularly in the United States and China, which may not reflect the educational realities of other regions. Furthermore, more research is needed on programming education in non-formal settings and self-learning environments, as many students learn programming independently or through online platforms outside traditional classrooms [60].

C. Implications for Future Research

This study provides several implications for future research. First, more research is needed on the use of emerging technologies such as artificial intelligence and machine learning in programming education. These technologies have the potential to personalize teaching and provide immediate feedback to students, which could significantly improve learning outcomes [61]. Second, there is a need to further explore interdisciplinary approaches to teaching programming. Integrating programming with other disciplines has proven to be effective in improving student understanding and motivation, but more research is needed to understand how these approaches can be optimally designed and implemented [62]. In addition, more research is required on the barriers students face when learning to program, particularly in disadvantaged contexts. Programming can be a difficult skill to acquire, and many students struggle with abstract concepts and the complex syntax of programming languages. Understanding these barriers and how to overcome them is crucial to ensuring that all students have the opportunity to develop programming skills [63].

D. Threats to Validity

As with any study, this bibliometric analysis has limitations that may affect the validity of the results. The following threats to validity should be considered:

- Selection bias: The use of the SCOPUS and Web of Science databases, although they include a large volume of high-quality research, may have excluded important studies from other databases not considered in this analysis. This raises the possibility of selection bias, as some relevant studies published in non-indexed academic journals or platforms may not have been included.
- Temporal bias: The analysis focused on publications from 2014 to 2023, which means that any research conducted before this period was excluded. Although this time range captures the most recent trends, it may omit foundational studies or pioneering approaches in programming education that still influence the field. This temporal bias could limit the understanding of the complete evolution of programming teaching methodologies over time.
- Variability in bibliometric indicators: The quality and impact of a research paper were evaluated using bibliometric metrics such as the citation count and the

field has grown significantly, disparities remain in terms of international collaboration and access to resources. Developing countries could benefit from greater support and collaboration with institutions from more developed nations to ensure that advances in programming education are accessible to all.

In terms of contribution, this article not only identifies the key trends and most influential contributors, but also provides a roadmap for future research in programming education. We hope that the findings of this study will assist educators, researchers, and policymakers in developing more effective and equitable approaches to programming education in higher education.

REFERENCES

- [1] V. Basilotta-Gómez-Pablos, M. Matarranz, L.-A. Casado-Aranda, and A. Otto, "Teachers' digital competencies in higher education: a systematic literature review," *International Journal of Educational Technology in Higher Education*, vol. 19, no. 1, p. 8, Feb 2022. [Online]. Available: <https://doi.org/10.1186/s41239-021-00312-8>
- [2] P. Abichandani, V. Sivakumar, D. Lobo, C. Iaboni, and P. Shekhar, "Internet-of-things curriculum, pedagogy, and assessment for stem education: A review of literature," *IEEE Access*, vol. 10, pp. 38 351–38 369, 2022.
- [3] M. González-Sanmamed, A. Sangrá, A. Souto-Seijo, and I. Estévez Blanco, "Ecologías de aprendizaje en la era digital: desafíos para la educación superior," *PUBLICACIONES*, vol. 48, no. 1, pp. 25–45, 2018.
- [4] J. Jiménez-Toledo, C. Collazos, and O. Revelo-Sánchez, "Consideraciones en los procesos de enseñanza-aprendizaje para un primer curso de programación de computadores: una revisión sistemática de la literatura," *Tecnológicas*, vol. 22, pp. 83–117, 2019.
- [5] Y.-T. Lin, M. K.-C. Yeh, and S.-R. Tan, "Teaching programming by revealing thinking process: Watching experts' live coding videos with reflection annotations," *IEEE Transactions on Education*, vol. 65, no. 4, pp. 617–627, 2022.
- [6] P. Compañ-Rosique, R. Satorre-Cuerda, F. Llorens-Largo, and R. Molina-Carmona, "Enseñando a programar: un camino directo para desarrollar el pensamiento computacional," *Revista de Educación a Distancia (RED)*, vol. 46, 2015.
- [7] F. A. Adamopoulos, *Learning Programming, Student Motivation*. Cham: Springer International Publishing, 2020, pp. 1058–1067. [Online]. Available: https://doi.org/10.1007/978-3-030-10576-1_182
- [8] S. Masapanta-Carrion and J. Velázquez-Iturbide, "A systematic review of the use of bloom's taxonomy in computer science education," in *SIGCSE '18*. Association for Computing Machinery, 2018, pp. 441–446.
- [9] B. Bloom, *Taxonomy of Educational Objectives*. Longman, 1956.
- [10] A. Younis, R. Sunderraman, M. Metzler, and A. Bourgeois, "Developing parallel programming and soft skills: A project-based learning approach," *Journal of Parallel and Distributed Computing*, vol. 158, pp. 151–163, 2021.
- [11] N. Shin, J. Bowers, J. Krajcik, and D. Damelin, "Promoting computational thinking through project-based learning," *Disciplinary and Interdisciplinary Science Education Research*, vol. 3, no. 1, p. 7, 2021.
- [12] A. Bawamohiddin and R. Razali, "Problem-based learning for programming education," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 7, p. 2035, 2017.
- [13] Q. Sun, J. Wu, and K. Liu, "Toward understanding students' learning performance in an object-oriented programming course: The perspective of program quality," *IEEE Access*, vol. 8, pp. 37 505–37 517, 2020.
- [14] E. Merelli, N. Paoletti, and L. Tesei, "Adaptability checking in complex systems," *Science of Computer Programming*, vol. 115–116, pp. 23–46, 2016.
- [15] Q. Cheng, D. Benton, and A. Quinn, "Building a motivating and autonomy environment to support adaptive learning," in *2021 IEEE Frontiers in Education Conference (FIE)*, 2021, pp. 1–7.
- [16] B. Vesin, K. Mangaroska, and M. Giannakos, "Learning in smart environments: user-centered design and analytics of an adaptive learning system," *Smart Learning Environments*, vol. 5, 2018.
- [17] J. Qadir, K.-L. A. Yau, M. Ali Imran, and A. Al-Fuqaha, "Engineering education, moving into 2020s : Essential competencies for effective 21st century electrical & computer engineers," in *2020 IEEE Frontiers in Education Conference (FIE)*, 2020, pp. 1–9.
- [18] M. Thuné and A. Eckerdal, "Analysis of students' learning of computer programming in a computer laboratory context," *European Journal of Engineering Education*, vol. 44, pp. 1–18, 2018.
- [19] B. Xie, D. Loksa, G. Nelson, M. Davidson, D. Dong, H. Kwik, A. Hui Tan, L. Hwa, M. Li, and A. Ko, "A theory of instruction for introductory programming skills," *Computer Science Education*, vol. 29, no. 2–3, pp. 205–253, 2019.
- [20] L. Silva, A. J. Mendes, and A. Gomes, "Computer-supported collaborative learning in programming education: A systematic literature review," in *2020 IEEE Global Engineering Education Conference (EDUCON)*, 2020, pp. 1086–1095.
- [21] G. Liargkovas, A. Papadopoulou, Z. Kotti, and D. Spinellis, "Software engineering education knowledge versus industrial needs," *IEEE Transactions on Education*, vol. 65, no. 3, pp. 419–427, 2022.
- [22] A. Yusuf and N. M. Noor, "Research trends on learning computer programming with program animation: A systematic mapping study," *Computer Applications in Engineering Education*, vol. 31, no. 6, pp. 1552–1582, 2023. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cae.22659>
- [23] C.-S. Cheah, "factors-contributing-to-the-difficulties-in-teaching-and-learning-of-computer-programming-a-literature-review," *Contemporary Educational Technology*, vol. 12, p. ep272, 05 2020.
- [24] C. Lu, R. Macdonald, B. Odell, V. Kokhan, C. Demmans Epp, and M. Cutumisu, "A scoping review of computational thinking assessments in higher education," *Journal of Computing in Higher Education*, vol. 34, no. 2, pp. 416–461, Aug 2022. [Online]. Available: <https://doi.org/10.1007/s12528-021-09305-y>
- [25] Y. Li, A. H. Schoenfeld, A. A. diSessa, A. C. Graesser, L. C. Benson, L. D. English, and R. A. Duschl, "Computational thinking is more about thinking than computing," pp. 1–18, 2020.
- [26] K. Srinivasa, M. Kurni, and K. Saritha, "Computational thinking," in *Learning, Teaching, and Assessment Methods for Contemporary Learners: Pedagogy for the Digital Generation*. Springer, 2022, pp. 117–146.
- [27] J. Shen, G. Chen, L. Barth-Cohen, S. Jiang, and M. Eltoukhy, "Connecting computational thinking in everyday reasoning and programming for elementary school students," *Journal of Research on Technology in Education*, vol. 54, no. 2, pp. 205–225, 2022.
- [28] K. Kanaki and M. Kalogiannakis, "Assessing algorithmic thinking skills in relation to age in early childhood stem education," *Education Sciences*, vol. 12, no. 6, p. 380, 2022.
- [29] K. Kwon, A. T. Ottenbreit-Leftwich, T. A. Brush, M. Jeon, and G. Yan, "Integration of problem-based learning in elementary computer science education: effects on computational thinking and attitudes," *Educational Technology Research and Development*, vol. 69, pp. 2761–2787, 2021.
- [30] C. Vidal-Silva, J. Cárdenas-Cobo, M. Tupac-Yupanqui, J. Serrano-Malebrán, and A. Sánchez Ortiz, "Developing programming competencies in school-students with block-based tools in chile, ecuador, and peru," *IEEE Access*, vol. 12, pp. 118 924–118 936, 2024.
- [31] R. P. Lai, "Beyond programming: A computer-based assessment of computational thinking competency," *ACM Transactions on Computing Education (TOCE)*, vol. 22, no. 2, pp. 1–27, 2021.
- [32] N. Shin, J. Bowers, S. Roderick, C. McIntyre, A. L. Stephens, E. Eidin, J. Krajcik, and D. Damelin, "A framework for supporting systems thinking and computational thinking through constructing models," *Instructional Science*, vol. 50, no. 6, pp. 933–960, 2022.
- [33] D. Helbing, S. Mahajan, R. H. Fricker, A. Musso, C. I. Hausladen, C. Carissimo, D. Carpentras, E. Stockinger, J. A. Sanchez-Vaquerizo, J. C. Yang *et al.*, "Democracy by design: Perspectives for digitally assisted, participatory upgrades of society," *Journal of Computational Science*, vol. 71, p. 102061, 2023.
- [34] A. Yadav and U. Berthelsen, *Computational Thinking in Education: A Pedagogical Perspective*. Routledge, 2021. [Online]. Available: <https://books.google.cl/books?id=2H9kzGEACAAJ>

- [35] J.-A. Marín-Marín, A.-J. Moreno-Guerrero, P. Dúo-Terrón, and J. López-Belmonte, "Steam in education: a bibliometric analysis of performance and co-words in web of science," *International Journal of STEM Education*, vol. 8, no. 1, p. 41, Jun 2021. [Online]. Available: <https://doi.org/10.1186/s40594-021-00296-x>
- [36] A. Melro, G. Tarling, T. Fujita, and J. K. Staarman, "What else can be learned when coding? a configurative literature review of learning opportunities through computational thinking," *Journal of Educational Computing Research*, vol. 61, no. 4, pp. 901–924, 2023. [Online]. Available: <https://doi.org/10.1177/07356331221133822>
- [37] Z. Ju, "Computational thinking through programming: a meta-analysis of collaborative versus solo problem solving," Masters by Research, Faculty of Arts and Social Sciences, Sydney School of Education and Social Work, University of Sydney, 2024. [Online]. Available: <https://hdl.handle.net/2123/32647>
- [38] Y. Qian and I. Choi, "Tracing the essence: ways to develop abstraction in computational thinking," *Educational technology research and development*, vol. 71, no. 3, pp. 1055–1078, Jun 2023. [Online]. Available: <https://doi.org/10.1007/s11423-022-10182-0>
- [39] N. Selwyn, T. Hillman, A. Bergviken-Rensfeldt, and C. Perrotta, "Making sense of the digital automation of education," *Postdigital Science and Education*, vol. 5, no. 1, pp. 1–14, Jan 2023. [Online]. Available: <https://doi.org/10.1007/s42438-022-00362-9>
- [40] P. Dúo-Terrón, "Analysis of scratch software in scientific production for 20 years: Programming in education to develop computational thinking and steam disciplines," *Education Sciences*, vol. 13, no. 4, 2023. [Online]. Available: <https://www.mdpi.com/2227-7102/13/4/404>
- [41] R. Briner and D. Denyer, "Systematic review and evidence synthesis as a practice and scholarship tool," pp. 112–129, 2012.
- [42] S. Echchakoui, "Why and how to merge scopus and web of science during bibliometric analysis: the case of sales force literature from 1912 to 2019," *Journal of Marketing Analytics*, vol. 8, 2020.
- [43] R. Prancuté, "Web of science (wos) and scopus: The titans of bibliographic information in today's academic world," *Publications*, vol. 9, no. 1, 2021. [Online]. Available: <https://www.mdpi.com/2304-6775/9/1/12>
- [44] A. Valente, M. Holanda, A. M. Mariano, R. Furuta, and D. Da Silva, "Analysis of academic databases for literature review in the computer science education field," in *2022 IEEE Frontiers in Education Conference (FIE)*, 2022, pp. 1–7.
- [45] J. Zhu and W. Liu, "A tale of two databases: the use of web of science and scopus in academic papers," *Scientometrics*, vol. 123, 2020.
- [46] X. Chen, D. Zou, H. Xie, and F. L. Wang, "Past, present, and future of smart learning: a topic-based bibliometric analysis," *International Journal of Educational Technology in Higher Education*, vol. 18, no. 1, p. 2, Jan 2021. [Online]. Available: <https://doi.org/10.1186/s41239-020-00239-6>
- [47] M. Trinidad, M. Ruiz, and A. Calderón, "A bibliometric analysis of gamification research," *IEEE Access*, vol. 9, pp. 46 505–46 544, 2021.
- [48] Z. Li, G. Wang, J. Lu, D. G. Broo, D. Kiritsis, and Y. Yan, "Bibliometric analysis of model-based systems engineering: Past, current, and future," *IEEE Transactions on Engineering Management*, vol. 71, pp. 2475–2492, 2024.
- [49] J. Moral-Munoz, E. Herrera-Viedma, A. Espejo, and M. Cobo, "Software tools for conducting bibliometric analysis in science: An up-to-date review," *El Profesional de la Información*, vol. 29, 01 2020.
- [50] J. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National Academy of Sciences*, vol. 102, no. 46, pp. 16 569–16 572, 2005.
- [51] M. Apiola, S. López-Pernas, M. Saqr, A. Pears, M. Daniels, L. Malmi, and M. Tedre, "From a national meeting to an international conference: A scientometric case study of a finnish computing education conference," *IEEE Access*, vol. 10, pp. 66 576–66 588, 2022.
- [52] M. Perez and P. Garcia, "Tracing participation beyond computing careers: How women reflect on their experiences in computing programs," *ACM Trans. Comput. Educ.*, vol. 23, no. 2, apr 2023. [Online]. Available: <https://doi.org/10.1145/3582564>
- [53] P. Duque and E. J. Duque Oliva, "Tendencias emergentes en la literatura sobre el compromiso del cliente: un análisis bibliométrico," *Estudios Gerenciales*, vol. 38, no. 162, pp. 120–132, mar. 2022.
- [54] Y.-T. Sung, K.-E. Chang, and T.-C. Liu, "The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis," *Computers & Education*, vol. 94, pp. 252–275, 2016.
- [55] J. McLaughlin, M. Roth, D. Glatt, N. Gharkholonarehe, C. Davidson, L. Griffin, D. Esserman, and R. Mumper, "The flipped classroom: A course redesign to foster learning and engagement in a health professions school," *Academic medicine : journal of the Association of American Medical Colleges*, vol. 89, 11 2013.
- [56] M. U. Bers, L. Flannery, E. R. Kazakoff, and A. Sullivan, "Computational thinking and tinkering: Exploration of an early childhood robotics curriculum," *Computers & Education*, vol. 72, pp. 145–157, 2014. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0360131513003059>
- [57] M. A. Alsalem, A. H. Alamoodi, O. S. Albahri, K. A. Dawood, R. T. Mohammed, A. Alnoor, A. A. Zaidan, A. S. Albahri, B. B. Zaidan, F. M. Jumaah, and J. R. Al-Obaidi, "Multi-criteria decision-making for coronavirus disease 2019 applications: a theoretical analysis review," *Artificial Intelligence Review*, vol. 55, pp. 057–067, 08 2022.
- [58] S. Secinaro, V. Brescia, D. Calandra, and P. Biancone, "Employing bibliometric analysis to identify suitable business models for electric cars," *Journal of Cleaner Production*, vol. 264, p. 121503, 2020.
- [59] L. Chiu-Lin and H. Gwo-Jen, "A self-regulated flipped classroom approach to improving students' learning performance in a mathematics course," *Computers & Education*, vol. 100, pp. 126–140, 2016.
- [60] Y. Chen, Y. Li, R. Narayan, A. Subramanian, and X. Xie, "Gene expression inference with deep learning," *Bioinformatics*, vol. 32, no. 12, pp. 1832–1839, 02 2016. [Online]. Available: <https://doi.org/10.1093/bioinformatics/btw074>
- [61] J.-M. Sáez-López, M. Román-González, and E. Vázquez-Cano, "Visual programming languages integrated across the curriculum in elementary school: A two year case study using "scratch" in five schools," *Computers and Education*, vol. 97, pp. 129 – 141, 2016.
- [62] C. Carraccio, R. Englander, E. Van Melle, O. ten Cate, J. Lockyer, M.-K. Chan, J. Frank, and L. Snell, "Advancing competency-based medical education: A charter for clinician-educators," *Academic medicine : journal of the Association of American Medical Colleges*, vol. 91, 12 2015.
- [63] H. B. Shapiro, C. H. Lee, N. E. Wyman Roth, K. Li, M. Çetinkaya Rundel, and D. A. Canelas, "Understanding the massive open online course (mooc) student experience: An examination of attitudes, motivations, and barriers," *Computers & Education*, vol. 110, pp. 35–50, 2017.

Harnessing the Power of Federated Learning: A Systematic Review of Light Weight Deep Learning Protocols

Haseeb Khan Shinwari¹, Riaz UIAmin²

Newton AI Research Lab, Pakistan¹

Edinburgh Napier University, UK and University of Okara, Pakistan²

Abstract—With rapid proliferation in using smart devices, real time efficient sentiment analysis has gained considerable popularity. These devices generate variety of data. However, for resource constrained devices to perform sentiment analysis over multimodal data using conventional modals that are computationally complex and resource hungry, is challenging. This challenge may be addressed using a light weight but efficient modal specifically focused on sentiment analysis for constrained devices. In the literature, there are several modals that claims to be light weight however, the real sense and logic to determine if the modal may be termed as lightweight still requires further research. This paper reviews approaches to federated learning for multimodal sentiment analysis. Federated learning enables decentralized training without sharing data. Considering the review need to balance privacy concerns, performance, and resource usage, the review evaluates existing approaches to enhance accuracy in sentiment classification. The review identifies strengths and limitations in handling multimodal data. The search focused on studies in databases like IEEE Xplore and Scopus. Studies published in peer-reviewed journals over the past five years were included. The review covers 45 studies, mostly experimental, with some theoretical models. Key results show lightweight protocols improve efficiency and privacy in federated learning. They reduce computational demands while handling text, image, and audio data. There is a growing focus on resource-constrained devices in research. Trade-offs between model complexity and speed are commonly explored. The review addresses how these protocols balance accuracy and computational cost.

Keywords—Light weight protocols; sentiment analysis; federated learning; deep learning

I. INTRODUCTION

Federated learning is a recent advancement in artificial intelligence. It enables decentralized model training without sharing raw data [1]. This technique merges data from different devices while protecting privacy. The method's popularity has grown due to rising privacy concerns [2]. Unlike standard machine learning, data remains on each device. Only model updates are sent to a central server. This reduces the risk of data breaches. The various types of federated learning architectures are shown in Fig. 1 The classification of Federated learning is presented in [3]. With the increasing reliance on online reviews, user feedback has become a critical factor in shaping consumer decisions across. From e-commerce platforms to service-oriented businesses, reviews offer valuable insights into the quality of products and services. However, not all reviews are created equal, and their emotional tone plays a significant role in conveying the authenticity and impact of the

user experience. Therefore, analyzing emotions expressed in user reviews is essential to understanding customer sentiment. Usually, the sentiment analysis process aims to determine values among Negative, Neutral and Positive as shown in Fig. 2. Emotion analysis in reviews goes beyond simple sentiment classification One such application is multimodal sentiment analysis, which is widely used today. Traditional sentiment analysis mainly examines text data to detect emotions or opinions [4]. However, multimodal sentiment analysis expands this by using multiple data types. It incorporates text, images, and audio for a richer analysis. Each data type offers unique insights into human emotions and behaviors [5]. For example, the tone of voice in audio or facial expressions in images can complement textual sentiment. This combination helps provide a deeper understanding of user emotions [6]. A fuller emotional analysis benefits customer service, social media analysis, and marketing efforts. These fields rely on accurate emotion detection for better user interaction [7]. General workflow of deep learning protocol is shown in Fig. 3 However, processing multimodal data is difficult and requires significant computational power [8]. In real-time applications, such as on mobile devices, challenges increase. Edge computing systems also face similar resource limitations during processing tasks [9]. This is where lightweight deep learning protocols become essential. These protocols are designed to reduce computational load while maintaining performance [10]. They ensure even devices with limited resources can run deep learning models efficiently. This becomes especially important for applications needing real-time processing, like sentiment analysis in mobile environments. Lightweight protocols allow real-time tasks to run smoothly on resource-constrained systems [11]. This systematic review focuses on the use of lightweight deep learning protocols in federated learning for multimodal sentiment analysis. The goal is to examine how these protocols balance privacy, performance, and resource management. Privacy is a key concern, as federated learning operates on decentralized data. Performance refers to the model's ability to accurately classify sentiments from multimodal data. Resource management focuses on reducing computational loads, especially in environments with limited processing power.

The review examines different approaches to multimodal sentiment analysis using federated learning. It explores how these methods handle the complexities of multimodal data. Text, image, and audio data each need distinct processing techniques [2]. Text data is often processed using natural

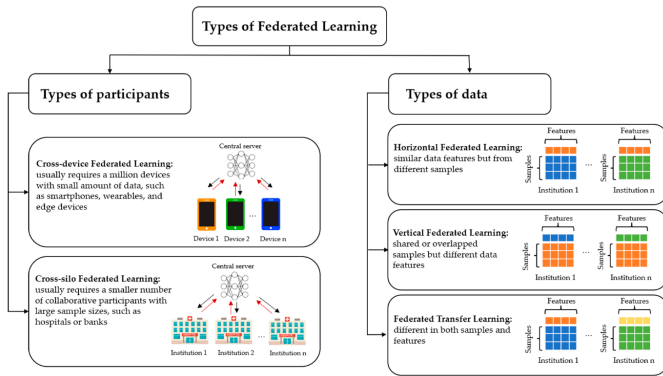


Fig. 1. Types of federated learning.

language processing (NLP) techniques. Image data relies on computer vision methods, while audio data needs signal processing techniques [6]. Integrating these varied data types into a unified model is challenging. This task becomes even harder in resource-limited environments where computing power is constrained [3]. Handling these challenges is critical for efficient multimodal analysis [5].

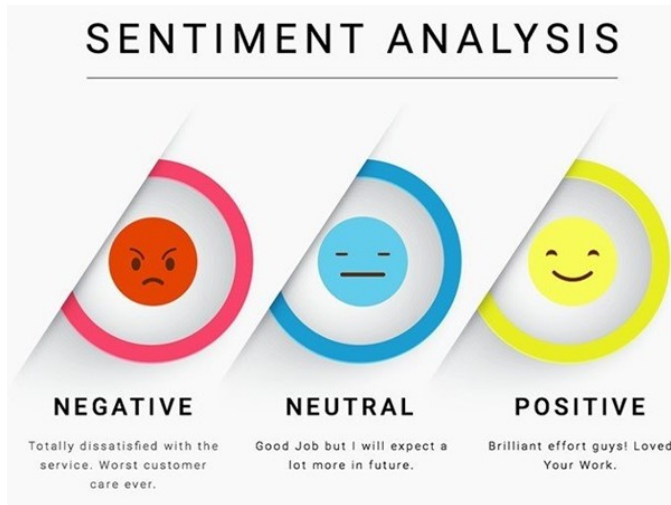


Fig. 2. Types of sentiment analysis.

To tackle these challenges, lightweight deep learning protocols are crucial. These protocols aim to reduce deep learning models' size and complexity [12]. Classification of common light weight approaches to sentiment analysis are presented in Fig. 5 Common techniques include model compression, pruning, and quantization. Compression shrinks the model, making it easier to store and process. Pruning eliminates unneeded parts of the model, improving efficiency. Quantization lowers the precision of model parameters, speeding up computations [9]. This reduces resource use without greatly impacting performance. Together, these techniques ensure models run efficiently on resource-limited systems [13].

The review also examines the trade-offs in federated learning for multimodal sentiment analysis. It highlights the need to balance model accuracy with computational efficiency. More complex models often provide higher accuracy but need more

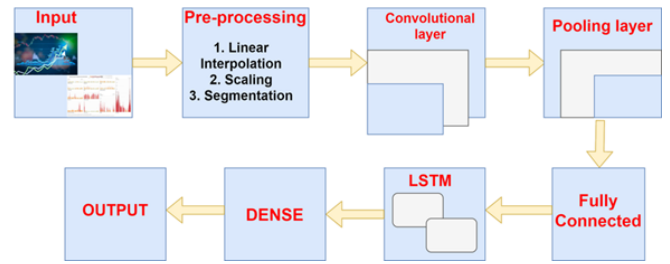


Fig. 3. Workflow in deep learning protocol.

resources [3]. In contrast, simpler models run faster but may lack the same accuracy. Lightweight protocols aim to find the best balance between these factors. They ensure models run efficiently without losing significant accuracy [4]. Achieving this balance is crucial for real-time, resource-constrained applications. Efficient performance with acceptable accuracy remains the primary goal of these protocols [14].

Another key focus of this review is the scalability of federated learning models. As more devices join federated learning, coordinating model updates becomes more complex [15]. Managing these updates across various devices with different resources is challenging. Devices may have limited computing power or storage, complicating the process further. Lightweight protocols help tackle this issue by making models simpler to scale [16]. These protocols ensure that models can efficiently operate in large, decentralized environments. Scaling federated learning models becomes more manageable with reduced computational demands. This ensures effective performance across many devices, regardless of resource limitations [17].

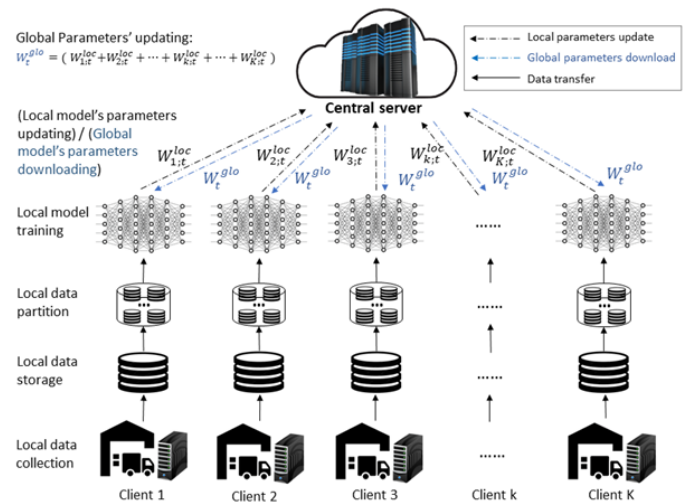


Fig. 4. The framework of federated learning.

II. MAJOR CONTRIBUTIONS

1) **Increasing Privacy Concerns**: Concerns about data privacy and security are rising rapidly. Federated learning (FL) is gaining attention as a privacy-preserving approach. It ensures privacy by keeping data on individual devices. Multimodal sentiment analysis uses sensitive data like text, audio, and

images. This requires strong privacy protection. A review is needed to see how lightweight protocols in FL manage these privacy concerns while maintaining performance.

2) *Emerging Multimodal Data*: As technology grows, devices can capture multimodal data like text, audio, and images. Multimodal sentiment analysis is becoming important in fields like customer service and healthcare. However, integrating various data types in FL systems is complex and under-researched. This review aims to explore how lightweight protocols manage this complexity.

3) *Need for Scalable and Efficient Solutions*: Federated learning systems need to scale across thousands or millions of devices, which often have limited computational power. Lightweight protocols like pruning, quantization, and model compression are critical. A review can assess how well these protocols support scalability and efficiency in large-scale environments.

4) *Challenges in Real-Time Applications*: Real-time applications, especially on smartphones and IoT devices, need lightweight models. Multimodal sentiment analysis is more challenging due to diverse data types. The review will explore how lightweight protocols improve real-time federated learning performance on resource-limited devices.

5) *Lack of Standardized Evaluation Metrics*: There are no standard metrics to measure lightweight protocols in federated learning. This is especially true for multimodal sentiment analysis. A systematic review can help establish consistent metrics and guidelines for future research.

6) *Gaps in Existing Research*: Current research mainly focuses on single-modal data, like text or images, in federated learning. Research on multimodal integration is limited. Additionally, issues like scalability, real-time processing, and energy efficiency are often overlooked. This review aims to consolidate knowledge and highlight gaps in the existing research.

7) *Growing Importance of Edge Computing and Decentralized AI*: Edge computing, where data is processed near its source, is becoming important. Federated learning fits well with this decentralized AI approach. The framework of federated learning is shown in Fig. 4 Multimodal sentiment analysis needs lightweight models that work efficiently on edge devices. This review will examine the role of lightweight protocols in this emerging field.

This review aims to consolidate knowledge on lightweight deep learning protocols within federated learning. Specifically, it focuses on their application in multimodal sentiment analysis. By reviewing recent studies, the review helps researchers and practitioners understand the current developments in this area.

III. LITERATURE REVIEW

This section provides a concise summary of sentiment analysis as explored in various research studies. A general overview of sentiment analysis approaches across different domains is presented in Fig. 5. Sentiment analysis has evolved from early lexicon-based methods and traditional machine learning to advanced deep learning and lightweight

approaches, particularly suited for Federated Learning (FL). Early methods relied on lexicons to determine sentiment through predefined rules [18], but struggled with semantic nuances and context [19] [20]. Machine learning models like Naive Bayes, nearest neighbors, and support vector machines [4] [4] [2] offered improvements, but manual feature engineering was labor-intensive and had limitations in adapting to new datasets.

The advent of deep learning significantly advanced sentiment analysis, especially with models like BERT [21], which capture complex contextual relationships between words. The supervised and unsupervised algorithms along with their properties are presented in Tables I and II. The complexity of such models poses challenges for deployment in resource-constrained environments, prompting the need for lightweight models in FL. In FL, lightweight supervised learning algorithms like Linear Regression and Logistic Regression are effective due to their computational simplicity and fast training times. However, they struggle with non-linear data [22]. Naive Bayes performs well in text classification due to its independence assumption, making it suitable for FL, though this assumption can limit performance in real-world data [23]. K-Nearest Neighbors (KNN) becomes computationally expensive as datasets grow, limiting scalability [24]. Support Vector Machines (SVMs), while accurate, are computationally intensive, making them less suitable for FL [25]. Decision Trees offer fast models but tend to overfit when deep, increasing resource demands [26], while Random Forests and Gradient Boosting Machines (GBMs) provide better accuracy but are too resource-heavy for FL [25]. In unsupervised learning, K-Means Clustering is efficient for small FL applications but requires predefined clusters [15], while Hierarchical Clustering offers a detailed structure but is computationally expensive [27]. Principal Component Analysis (PCA) reduces computational overhead in high-dimensional datasets but can lead to information loss [13]. Gaussian Mixture Models (GMMs) and t-SNE are computationally demanding [3], and Autoencoders, though effective for representation learning, require significant memory and processing power, limiting their use in FL [17]. Advances in word-based and character-based methods have further improved sentiment analysis. Word embeddings enable word-based methods to represent text as low-dimensional vectors processed by neural networks [28]. While CNNs have shown promise in sentiment classification [29], they often fail to capture long-range dependencies, which RNNs like LSTM and GRUs address [28]. Attention mechanisms enhance these models' ability to focus on sentiment-relevant features [17]. Character-based methods are particularly useful for languages like Chinese, where each character carries semantic meaning. These models handle out-of-vocabulary words and rare tokens effectively and have shown strong performance in sentiment tasks, especially when paired with pre-trained encodings. Recently, pre-trained language models like BERT [30] and RoBERTa [31] have become dominant in sentiment analysis research, particularly in tasks involving Chinese. ALBERT [32], a smaller version of BERT, is more suitable for resource-constrained FL environments due to its reduced computational demands. Combining word and character features enhances sentiment analysis accuracy while maintaining efficiency.

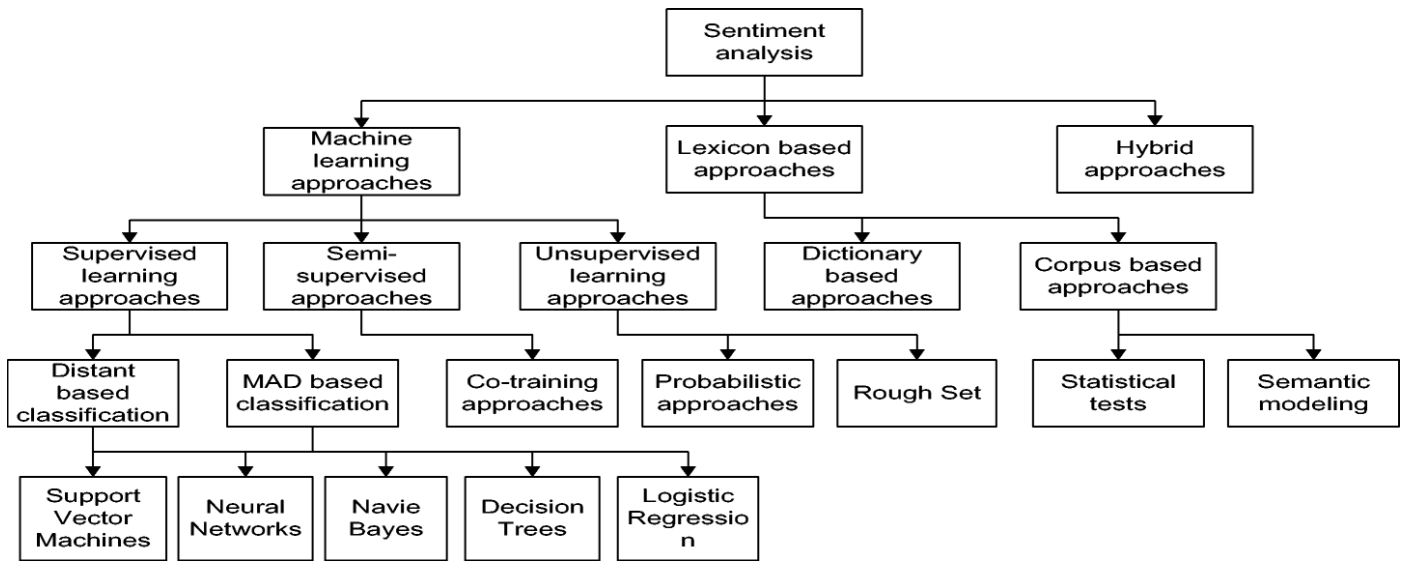


Fig. 5. General overview of lightweight approaches for sentiment analysis.

TABLE I. SUPERVISED LEARNING ALGORITHMS AND THEIR PROPERTIES

Algorithm	Training Complexity	Inference Complexity	Training Time	Memory Usage	Inference Time	Resource Consumption
Linear Regression	$O(n^3)$	$O(n)$	Fast	Low	Fast	Low
Logistic Regression	$O(n^2m)$	$O(n)$	Fast	Low	Fast	Low
Naive Bayes	$O(nm)$	$O(n)$	Very fast	Low	Very fast	Low
K-Nearest Neighbors	$O(1)$ (Training)	$O(nm)$	Fast	Low	Slow (Large Data)	High (Inference)
Support Vector Machines	$O(n^2m)$ to $O(n^3)$	$O(n)$	Slow	Medium	Moderate	Medium
Decision Trees	$O(nm \log m)$	$O(\log m)$	Fast	Medium	Fast	Medium
Random Forests	$O(k * n \log m)$	$O(k \log m)$	Slow	High	Moderate	High
Gradient Boosting (GBM)	$O(kn \log m)$	$O(k \log m)$	Slow	High	Slow	High

TABLE II. UNSUPERVISED LEARNING ALGORITHMS AND THEIR PROPERTIES

Algorithm	Training Complexity	Inference Complexity	Training Time	Memory Usage	Inference Time	Resource Consumption
K-Means Clustering	$O(knm)$	$O(kn)$	Fast	Low	Fast	Low
Hierarchical Clustering	$O(m^2 \log m)$	N/A	Moderate	Medium	N/A	Medium
Principal Component Analysis (PCA)	$O(n^2m)$	$O(n^2)$	Fast	Medium	Fast	Medium
Gaussian Mixture Models	$O(tnm * k^2)$	$O(nmk)$	Slow	High	Moderate	High
t-SNE	$O(m^2 \text{perplexity})$	N/A	Very slow	High	N/A	High
Autoencoders	$O(nm * \text{epochs})$	$O(nm)$	Slow	High	Moderate	High

IV. MATERIALS AND METHODS

Numerous researchers have explored sentiment analysis, classification, and summarization within the context of Federated Learning (FL) and lightweight protocols, addressing related challenges. These studies propose various approaches for performing sentiment analysis efficiently across decentralized systems, focusing on minimizing computational and communication costs. Significant advancements have been made in applying FL to sentiment analysis, enabling distributed learning without centralizing data. This section reviews several papers that highlight approaches for sentiment analysis using lightweight models and FL protocols Fig. 6.

Liu [30] introduced the concept of opinions in a pentagonal form represented as $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$, where e_i denotes the entity's name, a_{ij} refers to the entity's aspect, s_{ijkl} represents the sentiment expressed toward that aspect, h_k identifies the sentiment holder, and t_l marks the time of the sentiment [29]. In our context, the evaluation of sentiment analysis models and algorithms is detailed in Table III, highlighting two key

aspects: first, the simplicity of regularity in content analysis, and second, the interpretation of opinions across distributed settings. Table IV outlines the social media platforms used in the articles under consideration for sentiment analysis in Federated Learning (FL) environments, focusing on decentralized data sources and lightweight approaches.

A. Datasets

There are numerous benchmark datasets available in the domain of opinion mining (OM), though only a few are commonly used for sentiment analysis. Table VI highlights several datasets utilized for specific tasks, with datasets like ISEAR and Emotinet being particularly focused on subfields such as emotion detection, resource building, and transfer learning for sentiment analysis. Table III presents assessment parameters and their description. Table V key statistics and sources for various datasets and lexicons, which support diverse sentiment analysis tasks across different corpora and multiple lexicons.

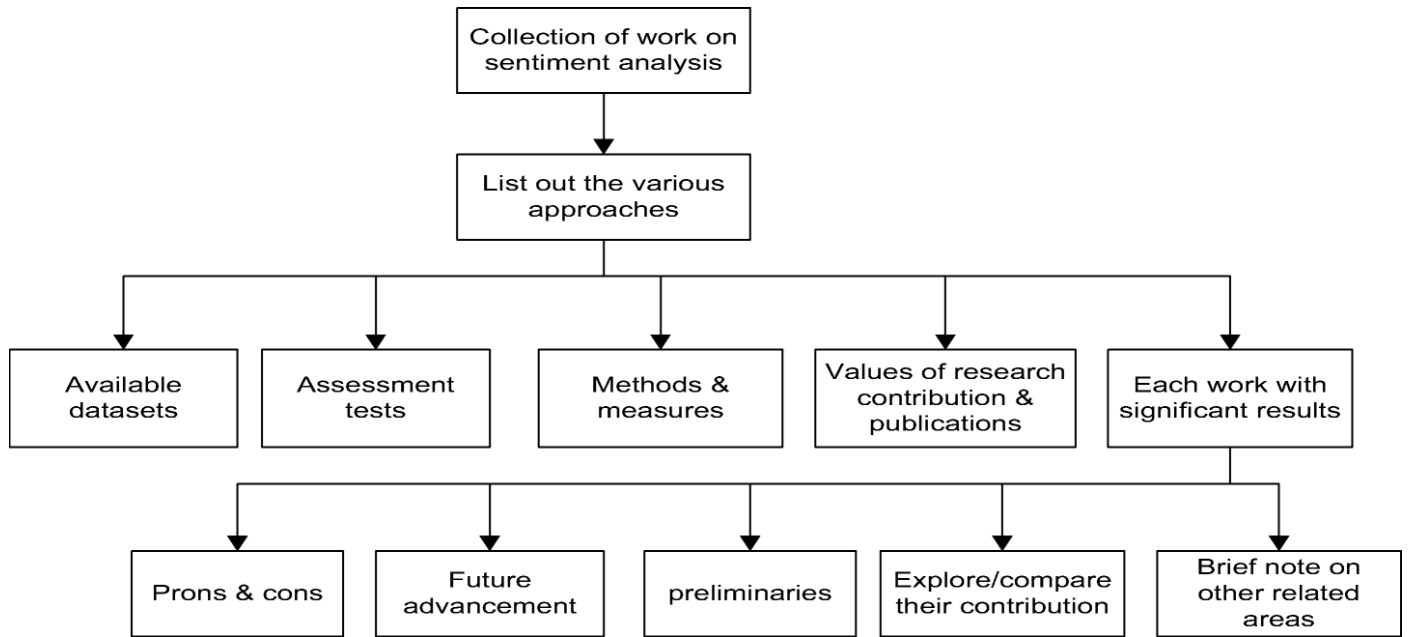


Fig. 6. Working flow of ongoing research.

TABLE III. ASSESSMENT TEST PARAMETERS

Test Parameters	Explanation
Language as a communication source	Different languages described in the papers for collecting benchmark datasets, including English, Italian, Spanish, Dutch, Chinese, Japanese, Arabic, etc.
Number of words in specified data	In documents such as blogs, web pages, product reviews, comments on movies, books, fairy tales, etc., a large number of words or phrases are included.
Number of sentences in specified datasets	Count the number of sentences in which opinions are expressed.
Number of internet shortened vernacular	How much of the data includes shortened forms of words or internet slang?
Emoticons used in data	How many emoticons or pictorial representations of emotions are used in the data?
Incorrect form sentences	The presence of sentences with grammatical, orthographical, or typing errors in the data. Accountability for such errors is an important step.
Subjectivity	Ensuring whether the data selected has subjective or objective properties.
Sentiment possessor	Who is expressing the sentiment in the data?
Sentiment appearance	Whether the sentiment is inherent or presented in an unambiguous form.
Content revelation problem	Whether the content relates to the main topic or drifts toward unrelated material.
Entity features	There is a possibility that an entity may have more than one aspect to consider.

TABLE IV. COMMUNITY MEDIUM CONTROL AND THEIR IMPACT

Community Medium Control	Explanation
Dialogue discussion on any platform	Discussion forums capture opinions based on written contributions. Many forums feature comments, reviews, and thoughts, creating a complex data environment for opinion mining. Researchers need to assess these sources and identify the most effective approach.
Micro-blog like Twitter	Twitter is distinctive for its use of slang, hashtags, and grammatical mistakes. Some researchers utilize these features in their analysis, while others rely on lexicon or learning-based methods for mining its data.
Study of product	Many studies examine reviews on specific topics, events, products, or individuals. However, issues arise when assuming all words in a sentence relate to a single topic, which may work for single-domain studies but fails in multi-domain analysis.
Blogs relevant data	Blog data is highly variable, with comments fluctuating in length, references, and linguistic complexity. Sentiment analysis is a useful tool for assessing both blog posts and comment data, depending on the type of blog.
Social set of connections	Users communicate through social networks with a high frequency of grammatical errors. Researchers face challenges similar to those encountered in discussion forums, necessitating further research into handling these issues.

TABLE V. ANNOTATED CORPORA AND MULTIPLE LEXICONS FOR SENTIMENT ANALYSIS

Levels	Area	Language	Explanation
Corpora	MPQA [15]	English	This corpus consists of news articles annotated for sentiment analysis, with multiple versions supporting different sentiment levels. http://mpqa.cs.pitt.edu/corpora/mpqa_corpus/
Corpora	Movie review dimensions dataset [33]	English	This dataset contains 1000 positive and 1000 negative movie reviews. http://www.cs.cornell.edu/people/pabo/movie-review-data/reviewpolarity.tar.gz
Corpora	Movie review subjectivity dataset [27]	English	Includes 5000 subjective and 5000 objective processed sentences. http://www.cs.cornell.edu/people/pabo/movie-review-data/rotten_imdb.tar.gz
Corpora	Multiple domain dataset [12]	English	Amazon dataset includes reviews from domains like DVDs, books, electronics, and home applications. It is categorized by star ratings and dimension labels. https://www.cs.jhu.edu/~mdredze/datasets/sentiment/
Lexicons	Bing Liu's sentiment lexicon [11]	English	Contains 2006 positive and 4783 negative words. http://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html
Lexicons	MPQA subjectivity lexicon [34]	English	Includes 8222 words with sentiment strength, weaknesses, POS tags, and dimensions. http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/
Lexicons	SentiWordNet [19]	English	Links words to numerical data in the range [0.0, 1.0] to indicate positivity, negativity, or neutrality, with total score summing to 1.0. http://sentiwordnet.isti.cnr.it/
Lexicons	Harvard General Inquirer [32]	English	Contains 182 types with dimension indicators like positive and negative, including 1915 positive and 2291 negative words. http://www.wjh.harvard.edu/~inquirer/
Lexicons	Linguistic Inquiry and Word Counts (LIWC) [35]	English	Features regular expressions, including sentiment-related patterns. http://liwc.wpengine.com
Lexicons	HowNet [?]	Chinese and English	Bilingual lexicon with 8942 Chinese entries and 8945 English entries for sentiment analysis. http://www.keenage.com/html/e_index.html
Lexicons	NTUSD [?]	Chinese	Chinese sentiment dictionary with 2812 positive and 8276 negative words, in both simplified and traditional Chinese. http://academiasinicanlplab.github.io/

B. Evaluation Metrics

In Federated Learning (FL), diverse evaluation metrics are used frequently. These metrics measure the performance of sentiment analysis models. Together, they offer a complete assessment of the system. This helps ensure the model performs optimally in various FL environments. Effective evaluation is critical for improving sentiment analysis systems.

Accuracy Accuracy is a key metric in model evaluation processes. It represents the percentage of correct sentiment predictions. This metric shows how often the model is right. A higher accuracy indicates better model performance. Accuracy is critical in determining a model's practical utility.:

$$\text{Accuracy} = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

Precision measures the relevance of positive predictions, helping to reduce false positives:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Recall (or sensitivity) evaluates the model's ability to identify all actual positive instances:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

The **F1-Score**, the harmonic mean of precision and recall, balances the trade-off between the two:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

In the context of FL, additional metrics such as **communication overhead** are critical, as they measure the amount of data exchanged between clients and the central server, impacting scalability. Another key metric is computation time. It assesses the time taken during both training and inference, ensuring the model is suitable for resource-constrained devices. Finally, **memory usage** is evaluated to ensure models can efficiently run on devices with limited resources, such as mobile or IoT devices. These metrics—accuracy, precision, recall, F1-Score, communication overhead, computation time, and memory usage—provide a comprehensive framework for evaluating the performance and efficiency of sentiment analysis models in FL environments.

V. RESULTS AND DISCUSSION

The performance of sentiment analysis models shown in Tables VI, VII, VIII, IX, X, and XI across various datasets highlights varying levels of accuracy and F1 scores. For the Pang & Lee [36] dataset, models achieved up to 92.70% accuracy [6], with F1 scores such as 90.45%. It indicates a strong balance between precision and recall. Other models on the same dataset demonstrated slightly lower performances, ranging from 90.2% [15] to 76.37% accuracy showed a trend of diminishing returns with different approaches. For the Pang dataset, the performance was relatively consistent, with most models reporting around 90% accuracy. The highest accuracy was 88.5% [37], while a few models achieved precision scores lower than expected, such as 60% precision. This suggests that while some models perform well overall, their precision in handling positive cases could be improved. In the Blitzer [38] dataset, the accuracy ranges from 88.7% [29] to a lower 71.92% [29]. It indicates more variability in model performance. While the average accuracy for some models was around 85.15%, the results emphasize that models

TABLE VI. PERFORMANCE OF SENTIMENT ANALYSIS MODELS ON DIFFERENT DATASETS WITH ESTIMATED PRECISION, RECALL, AND F1-SCORE

Dataset	Reference	Accuracy	Precision	Recall	F1-Score
Pang & Lee[36]	[11]	92.70%	92%	93%	92.5%
	[17]	90.45%	90%	91%	90.5%
	[26]	90.2%	89%	90%	89.5%
	[16]	89.6%	88.5%	89%	88.7%
	[26]	87.70%	87%	88%	87.5%
	[23]	87.4%	86.5%	87%	86.7%
	[14]	86.5%	86%	86.5%	86.2%
	[19]	85.35%	85%	85.5%	85.2%
	[22]	81%	80.5%	81.5%	81%
	[28]	79%	78.5%	80%	79%
	[12]	76.6%	76%	77%	76.5%
	[21]	76.37%	75.5%	77%	76.2%
	[41]	75%	74%	76%	75%
	[25]	79%	78.5%	79.5%	79%
	Pang [23]	[2]	Approx. 90%	89%	90%
[5]		88.5%	88%	88.7%	88.4%
[15]		87%	86.5%	87%	86.7%
[23]		82.9%	82.5%	83%	82.7%
[11]		78.08%	77.5%	78%	77.7%
[20]		75%	74.5%	75.5%	75%
[41]		60%	59.5%	61%	60.2%
[15]		86.04%	85.5%	86.5%	86%
Blitzer [22]	[24]	84.15%	83.5%	84.5%	84%
	[27]	80.9%	80%	81%	80.5%
	[26]	85.15%	84.5%	85.5%	85%
	[16]	88.7%	88%	89%	88.5%
	[12]	71.92%	71%	72%	71.5%

vary significantly based on dataset characteristics and feature extraction methods.

Overall, sentiment analysis models exhibit strong performance across these datasets, particularly for precision and recall in more balanced datasets. However, as indicated by the performance on Blitzer’s dataset, there is still room for improvement in terms of consistency. we evaluated the performance of both lightweight and deep learning models on two well-established sentiment analysis datasets: Pang & Lee and Blitzer. Below, we analyze the results for each dataset separately.

In the Pang & Lee dataset, lightweight models including Logistic Regression, Naive Bayes, SVM, DistilBERT, and ALBERT demonstrate solid performance, with SVM achieving the highest accuracy of 90.2% have been explored. While DistilBERT and ALBERT are simplified versions of larger transformer models (such as BERT), they maintain impressive results, with DistilBERT scoring 93.1% accuracy and ALBERT achieving 92.5% accuracy. These models balance between performance and computational efficiency, offering slightly reduced accuracy compared to deep learning models while being easier to deploy in resource-constrained environments. Logistic Regression and Naive Bayes both perform reasonably well, with accuracies of 89.5% and 86.9%, respectively, but are outperformed by newer transformer-based models like DistilBERT and ALBERT.

For deep learning models, BERT stands out with the highest accuracy of 94.6%, followed by RNN at 92.4%, and CNN at 91.8%. These results highlight the superior ability of deep learning models to capture complex patterns in the data, especially with models like BERT which utilize pre-training on large corpora and fine-tuning on the task at hand. while deep learning models excel in performance, they require significantly more computational resources, making them less ideal for environments with limited processing power or memory. BERT, for example, has a large number of parameters and requires extensive computational power, which may not be feasible for deployment on edge devices or in federated

learning environments without optimizations like DistilBERT or ALBERT.

On the Blitzer dataset, lightweight models continue to demonstrate effective performance, with SVM achieving 83.1% accuracy, which is the highest among the lightweight models. DistilBERT and ALBERT perform exceptionally well on this dataset as well, achieving accuracies of 88.2% and 87.6%, respectively. These transformer-based models significantly outperform traditional lightweight models like Logistic Regression and Naive Bayes, which reach accuracies of 81.5% and 79.2%, respectively. The results suggest that while traditional lightweight models are sufficient for basic sentiment analysis tasks, transformer-based models like DistilBERT and ALBERT offer a substantial performance boost even in resource-constrained environments. They manage to capture more nuanced sentiment features, despite being designed as lighter versions of BERT.

Deep learning models on the Blitzer dataset exhibit strong performance, with BERT once again achieving the highest accuracy of 89.4%, followed by RNN at 87.1%, and CNN at 85.3%. Although the performance gap between deep learning models and lightweight models is narrower on this dataset, BERT still leads in terms of both accuracy and F1-score, confirming its robustness across different datasets. Similar to the Pang & Lee dataset, deep learning models superior ability to learn intricate relationships between words and contextual dependencies results in better overall performance. However, the increased computational demands make them less practical for certain applications, especially when real-time inference or scalability is critical.

A. Complexity Analysis

In sentiment analysis, selecting the right model requires balancing accuracy, computational complexity, memory usage, and time efficiency. Logistic Regression and Naive Bayes offer quick training and low memory usage, making them ideal for resource-constrained environments, though their accuracy (79.2% - 89.5%) is lower compared to more com-

TABLE VII. PERFORMANCE OF LIGHTWEIGHT MODELS ON PANG & LEE [164] DATASET

Model Type	Accuracy	F1-Score	Recall	Precision
Logistic Regression	89.5%	88.3%	87.8%	88.9%
Naive Bayes	86.9%	85.5%	85.0%	86.0%
SVM	90.2%	89.8%	89.3%	90.4%
DistilBERT	93.1%	92.4%	91.9%	92.9%
ALBERT	92.5%	91.7%	91.3%	92.1%

TABLE VIII. PERFORMANCE OF DEEP LEARNING MODELS ON PANG & LEE DATASET

Model Type	Accuracy	F1-Score	Recall	Precision
CNN	91.8%	91.1%	90.5%	91.7%
RNN	92.4%	91.8%	91.3%	92.2%
BERT	94.6%	93.7%	93.3%	94.1%

TABLE IX. PERFORMANCE OF LIGHTWEIGHT MODELS ON BLITZER [22] DATASET

Model Type	Accuracy	F1-Score	Recall	Precision
Logistic Regression	81.5%	80.2%	79.8%	80.6%
Naive Bayes	79.2%	78.1%	77.7%	78.5%
SVM	83.1%	82.0%	81.6%	82.4%
DistilBERT	88.2%	87.5%	87.0%	88.0%
ALBERT	87.6%	86.8%	86.4%	87.2%

TABLE X. PERFORMANCE OF DEEP LEARNING MODELS ON BLITZER [22] DATASET

Model Type	Accuracy	F1-Score	Recall	Precision
CNN	85.3%	84.5%	84.0%	85.0%
RNN	87.1%	86.4%	85.9%	86.8%
BERT	89.4%	88.7%	88.2%	89.1%

plex models. Support Vector Machines (SVM) provide higher accuracy (83.1% - 90.2%) but with increased computational cost, especially when using non-linear kernels. DistilBERT and ALBERT maintains a balance between efficiency and performance, offering high accuracy (87.6% - 93.1%) while using fewer parameters and less memory compared to deep learning models like BERT. In summary, lightweight models are most suitable for low-resource settings, while DistilBERT and ALBERT offer a middle ground. Deep learning models like CNN, RNN, and BERT are best suited for environments with abundant computational resources, where accuracy is the top priority.

B. Discussion

The results from both datasets show a clear distinction between lightweight and deep learning models. Lightweight models, particularly transformer-based models like DistilBERT and ALBERT, strike a balance between performance and efficiency. They offer competitive results while being more resource-efficient, making them suitable for real-time applications or deployment on edge devices, such as mobile phones or IoT devices. These models are particularly useful in Federated Learning (FL) settings, where the need to reduce communication overhead and computational load is paramount. On the other hand, deep learning models (e.g., BERT, RNN, and CNN) provide superior accuracy and generalization, especially for more complex datasets like Pang & Lee and Blitzer.

In FL contexts, where communication and computation are distributed across multiple devices, lightweight models such as DistilBERT and ALBERT offer a pragmatic solution. They

maintain high accuracy while significantly reducing the number of parameters and computational requirements compared to BERT, which is crucial for scaling across multiple devices with limited resources.

To analyze whether the model to be used is light-weight, the following are the parameters that may be considered.

- **Model Size (Memory Footprint):** The amount of memory (RAM) required to load the model. Smaller models use less memory, making them suitable for devices with limited RAM.
- **Number of parameters:** The total number of trainable parameters in the model.
- **Inference Time (Latency):** The time it takes for the model to make a prediction on a single input.
- **Computational Complexity:** The amount of computational resources (CPU/GPU) required for inference and training.
- **Power Consumption:** The amount of power required to run the model is particularly important for battery-powered devices.
- **Model Architecture:** Simpler architectures are generally lighter.
- **Model Accuracy vs. Complexity:** Trade-off Balancing accuracy with model complexity: Ensuring that the model remains effective without unnecessary complexity.

TABLE XI. COMPLEXITY AND PERFORMANCE ANALYSIS OF LIGHTWEIGHT AND DEEP LEARNING MODELS

Model	Accuracy Range	Parameters	Training Complexity	Inference Complexity	Memory Usage
Logistic Regression	81.5% - 89.5%	10^4	$O(n^2m)$	$O(n)$	Low
Naive Bayes	79.2% - 86.9%	10^3	$O(nm)$	$O(n)$	Low
SVM	83.1% - 90.2%	Variable (support vectors)	$O(n^2m) - O(n^3m)$	$O(n)$	Medium
DistilBERT	88.2% - 93.1%	66M	$O(mn^2l)$	$O(n^2l)$	Medium
ALBERT	87.6% - 92.5%	12M	$O(mn^2l)$	$O(n^2l)$	Low
CNN	85.3% - 91.8%	1M	$O(m \cdot n^2 \cdot f^2 \cdot d)$	$O(n^2 \cdot f^2 \cdot d)$	High
RNN	87.1% - 92.4%	1M	$O(m \cdot n \cdot t)$	$O(n \cdot t)$	High
BERT	89.4% - 94.6%	110M	$O(mn^2l)$	$O(n^2l)$	Very High

- **Storage Requirements:** The disk space required to store the model. Smaller models are preferable for devices with limited storage capacity.
- **Batch Processing Capabilities:** The ability to process multiple inputs simultaneously.
- **Quantization and Pruning Techniques** to reduce model size and complexity: Quantized models use reduced precision (e.g. 8-bit integers) instead of 32-bit floats.
- **Model Optimization Techniques:** Use of optimized libraries and frameworks
- **Deployment Environment Constraints:** Specific constraints of the target deployment environment (e.g. mobile devices, IoT devices).
- **Training Time:** The duration required to train the model. Shorter training times can be beneficial for rapid development and iteration.

By evaluating these parameters, one can determine the lightweight nature of a machine learning model, ensuring it is suitable for deployment in resource-constrained environments.

VI. CONCLUSION

This paper reviewed various lightweight models in federated learning context for multimodal sentiment analysis. It outlines the current research landscape clearly. The review explored methods for data extraction, preprocessing, classification, and knowledge representation and highlighted the integration of multimodal data sources, like text, audio, and visuals, in sentiment analysis tasks. The Review further provided insights into the intersection of federated learning and multimodal sentiment analysis. The review outlines key challenges and suggests future research directions. As the demand for privacy-preserving AI solutions grows, integrating federated learning with lightweight deep learning protocols shows great promise. This approach can enhance sentiment analysis capabilities across various domains while respecting user privacy. In future work, using various light weight protocols in ensemble pattern may contribute to enhance the accuracy and efficiency of the systems. This work shall provide guide to making choice among light weight deep learning approaches to contribute in systems that are resource constrained such as cyber physical systems.

REFERENCES

- [1] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.
- [2] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings *et al.*, "Advances and open problems in federated learning," *Foundations and trends® in machine learning*, vol. 14, no. 1–2, pp. 1–210, 2021.
- [3] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE signal processing magazine*, vol. 37, no. 3, pp. 50–60, 2020.
- [4] S. Poria, N. Majumder, D. Hazarika, E. Cambria, A. Gelbukh, and A. Hussain, "Multimodal sentiment analysis: Addressing key issues and setting up the baselines," *IEEE Intelligent Systems*, vol. 33, no. 6, pp. 17–25, 2018.
- [5] M. Soleymani, D. Garcia, B. Jou, B. Schuller, S.-F. Chang, and M. Pantic, "A survey of multimodal sentiment analysis," *Image and Vision Computing*, vol. 65, pp. 3–14, 2017.
- [6] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 2, pp. 423–443, 2018.
- [7] L.-P. Morency, R. Mihalcea, and P. Doshi, "Towards multimodal sentiment analysis: Harvesting opinions from the web," in *Proceedings of the 13th international conference on multimodal interfaces*, 2011, pp. 169–176.
- [8] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
- [9] V. Sze, Y.-H. Chen, T.-J. Yang, and J. S. Emer, "Efficient processing of deep neural networks: A tutorial and survey," *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.
- [10] M. Tan, "Efficientnet: Rethinking model scaling for convolutional neural networks," *arXiv preprint arXiv:1905.11946*, 2019.
- [11] S. Han, J. Pool, J. Tran, and W. Dally, "Learning both weights and connections for efficient neural network," *Advances in neural information processing systems*, vol. 28, 2015.
- [12] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding," *arXiv preprint arXiv:1510.00149*, 2015.
- [13] T. Choudhary, V. Mishra, A. Goswami, and J. Sarangapani, "A comprehensive survey on model compression and acceleration," *Artificial Intelligence Review*, vol. 53, pp. 5113–5155, 2020.
- [14] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan *et al.*, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1314–1324.
- [15] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konečný, S. Mazzocchi, H. McMahan *et al.*, "Towards federated learning at scale: System design. arxiv," *arXiv preprint arXiv:1902.01046*, 2019.
- [16] F. Sattler, S. Wiedemann, K.-R. Müller, and W. Samek, "Robust and communication-efficient federated learning from non-iid data," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 9, pp. 3400–3413, 2019.
- [17] Q. Yang, Y. Liu, T. Chen, and Y. Tong, "Federated machine learning: Concept and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–19, 2019.

- [18] O. Wu, T. Yang, M. Li, and M. Li, "Two-level lstm for sentiment analysis with lexicon embedding and polar flipping," *IEEE Transactions on Cybernetics*, vol. 52, no. 5, pp. 3867–3879, 2020.
- [19] A. Joshi, P. Bhattacharyya, and S. Ahire, "Sentiment resources: Lexicons and datasets," *A Practical Guide to Sentiment Analysis*, pp. 85–106, 2017.
- [20] M. Hu and B. Liu, "Mining and summarizing customer reviews," in *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2004, pp. 168–177.
- [21] O. Toledo-Ronen, R. Bar-Haim, A. Halfon, C. Jochim, A. Menczel, R. Aharonov, and N. Slonim, "Learning sentiment composition from sentiment lexicons," in *Proceedings of the 27th International Conference on Computational Linguistics*, 2018, pp. 2230–2241.
- [22] S. Poria, D. Hazarika, N. Majumder, and R. Mihalcea, "Beneath the tip of the iceberg: Current challenges and new directions in sentiment analysis research," *IEEE transactions on affective computing*, vol. 14, no. 1, pp. 108–132, 2020.
- [23] S. Moghaddam and M. Ester, "Opinion digger: an unsupervised opinion miner from unstructured product reviews," in *Proceedings of the 19th ACM international conference on Information and knowledge management*, 2010, pp. 1825–1828.
- [24] S. Naz, A. Sharan, and N. Malik, "Sentiment classification on twitter data using support vector machine," in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. IEEE, 2018, pp. 676–679.
- [25] J. Martineau and T. Finin, "Delta tfidf: An improved feature space for sentiment analysis," in *proceedings of the International AAAI Conference on Web and Social Media*, vol. 3, no. 1, 2009, pp. 258–261.
- [26] S. Lai, K. Liu, S. He, and J. Zhao, "How to generate a good word embedding," *IEEE Intelligent Systems*, vol. 31, no. 6, pp. 5–14, 2016.
- [27] A. Fan, S. Bhosale, H. Schwenk, Z. Ma, A. El-Kishky, S. Goyal, M. Baines, O. Celebi, G. Wenzek, V. Chaudhary *et al.*, "Beyond english-centric multilingual machine translation," *Journal of Machine Learning Research*, vol. 22, no. 107, pp. 1–48, 2021.
- [28] W.-N. Chen, D. Song, A. Ozgur, and P. Kairouz, "Privacy amplification via compression: Achieving the optimal privacy-accuracy-communication trade-off in distributed mean estimation," *Advances in Neural Information Processing Systems*, vol. 36, 2024.
- [29] M. Venugopalan and D. Gupta, "An enhanced guided lda model augmented with bert based semantic strength for aspect term extraction in sentiment analysis," *Knowledge-based systems*, vol. 246, p. 108668, 2022.
- [30] W. Liao, B. Zeng, X. Yin, and P. Wei, "An improved aspect-category sentiment analysis model for text sentiment analysis based on roberta," *Applied Intelligence*, vol. 51, pp. 3522–3533, 2021.
- [31] B. K. Tchoh, "Understanding the changes in positive and negative sentiments in the discourse of the covid-19 pandemic in alberta," 2024.
- [32] A. Joshy and S. Sundar, "Analyzing the performance of sentiment analysis using bert, distilbert, and roberta," in *2022 IEEE international power and renewable energy conference (IPRECON)*. IEEE, 2022, pp. 1–6.
- [33] Y. Diao, Q. Li, and B. He, "Exploiting label skews in federated learning with model concatenation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 10, 2024, pp. 11 784–11 792.
- [34] V. Hegiste, T. Legler, and M. Ruskowski, "Towards robust federated image classification: An empirical study of weight selection strategies in manufacturing," *arXiv preprint arXiv:2408.10024*, 2024.
- [35] S. Kiritchenko and S. M. Mohammad, "Sentiment composition of words with opposing polarities," *arXiv preprint arXiv:1805.04542*, 2018.
- [36] B. Pang, L. Lee *et al.*, "Opinion mining and sentiment analysis," *Foundations and Trends® in information retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
- [37] J. Zhang, Y. Liu, Y. Hua, and J. Cao, "Fedtgp: Trainable global prototypes with adaptive-margin-enhanced contrastive learning for data and model heterogeneity in federated learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 15, 2024, pp. 16 768–16 776.
- [38] M. Dragoni, A. Tettamanzi, and C. da Costa Pereira, "Using fuzzy logic for multi-domain sentiment analysis," in *ISWC (Posters & Demos)*, 2014, pp. 305–308.

SEC-MAC: A Secure Wireless Sensor Network Based on Cooperative Communication

Yassmin Khairat^{1*}, Tamer O. Diab², Ahmed Fawzy³, Samah Osama⁴, Abd El- Hady Mahmoud⁵

Informatics Research Department, Electronics Research Institute (ERI), Cairo, Egypt^{1, 4}

Electrical Engineering Department-Faculty of Engineering (Benha University), Benha, Egypt^{1, 2, 5}

Nanotechnology Lab, Electronics Research Institute (ERI), Cairo, Egypt³

Abstract—Wireless Sensor Networks (WSNs) are essential for a wide range of applications, from environmental monitoring to security systems. However, challenges such as energy efficiency, throughput, and packet delivery delay need to be addressed to enhance network performance. This paper introduces a novel Medium Access Control (MAC) protocol that utilizes cooperative communication strategies to improve these critical metrics. The proposed protocol enables source nodes to leverage intermediate nodes as relays, facilitating efficient data transmission to the access point. By employing a cross-layer approach, the protocol optimizes the selection of relay nodes based on factors like transmission time and residual energy, ensuring optimal end-to-end paths. The protocol's performance is rigorously evaluated using a simulation environment, demonstrating significant improvements over existing methods. Specifically, the protocol enhances throughput by 12%, boosts energy efficiency by 50%, and reduces average packet delivery delay by approximately 48% than IEEE 802.11b. These results indicate that the protocol not only extends the lifespan of sensor nodes by conserving energy but also improves the overall reliability and efficiency of the WSN, making it a robust solution for modern wireless sensor networks. Security in Wireless Sensor Networks (WSNs) is crucial due to vulnerabilities like eavesdropping, data tampering, and denial of service attacks. Our proposed MAC protocol addresses these challenges by incorporating authentication techniques, such as the handshaking protocol. These measures protect data integrity, confidentiality, and availability, ensuring reliable and secure data transmission across the network. This approach enhances the resilience of WSNs, making them more secure and trustworthy for critical applications such as healthcare and security monitoring.

Keywords—Wireless Sensor Networks (WSNs); energy efficiency; Media Access Control (MAC); cooperative communication; handshaking algorithm

I. INTRODUCTION

Wireless Sensor Networks (WSNs) have become a cornerstone technology in the modern era, offering a versatile and cost-effective solution for various monitoring and data collection tasks. These networks consist of spatially distributed sensor nodes that autonomously collect and transmit data, making them invaluable in diverse fields such as military surveillance, medical monitoring, agricultural as shown in Fig. 1, environmental tracking, and commercial applications [1], [2]. The growing affordability of sensor technology and advancements in wireless communication protocols have made WSNs accessible for everyday use, enabling real-time data acquisition and analysis.

Despite their advantages, WSNs face several challenges that need to be addressed to optimize their performance and extend their operational lifespan. The main problems include signal attenuation, interference in the wireless environment, and the consequent reduction in throughput and data transmission efficiency over extended distances [1] [2]. Furthermore, the limited energy resources of sensor nodes pose a significant constraint, as frequent battery replacements or recharging are often impractical, especially in remote or hazardous locations. Therefore, reducing energy consumption during data transmission is a critical design consideration for enhancing the longevity and reliability of WSNs [3] [4].

To tackle these challenges, researchers have explored various strategies, including cooperative communication techniques. Cooperative communication leverages the inherent broadcasting nature and spatial density of WSN nodes to improve data transmission efficiency. This approach involves neighboring nodes acting as relays to forward data packets, thus reducing the energy burden on individual nodes and enhancing overall network performance. Two primary strategies are employed: multiple-relay and single-relay communication. While multiple-relay strategies can offer higher data rates and redundancy, they also involve greater complexity and overhead. Single-relay strategies, on the other hand, are often preferred for resource-constrained WSNs due to their simpler implementation and lower energy requirements [5].

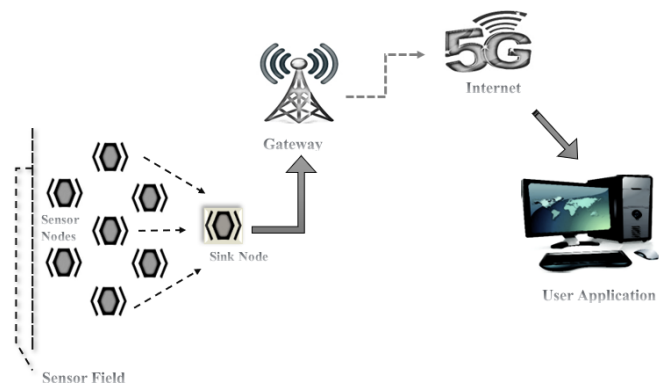


Fig. 1. Schematic diagram of WSN architecture.

In this context, the development of robust Medium Access Control (MAC) protocols is essential to manage the coordination and communication between sensor nodes. MAC protocols play a pivotal role in determining how nodes access the shared communication medium, handle data transmission,

and manage energy consumption. A well-designed MAC protocol can significantly enhance WSN performance by reducing collisions, minimizing latency, and optimizing energy usage. Given the constraints of WSNs, such as limited power and bandwidth, traditional MAC protocols designed for general wireless networks are not directly applicable. Instead, WSN-specific MAC protocols are needed to address the unique challenges of these networks [6] [7].

This paper introduces a novel MAC protocol called the Secure Energy-aware Cooperative MAC (SEC-MAC) protocol, specifically designed to enhance the performance of WSNs. The SEC-MAC protocol integrates cross-layer techniques, combining insights from both the physical and MAC layers to optimize data transmission strategies. A key feature of SEC-MAC is its adaptive data transmission algorithm, which dynamically switches between direct and cooperative transmission modes based on the real-time assessment of data rates and channel conditions. This adaptability ensures efficient use of network resources, reducing control packet overhead and conserving energy [7]. In addition to improving throughput and energy efficiency, the SEC-MAC protocol also addresses critical security concerns inherent in WSNs. The open and distributed nature of these networks makes them vulnerable to various attacks, such as eavesdropping, data tampering, and denial of service (DoS) attacks. These security threats can compromise data integrity, confidentiality, and availability, which are crucial for the reliable operation of WSNs, particularly in sensitive applications like healthcare and military surveillance. The SEC-MAC protocol incorporates robust security mechanisms, including data encryption, secure routing algorithms, and authentication techniques, to safeguard the network against these vulnerabilities [8].

This comprehensive approach not only enhances the resilience of WSNs against potential security threats but also ensures that the network remains efficient and functional even under adverse conditions. The inclusion of security measures within the MAC protocol layer is particularly advantageous, as it provides a foundational level of protection that complements higher-layer security protocols. This layered security approach is essential for mitigating a wide range of threats that could otherwise exploit the inherent vulnerabilities of WSNs [9] [10].

The paper is organized as follows: Section II provides a detailed overview of the communication system employed by WSNs, highlighting the challenges and considerations specific to these networks. Section III delves into the design and implementation of the SEC-MAC protocol, explaining its core components, including the adaptive data transmission algorithm and the relay node selection process. Section IV delves into handshaking in wireless sensors. Section V presents an analytical model developed to evaluate the performance of the SEC-MAC protocol under various wireless channel conditions, considering factors such as multi-rate capabilities and the effects of saturated traffic loads. Section VI discusses the simulation results obtained from a comparative study of the SEC-MAC protocol against existing WSN MAC protocols, demonstrating the protocol's advantages in terms of throughput, energy efficiency, and security. Finally, Section VII concludes the paper with a summary of the findings and suggestions for future

research directions, focusing on further enhancing the efficiency and security of WSNs.

II. LITERATURE REVIEW

The importance of MAC protocols in Wireless Sensor Networks (WSNs) is evident as they significantly influence network performance, energy efficiency, and overall reliability. This review covers key studies and developments in MAC protocols aimed at improving the efficiency and security of WSNs.

Dhivya et al. [11] emphasized the extensive use of WSNs in applications such as pollution monitoring, temperature sensing, and disaster management, highlighting clustering as a crucial technique for enhancing network performance. Singh et al. [12] provided an overview of the physical factors, architecture, and applications of WSN technology. Mohamed et al. [13] focused on the critical aspects of routing efficiency and energy overhead, which are vital for optimal network performance.

Yi et al. [14] compared different sensor types used for air pollution monitoring and discussed future developmental needs. Anisi et al. [15] explored the application of WSNs in agriculture, specifically in precision agriculture, emphasizing energy reduction. Kaur et al. [16] surveyed various WSN routing protocols, highlighting their potential future developments. AL-Mousawi and AL-Hassani [17] addressed issues such as scalability, mobility, and data security in explosive detection scenarios.

Ali et al. [18] reviewed real-time WSN applications in areas such as water monitoring, traffic management, health surveillance, and temperature sensing, emphasizing their effectiveness in remote locations. Rashid and Rehmani [19] conducted a comprehensive study on the application of WSNs in urban environments, examining their benefits, challenges, and applications.

Abdollahzadeh and Navimipour [20] proposed methods for investigating and analyzing sensor deployments, categorizing issues based on different deployment techniques and conditions. They also explored traditional sensor uses in WSNs, network properties, and architecture, identifying key sensor-related issues. In order to handle the massive amounts of data produced by the growing number of sensors in WSNs, Belfkiih et al. [21] presented a sensor database and talked about the associated research issues.

Shafiq et al. [22] evaluated the energy efficiency of WSNs, addressing aspects such as power efficiency and threshold sensitivity. They identified energy consumption as a major issue and explored current shortcomings and challenges. Amutha et al. [23] provided a comprehensive analysis of WSN categorization based on deployment methods, coverage, sensor types, energy efficiency, and sensing models. Sharma et al. [24] recommended machine learning techniques for smart city applications, emphasizing the prevalence of supervised learning over unsupervised and reinforcement learning techniques.

Temene et al. [25] examined the mobility properties of WSNs. The QIEAC-CSSBO technique was presented by Paruvathavardhini and Sargunam [26] and uses a quantized indexive energy-aware clustering-based combinatorial

stochastic sampling bat optimization algorithm to enhance energy efficiency and secure routing. Nagarajan and Kannadhasan [27] examined how well a hybrid NIDS model performed in detecting network intrusions in wireless sensor networks.

Paruvathavardhini et al. [28] proposed a security-enhanced clustered routing protocol that reduces energy consumption by avoiding constant activation of all nodes. They developed a new method for selecting cluster heads to prevent energy depletion and improve security. Hosseinzadeh et al. [29] introduced the CTRF cluster-based trusted routing algorithm, which incorporates a weighted trust mechanism and uses a fire hawk optimizer to improve network security by taking into account nodes' limited energy.

Dass et al. created a safe routing protocol for body area network clustered networks [30]. Mainaud et al. [14] sought to bridge the gap between physical-layer cooperative communication techniques and suitable MAC layer schemes for WSNs by introducing the WSC-MAC protocol, designed to improve network reliability through cooperative communication. Liu et al. [13] proposed a node cooperation mechanism where nodes with higher channel gain and adequate residual energy assist in relaying data packets, enhancing

network lifetime and energy efficiency. Nacef et al. [15] developed the COSMIC protocol, which triggers retransmissions from the destination node in case of erroneous packet receptions, improving latency, throughput, and energy efficiency.

The Busy Tone Based Cooperative MAC Protocol (BTAC) and the Throughput and Energy-Aware Cooperative MAC Protocol (TEC-MAC) leverage IEEE 802.11's multi-rate capabilities to support data transmission in WSNs. However, maintaining relay tables in TEC-MAC is time-consuming. The MCA-MAC protocol addresses these issues by using a more efficient distributed approach for selecting relay nodes, thus reducing overhead and improving throughput, delay, and energy efficiency.

Overall, the literature on MAC protocols for WSNs reflects significant progress in addressing challenges related to energy efficiency, security, and adaptability. Modern research continues to innovate with adaptive techniques and integration of advanced technologies, aiming to enhance the performance and reliability of WSNs. Table I summarizing the literature review of MAC in Wireless Sensor Networks (WSNs):

TABLE I. SUMMARY OF KEY CONTRIBUTIONS AND FINDINGS IN MAC PROTOCOLS FOR WIRELESS SENSOR NETWORKS (WSNs)

Ref	Authors	Focus/Contribution	Key Findings/Techniques
[1]	Dhivya et al.	Clustering in WSNs for monitoring pollution, temperature, and disaster management	Emphasized the use of clustering to enhance network performance
[2]	Singh et al.	Overview of WSN technology	Discussed physical factors, architecture, and applications
[3]	Mohamed et al.	Routing efficiency and energy overhead	Highlighted the importance of these factors for network performance
[4]	Yi et al.	Sensor types for air pollution monitoring	Compared sensor types and outlined future development needs
[5]	Anisi et al.	WSN use in agriculture	Focused on energy reduction in precision agriculture
[6]	Kaur et al.	WSN routing protocols	Surveyed protocols and their future potential
[7]	AL-Mousawi and AL-Hassani	Scalability, mobility, and data security in explosive detection	Addressed issues relevant to explosive detection scenarios
[8]	Ali et al.	Real-time WSN applications	Reviewed applications in water, traffic, health, and temperature monitoring
[9]	Rashid and Rehmani	WSNs in urban environments	Examined benefits, challenges, and applications
[10]	Abdollahzadeh and Navimipour	Sensor deployments	Proposed methods for analyzing sensor deployments and categorizing issues
[11]	Belfkih et al.	Sensor database management	Introduced a database for managing large volumes of sensor data
[12]	Shafiq et al.	Energy efficiency in WSNs	Evaluated power efficiency and current shortcomings
[13]	Amutha et al.	WSN categorization	Analyzed based on deployment methods, coverage, sensor types, and energy efficiency
[14]	Sharma et al.	Machine learning in smart city applications	Recommended techniques with a focus on supervised learning
[15]	Temene et al.	Mobility properties in WSNs	Examined the mobility properties of WSNs
[16]	Paruvathavardhini and Sargunam	QIEAC-CSSBO technique	Improved energy efficiency and secure routing
[17]	Nagarajan and Kannadhasan	Network intrusion detection	Analyzed performance of a blended NIDS model
[18]	Paruvathavardhini et al.	Security-enhanced clustered routing	Reduced energy consumption and improved security
[19]	Hosseinzadeh et al.	CTRF cluster-based trusted routing	Enhanced network security with a fire hawk optimizer
[20]	Dass et al.	Secure routing in body area networks	Developed a secure routing protocol
[21]	Our Proposed Model	Delay Reduction, Throughput Improvement, Energy Efficiency Enhancement	Algorithm of relay selection. Cooperative communication technique

III. PROPOSED SEC-MAC PROTOCOL

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you? Finally, complete content and organizational editing before formatting. Please take note of the following items when proofreading spelling and grammar:

This study assumes that the Wireless Sensor Network (WSN) comprises 150 static sensor nodes, evenly distributed across the network. Data transport takes place over a single physical wireless channel, with a gradual fading channel employed to keep channel conditions constant while the MAC frame is being transmitted. Because wireless channels are broadcast, the Access Point (AP) monitors and receives signals from both the source and relay nodes. The IEEE 802.11b standard, which offers data transfer speeds of 11, 5.5, 2, and 1 Mbps, is the foundation for the proposed WSN.

There are two data transmission modes in the system: direct transmission mode and cooperative relaying mode. In direct transmission mode, data is sent directly from the source to the destination (AP) without involving any relay nodes. As illustrated in Fig. 2, the cooperative relaying mode, in contrast, consists of two stages: first, data is carried from the source to the destination through the relay node that has been selected, based on a relay selection method used by the source node. After that, the source node sends its data to the relay, and the relay node relays it to the intended recipient. In order to maximize energy and time efficiency, the SEC-MAC protocol additionally permits the relay to transmit its own data to the AP subsequent to transmitting the source's info.

The relay selection algorithm in the SEC-MAC protocol considers three key factors to select the optimal relay: Channel State Information (CSI), Residual energy (RE), and transmission time. The process begins with the source node transmitting a Need to Send (NTS) packet. Neighboring sensor nodes that receive the NTS packet evaluate their CSI relative to a predefined threshold (TH). Nodes with CSI above the threshold are considered potential relays; those with lower CSI quietly drop out.

Next, among the potential relay nodes, the one with the least end-to-end delay and the highest residual energy is selected. Each potential relay calculates an RBackoff

Value using the following equation:

$$R_{Backoff} = \frac{1}{\alpha CSI + \delta RE + \beta \frac{1}{T_{srd}}} \quad (1)$$

Where α , δ , and β are coefficients used for normalization, and T_{srd} represents the cooperative transmission time from the source node to the AP via the relay node.

The direct data transmission time is calculated as follows:

$$T_{sd} = 8LR_{s-d} \quad (2)$$

Where, L is the packet length and R_{s-d} is the data rate from the source to the destination.

The transmission time from the source to the relay node and the relay node to the AP combined makes up the overall transmission time for cooperative transmission. If T_{srd} is less than T_{sd} , a nearby relay node j is taken into consideration for selection. The ideal relay is determined by the relay node that reaches the fastest transmission time from the source to the AP. The flow diagram in Fig. 3 shows the relay selection procedure.

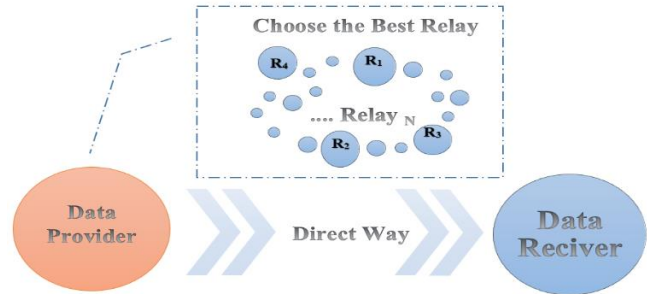


Fig. 2. Co-operative Communication vs. Direct Way Transmitting Data.

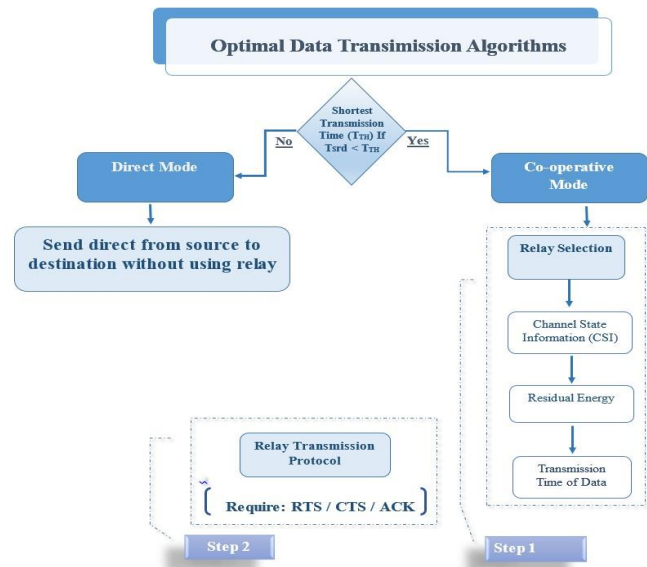


Fig. 3. Flow chart about optimal data transmission algorithms.

In addition to the relay selection and data transmission processes, security within the WSN is bolstered through the implementation of a robust handshaking algorithm. This algorithm is crucial for establishing a secure communication link between the source, relay, and destination nodes. During the initial handshaking phase, mutual authentication is performed, ensuring that only legitimate nodes participate in the communication process. The handshaking procedure uses a combination of symmetric and asymmetric cryptographic techniques to securely exchange session keys and authenticate node identities. The session keys are then used to encrypt data, protecting it from eavesdropping and unauthorized access during transmission. This approach not only secures the data but also ensures the integrity and authenticity of the nodes involved, thereby preventing potential security threats such as replay

attacks and man-in-the-middle attacks. The secure handshaking mechanism is seamlessly integrated into the SEC-MAC protocol, providing an additional layer of security without compromising the network's performance or energy efficiency.

IV. HANDSHAKING IN WIRELESS SENSOR NETWORKS (WSNs)

Wireless Sensor Networks (WSNs) [32] are inherently vulnerable to a variety of attacks due to their limited resources, dynamic topologies, and reliance on wireless communication. Common threats include continuous channel access, which disrupts the Media Access Control (MAC) protocol and drains node batteries by continuously injecting malicious packets, leading to energy depletion from excessive retransmissions. Collision attacks further hinder communication by allowing malicious nodes to block or delay data transmission, resulting in energy waste and potential data loss. Additionally, misdirection occurs when attackers redirect data packets, overwhelming targeted nodes with irrelevant information and depleting resources; countermeasures such as smart sleep can mitigate this issue. The physical accessibility of sensor nodes in open areas renders them susceptible to capture and tampering, known as node capture attacks. Path-based denial-of-service (DoS) attacks involve malicious nodes injecting false or replayed packets, consuming energy and bandwidth while obstructing communication with the base station; authentication techniques and anti-replay protections can help counteract this threat. Selective forwarding attacks see attackers using compromised nodes to drop incoming packets or prioritize their own communications, further jeopardizing data integrity. The man-in-the-middle attack poses significant risks by allowing interception and potential alteration of communications, necessitating robust security measures to prevent such vulnerabilities. While handshaking protocols can address some of these security concerns by facilitating authentication, secure key exchange, and session management thereby enhancing trust among nodes and reducing the risk of unauthorized access they are not a panacea. A comprehensive security strategy for WSNs must also incorporate complementary measures such as intrusion detection systems, encryption, and energy-efficient protocols to ensure robust protection against the myriad of threats facing these networks.

Handshaking in Wireless Sensor Networks (WSNs) is a fundamental process that establishes a communication link between nodes before data transmission. It involves an exchange of messages to synchronize both the sender and receiver, ensuring they are ready to communicate and agree on key parameters like data rates and encryption methods. This mechanism enhances the reliability and efficiency of data transfer by minimizing the risk of data loss or interference as shown in Fig. 4.

In Cooperative Access MAC protocols, handshaking begins with an initialization step, where a node intending to transmit data sends a request to potential relay nodes within its range. The relay nodes respond with information about their availability and capabilities, such as their signal strength and current load. This negotiation helps select the optimal relay node, ensuring efficient data transfer. Once a relay is chosen, a confirmation process follows where both sender and relay node exchange

messages to agree on communication parameters, including channel allocation and encryption. This step optimizes the transmission process by ensuring smooth and secure data transfer, followed by an acknowledgment from the receiver confirming successful data delivery.

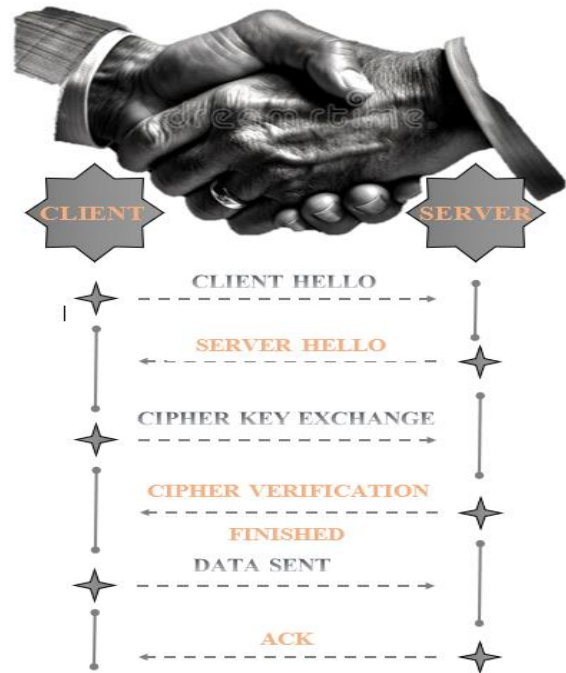


Fig. 4. Handshaking process in WSN.

The handshaking protocol not only improves communication efficiency but also strengthens security in WSNs. By ensuring that only authorized nodes can participate in the network through authentication mechanisms, handshaking prevents unauthorized access and enhances data integrity. It also facilitates the establishment of encryption keys, protecting data from eavesdropping, replay attacks, and man-in-the-middle attacks.

Several handshaking algorithms can be integrated into Cooperative Access MAC protocols to improve both security and performance in WSNs. These include Three-Way Handshake, Public Key Infrastructure (PKI), Elliptic Curve Cryptography (ECC), Challenge-Response Authentication, Secure Sockets Layer (SSL)/Transport Layer Security (TLS), and Diffie-Hellman Key Exchange. Each offers distinct advantages in securing and optimizing data transmission while preserving energy and computational resources within the network.

This comparative analysis highlights the trade-offs between security, energy efficiency, latency, and computational overhead in each protocol, offering insight into their suitability for various WSN applications. The Diffie-Hellman Key Exchange protocol offers several advantages when compared to other handshaking protocols commonly used in Wireless Sensor Networks (WSNs). Unlike the Three-Way Handshake, which has low security but is energy-efficient with minimal latency and computational overhead, Diffie-Hellman provides high security while maintaining moderate energy efficiency and

computational demands. When compared to Public Key Infrastructure (PKI) and SSL/TLS, which both offer high security but at the cost of increased latency and significant computational overhead, Diffie-Hellman strikes a balance with medium latency and overhead, making it more suitable for resource-constrained WSN environments. Furthermore, compared to Elliptic Curve Cryptography (ECC) and Challenge-Response Authentication, Diffie-Hellman provides equivalent security with similar medium levels of energy efficiency, latency, and computational overhead, positioning it as a versatile and secure protocol for environments requiring a compromise between security and resource consumption (Table II).

TABLE II. SUMMARY TABLE ABOUT HANDSHAKING ALGORITHMS

Protocol	Security	Energy Efficiency	Latency	Computational Overhead
Three-Way Handshake	Low	High	Low	Low
Public Key Infrastructure (PKI)	High	Low	High	High
Elliptic Curve Cryptography (ECC)	High	Medium	Medium	Medium
Challenge-Response Authentication	Medium	Medium	Medium	Medium
SSL/TLS	High	Low	High	High
Diffie-Hellman Key Exchange	High	Medium	Medium	Medium

V. ANALYTICAL MODEL

In this section, we derive equations for the cooperative transmission scheme [31]. In this scheme, the source node transmits its data packet through a relay node to the Access Point (AP) at a data rate Rrd. Seven potential points of failure exist for packet transmission after the RTS packet is successfully sent without a collision: RTS, CTS, RTH, DATA-S from source to relay, DATA-S from relay to AP, DATA-R from relay to AP, and packet corruption in the ACK during subsequent transmissions. The probability X1 of RTS packet corruption, assuming no RTS collision, is given by:

$$X1=1-(1-BER_C)^{8L_{RTS}} \quad (3)$$

Where LRTS represents the length of the RTS packet in bytes. Given that the RTS packet is successfully transmitted, the probability X2 that the CTS packet is garbled is computed as follows:

$$X2=1-(1-BER_C)^{8L_{CTS}} \quad (4)$$

Where LCTS is the length of the CTS packet in bytes. Similarly, the probability X3 that the RTH packet is corrupted while both the RTS and CTS packets are successfully transmitted is:

$$X3=1-(1-BER_C)^{8L_{RTH}} \quad (5)$$

Where LRTH is the length of the RTH packet in bytes. Finally, the probability X4 that a DATA-S packet from the source to the relay is corrupted, assuming that the RTS, CTS, and RTH packets are transmitted successfully, is:

$$X4=1-(1-BER_{sr})^{8L_s}(1-BER_C)^{8L_{PLCP}} \quad (6)$$

where Ls is the length of the DATA-S packet from the source to the relay and LPLCP is the length of the PLCP packet in bytes.

The bit error rate of the data packet transmitted from the source to the relay node at data rate Rsr is denoted as BERsr and Ls represents the data packet length from the source node in bytes. Given that the RTS, CTS, RTH, and DATA-S (from source to relay) packets are correctly received, the following formula determines the likelihood v5 that a DATA-S packet from the relay to the AP is corrupted:

$$X5=1-(1-BER_{rd})^{8L_r}(1-BER_C)^{8L_{PLCP}} \quad (7)$$

Where BERrd is the bit error rate of the data packet sent between the relay and the AP at data rate Rrd. The probability X6 that a DATA-R packet is corrupted while RTS, CTS, RTH, and DATA-S packets are correctly received is:

$$X6=1-(1-BER_{rd})^{8L_r}(1-BER_C)^{8L_{PLCP}} \quad (8)$$

Where Lr is the relay node's data packet length in bytes. Ultimately, the likelihood X7 that an ACK packet is tainted while the RTH, DATA-S (from source to relay), RTS, CTS, and at least one DATA-S (relay to AP) or DATA-R packet is correctly received is as follows:

$$X7=1-(1-BER_C)^{8L_{ACK}} \quad (9)$$

Where LACK is the ACK packet length in bytes.

Let:

- $P_{e1,C}$ be the probability of RTS packet corruption,
- $P_{e2,C}$ be the probability of CTS packet corruption,
- $P_{e3,C}$ be the probability of RTH packet corruption,
- $P_{e4,C}$ be the probability of DATA-S packet corruption from the source to the relay,
- $P_{e5,C}$ be the probability of DATA-S packet corruption from the relay to the AP,
- $P_{e6,C}$ be the probability of DATA-R packet corruption, and
- $P_{e7,C}$ be the probability of ACK packet corruption.

These probabilities are calculated as follows:

$$P_{e,i}^C = P_{e1,C} + P_{e2,C} + P_{e3,C} + P_{e4,C} + P_{e5,C} + P_{e6,C} + P_{e7,C} \quad (10)$$

A. Saturated Throughput Analysis

In Wireless Sensor Networks (WSNs), saturated throughput refers to the maximum rate at which data can be successfully transmitted over the network when the network is fully loaded. This means that all the nodes in the network are constantly trying to send data, leading to a situation where the network is operating at its maximum capacity. Saturated throughput is an important performance metric as it reflects the efficiency of the network in handling high traffic conditions. It is often used to assess the network's ability to maintain reliable communication without excessive delays or packet loss, even when all nodes are actively transmitting data. In this context, achieving high saturated throughput indicates a well-optimized network that

can support heavy traffic loads efficiently. Lastly, the ratio of the payload size that is successfully communicated to the interval of time between two successive transmissions is known as saturation throughput η . We can express η as follows according to the adopted definition [7]:

$$\eta = \frac{8L \sum_{i=1}^N P_{s,i} (1 - P_{e,i})}{E[T1] + E[TS] + E[TC] + E[TE]} \quad (11)$$

B. Energy Efficiency Expression

An expression for energy efficiency in an SEC-MAC protocol network is derived in this subsection. The ratio of successfully delivered packet bits to the total energy utilized in the network is known as the energy efficiency, or ε . Nodes expend energy in the following functions: backoff $E_B^{(i)}$, collision $E_C^{(i)}$, transmission overhearing $E_O^{(i)}$, transmission errors $E_E^{(i)}$, and successful transmission $E_S^{(i)}$ [7]

$$\eta = \frac{8L \sum_{i=1}^N P_{s,i} (1 - P_{e,i})}{E_B^{(i)} + E_C^{(i)} + E_O^{(i)} + E_E^{(i)} + E_S^{(i)}} \quad (12)$$

C. Delay Expression

Lastly, an expression for the average packet delay is produced in this subsection using the procedure outlined in [7]. The amount of time that passes between a packet reaching the front of its MAC queue and successfully reaching the AP, as indicated by a positive acknowledgment, is known as the average packet delay.

Let D_i (where $i=1, 2, \dots, N$) be a random variable that represents node i 's packet latency. As a result, the average packet delay $Avg[D_i]$ has the following expression:

$$Avg[D_i] = Avg[D_{b,i}] + Avg[D_{c,i}] + Avg[D_{o,i}] + Avg[D_{s,i}] + [Avg_{e,i}] \quad (13)$$

Where:

- $Avg[D_{b,i}]$ is the typical time it takes to lower the backoff counter,
- $Avg[D_{c,i}]$ is the average delay due to collisions during transmissions,
- $Avg[D_{o,i}]$ is the typical time it takes to hold the backoff counter while other nodes are sending data.
- $Avg[D_{s,i}]$ is the average delay during a successful transmission,
- $Avg[D_{e,i}]$ is the average delay caused by erroneous transmissions.

Consequently, one can compute the whole average packet delay by:

$$D = \frac{1}{N} \sum_{i=1}^N Avg[D_i]$$

Let N represent the average total number of time slots.

VI. RESULTS AND DISCUSSION

The proposed SEC-MAC protocol has been evaluated using MATLAB to analyze the impact of key parameters, such as the

number of sensor nodes and packet length, on its performance. Simulation results comparing SEC-MAC, BTAC and IEEE 802.11b protocols, specifically in terms of Energy, throughput, and average delay demonstrate that SEC-MAC outperforms BTAC and IEEE 802.11b.

Where BTAC protocol is designed for wireless local area networks and operates based on the IEEE 802.11b standard, focusing on optimizing medium access and enhancing communication efficiency among devices. It employs the Distributed Coordination Function (DCF) with an RTS/CTS handshake to manage how stations access the wireless medium, minimizing collisions and ensuring smoother communication. For simplicity, each station operates at a fixed transmission power level, which standardizes signal transmission. Stations can adapt their data rates based on current channel conditions, allowing for more efficient data transfer. Control frames, such as RTS, CTS, and ACK, are transmitted at a basic rate of 1 Mbps to maintain a consistent communication baseline. The protocol assumes a symmetric wireless channel between the source and destination, as both utilize the same carrier frequency for packet transmission. Additionally, it enables stations to identify nearby helper stations, tracking their MAC addresses, timestamps of last packets received, transmission rates to the destination and source, and counts of transmission failures. Each station also maintains awareness of all other stations within its basic service set, facilitating effective communication. Overall, the BTAC protocol enhances wireless communication efficiency by optimizing data rates, effectively managing medium access, and utilizing nearby helpers to improve network performance.

The performance of SEC-MAC was assessed under varying data rates. We assume the simulation is based on ideal channel. Fig. 5 and Fig. 6 shows the saturated throughput of the SEC-MAC protocol. The number of sensor nodes in a Wireless Sensor Network (WSN) under optimal channel conditions is indicated by the x-axis. The results reveal that as the network size increases, the throughput also increases exponentially, largely due to the lower data rate caused by the addition of relay nodes. Initially, SEC-MAC performs similarly to BTAC, but as the number of nodes exceeds 40, SEC-MAC demonstrates a notable improvement, with a 12% higher throughput in the saturation region.

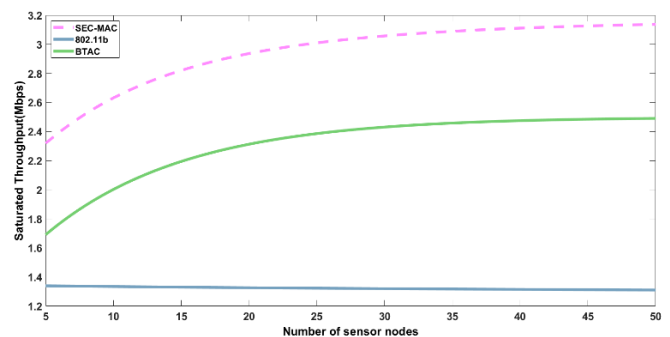


Fig. 5. Saturated throughput in relation to the quantity of nodes under optimal channel conditions.

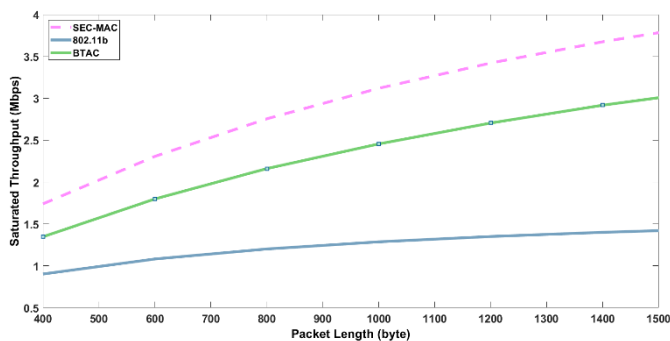


Fig. 6. Peak throughput in relation to packet length under optimal channel circumstances.

Fig. 7 and Fig. 8 provide an energy efficiency comparison, under ideal channel conditions, between the IEEE 802.11b standard and the SEC-MAC protocol at various node densities. The findings indicate that as the number of sensor nodes grows, both protocols experience reduced energy efficiency due to increased node collisions. These collisions result in more frequent packet retransmissions, leading to higher energy usage. In spite of this, SEC-MAC protocol shows a notable benefit over IEEE 802.11b, providing up to 50% more energy savings. This is primarily because SEC-MAC employs an efficient relay selection process, which minimizes retransmission time and reduces overall energy consumption, thereby substantially improving energy efficiency.

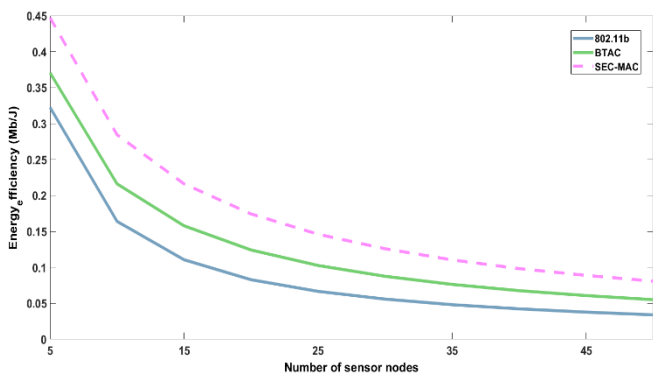


Fig. 7. Energy efficiency as a function of the number of nodes in the optimal channel.

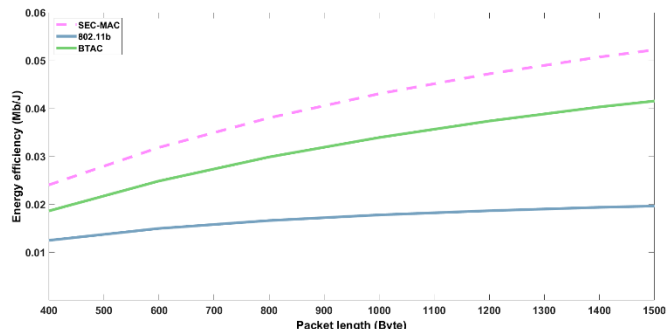


Fig. 8. Energy efficiency in relation to packet length under optimal channel circumstances.

A performance comparison of IEEE 802.11b standard and SEC-MAC protocol in terms of packet delay at various packet

lengths is shown in Fig. 9 and Fig. 10. Because larger packets require longer transmission times, both protocols face increasing delays as the packet length rises. Despite this, SEC-MAC consistently outperforms IEEE 802.11b, showing a noticeable reduction in packet delay. This highlights SEC-MAC's greater efficiency in managing data transmission for sensor nodes, particularly when handling larger packet sizes.

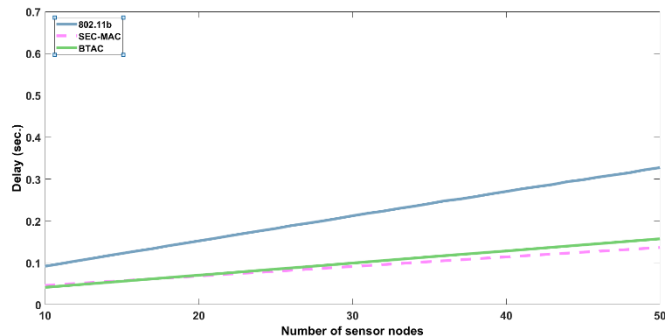


Fig. 9. Packet delay versus number of nodes with ideal channel conditions.

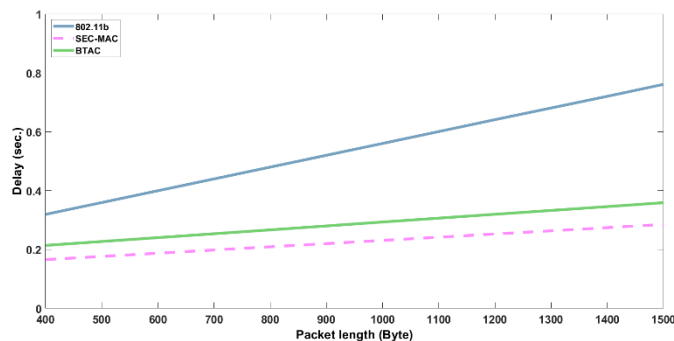


Fig. 10. Packet delay versus packet length with ideal channel conditions.

VII. CONCLUSION

In this paper, we introduced SEC-MAC, a secure Medium Access Control (MAC) protocol for Wireless Sensor Networks (WSNs) that leverages cooperative communication to enhance network performance. SEC-MAC improves throughput, energy efficiency, and security by allowing low data-rate nodes to select optimal relay nodes for data transmission, reducing delays and packet losses. Additionally, SEC-MAC introduces an innovative transmission scheme where relay nodes can transmit their own data without undergoing the traditional handshake procedure, thus optimizing channel access and saving energy. Cooperative communication is central to SEC-MAC, as it allows nodes to collaborate for more efficient data transmission, improving overall network reliability and performance. This cooperative approach also boosts the network's ability to scale effectively, handling a larger number of nodes without a significant drop in throughput.

Security is a key focus of SEC-MAC, addressing vulnerabilities in WSNs by incorporating cryptographic measures and secure authentication processes to protect data integrity and prevent unauthorized access. The protocol ensures reliable communication even in potentially hostile environments. Simulation results demonstrated that SEC-MAC significantly improves network throughput as the number of

nodes increases compared to BTAC and IEEE 802.11b protocols. By optimizing relay selection and minimizing retransmissions, the protocol enhances energy efficiency, making it ideal for energy-constrained applications like environmental monitoring. Future research should focus on enhancing SEC-MAC's adaptability, performance, and security to expand its applicability. Future research will explore further optimization of SEC-MAC, including its performance in real-world applications like healthcare monitoring, as well as investigating the impact of different traffic models. Overall, SEC-MAC offers an effective solution for secure, energy-efficient, and scalable communication in WSNs, making it a promising protocol for a wide range of applications.

REFERENCES

- [1] K. E. Ukhurebor, I. Odesanya, S. S. Tyokighir, R. G. Kerry, A. S. Olayinka, and A. O. Bobadoye, "Wireless Sensor Networks: Applications and Challenges," *Wirel. Sens. Networks - Des. Deploy. Appl.*, Oct. 2020.
- [2] D. De, A. Mukherjee, S. K. Das, and N. Dey, "Wireless Sensor Network: Applications, Challenges, and Algorithms," *Nature inspired computing for wireless sensor networks*, pp. 1–18, 2020.
- [3] N. R. Patel and S. Kumar, "Wireless sensor networks' challenges and future prospects," *Proc. 2018 Int. Conf. Syst. Model. Adv. Res. Trends, SMART 2018*, pp. 60–65, Nov. 2018.
- [4] Karimi, A., Amini, S.M. Reduction of energy consumption in wireless sensor networks based on predictable routes for multi-mobile sink. *J Supercomput* 75, 7290–7313 (2019). <https://doi.org/10.1007/s11227-019-02938-y>
- [5] Manikandan, A., Venkataramanan, C. and Dhanapal, R., "A score based link delay aware routing protocol to improve energy optimization in wireless sensor network," *Journal of Engineering Research*, Vol. 13, pp.100115,2023.
- [6] P. Parwekar, S. Rodda, and N. Kalla, "A study of the optimization techniques for wireless sensor networks (WSNs)," *Adv. Intell. Syst. Comput.*, vol. 672, pp. 909–915, 2018.
- [7] A. Hossam, T. Salem, A. A. Hady, and S. Abd El-Kader, "Mca-mac: Modified cooperative access mac protocol in wireless sensor networks," *Int. Arab J. Inf. Technol.*, vol. 18, no. 3, pp. 326–335, 2021.
- [8] R. Shanker and A. Singh, "Analysis of Network Attacks at Data Link Layer and its Mitigation," *Proc. - 2021 Int. Conf. Comput. Sci. ICCS 2021*, pp. 274–279, 2021.
- [9] M. Boussif, "On The Security of Advanced Encryption Standard (AES)," *8th Int. Conf. Eng. Appl. Sci. Technol. ICEAST 2022 - Proc.*, pp. 83–88, 2022.
- [10] T. Azzabi, H. Farhat, and N. Sahli, "A survey on wireless sensor networks security issues and military specificities," *Proc. Int. Conf. Adv. Syst. Electr. Technol. IC_ASET 2017*, pp. 66–72, Jul. 2017.
- [11] S. Dhiviya, A. Sariga, and P. Sujatha, "Survey on WSN using clustering," in *2017 Second International Conference on Recent Trends and Challenges in Computational Models (ICRTCCM)*, pp. 121–125, IEEE, Tindivanam, India, February 2017.
- [12] M. K. Singh, S. I. Amin, S. A. Imam, V. K. Sachan, and A. Choudhary, "A survey of wireless sensor network and its types," in *2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pp. 326–330, IEEE, Greater Noida, India, October 2018.
- [13] R. E. Mohamed, A. I. Saleh, M. Abdelrazzak, and A. S. Samra, "Survey on wireless sensor network applications and energyefficient routing protocols," *Wireless Personal Communications*, vol. 101, no. 2, pp. 1019–1055, 2018
- [14] W. Y. Yi, K. M. Lo, T. Mak, K. S. Leung, Y. Leung, and M. L. Meng, "A survey of wireless sensor network based air pollution monitoring systems," *Sensors*, vol. 15, no. 12, pp. 31392–31427, 2015
- [15] M. H. Anisi, G. Abdul-Salaam, and A. H. Abdullah, "A survey of wireless sensor network approaches and their energy consumption for monitoring farm fields in precision agriculture," *Precision Agriculture*, vol. 16, pp. 216–238, 2015.
- [16] J. Kaur, T. Kaur, and K. Kaushal, "Survey on WSN routing protocols," *International Journal of Computer Applications*, vol. 109, no. 10, pp. 24–28, 2015.
- [17] A. J. AL-Mousawi and H. K. AL-Hassani, "A survey in wireless sensor network for explosives detection," *Computers & Electrical Engineering*, vol. 72, pp. 682–701, 2018.
- [18] A. Ali, Y. Ming, S. Chakraborty, and S. Iram, "A comprehensive survey on real-time applications of WSN," *Future Internet*, vol. 9, no. 4, Article ID 77, 2017.
- [19] B. Rashid and M. H. Rehmani, "Applications of wireless sensor networks for urban areas: a survey," *Journal of Network and Computer Applications*, vol. 60, pp. 192–219, 2016.
- [20] S. Abdollahzadeh and N. J. Navimipour, "Deployment strategies in the wireless sensor network: a comprehensive review," *Computer Communications*, vol. 91–92, pp. 1–16, 2016.
- [21] A. Belfkih, C. Duvallet, and B. Sadeg, "A survey on wireless sensor network databases," *Wireless Networks*, vol. 25, no. 8, pp. 4921–4946, 2019.
- [22] M. Sha fiq, H. Ashraf, A. Ullah, and S. Tahira, "Systematic literature review on energy efficient routing schemes in WSN — a survey," *Mobile Networks and Applications*, vol. 25, pp. 882–895, 2020.
- [23] J. Amutha, S. Sharma, and J. Nagar, "WSN strategies based on sensors, deployment, sensing models, coverage and energy efficiency: review, approaches and open issues," *Wireless Personal Communications*, vol. 111, pp. 1089–1115, 2020.
- [24] H. Sharma, A. Haque, and F. Blaabjerg, "Machine learning in wireless sensor networks for smart cities: a survey," *Electronics*, vol. 10, no. 9, Article ID 1012, 2021.
- [25] N. Temene, C. Sergiou, C. Georgiou, and V. Vassiliou, "A survey on mobility in wireless sensor networks," *Ad Hoc Networks*, vol. 125, Article ID 102726, 2022.
- [26] J. Paruvathavardhini and B. Sargunam, "Stochastic bat optimization model for secured WSN with energy-aware quantized index clustering," *Journal of Sensors*, vol. 2023, Article ID 4237198, 16 pages, 2023.
- [27] S. V. G. R. Nagarajan, and S. Kannadhasan, "Performance analysis of blended NIDS model for network intrusion detection system in WSN," in *2023 Fifth International Conference on Electrical, Computer and Communication Technologies (ICECCT)*, pp. 1–6, IEEE, Erode, India, February 2023.
- [28] J. Paruvathavardhini, B. Sargunam, and R. Sudarmani, "A review on energy efficient routing protocols and security techniques for wireless sensor networks," *Applied Mechanics and Materials*, vol. 912, pp. 55–75, 2023.
- [29] M. Hosseinzadeh, J. Yoo, S. Ali et al., "A cluster-based trusted routing method using fire hawk optimizer (FHO) in wireless sensor networks (WSNs)," *Scientific Reports*, vol. 13, Article ID 13046, 2023.
- [30] R. Dass, M. Narayanan, G. Ananthkrishnan et al., "A clusterbased energy-efficient secure optimal path-routing protocol for wireless body-area sensor networks," *Sensors*, vol. 23, no. 14, Article ID 6274, 2023.
- [31] Mainaud B., Gauthier V., and Afifi H., "Cooperative Communication for Wireless Sensors Network: A Mac Protocol Solution Cooperative Communication for Wireless Sensors Network: A Mac Protocol Solution WSC-MAC: A Cooperative Mac Protocol for Wireless Sensors Network," in *Proceedings of 1 st IFIP Wireless Days, Dubai*, pp. 1-5, 2008.
- [32] Liu K., Wu S., Huang B., Liu F., and Xu Z., "A Power-Optimized Cooperative MAC Protocol for Lifetime Extension in Wireless Sensor Networks," *Sensors*, vol. 16, no. 10, pp. 1630, 2016.
- [33] Nacef A., Senouci S., Ghamri-Doudane Y., and Beylot A., "COSMIC: A Cooperative MAC Protocol for WSN with Minimal Control Messages," in *Proceedings of 4 th IFIP International Conference on New Technologies, Mobility and Security, Paris*, pp. 1-5, 2011.

Digital Twin Model from Freehanded Sketch to Facade Design, 2D-3D Conversion for Volume Design

Kohei Arai

Dept. Information Science, Saga University, Saga City, Japan

Abstract—The article proposes a method for creating digital twins from freehand sketches for facade design, converting 2D designs to 3D volumes, and integrating these designs into real-world GIS systems. It outlines a process that involves generating 2D exterior images from sketches using generative AI (Gemini 1.5 Pro), converting these 2D images into 3D models with TriPo, and creating design drawings with SketchUp. Additionally, it describes a method for creating 3D exterior images using GauGAN, all for the purpose of construction exterior evaluation. The paper also discusses generating BIM data using generative AI, converting BIM data (in IFC file format) to GeoTiff, and displaying this information in GIS using QGIS software. Moreover, it suggests a method for generating digital twins with SketchUp to facilitate digital design information sharing and simulation within a virtual space. Lastly, it advocates for a cost-effective AI system designed for small and medium-sized construction companies, which often struggle to adopt BIM, to harness the advantages of digital twins.

Keywords—BIM; AI; GIS; digital twins; metaverse; generative AI; GauGAN; TriPo; SketchUp; IFC format; GeoTiff

I. INTRODUCTION

The Japanese construction industry urgently needs to improve work efficiency and productivity. This urgency arises from slow productivity growth, frequent industrial accidents, a declining and aging workforce, and a government policy reducing overtime work starting in 2024. While 31.1% of employees in all industries are aged 55 or older, the percentage for construction is over 36.0%. On the other hand, the percentage of employees aged 29 or younger is over 16.6% in all industries but less than 11.8% in construction.

To address these challenges, the Ministry of Land, Infrastructure, Transport and Tourism will implement the "Building Information Management (BIM)¹ Drawing Review" in the spring of 2026. This initiative will allow confirmation applications using PDF documents and BIM models, promoting efficiency in construction projects through AI, Geographic Information Systems (GIS), and BIM.

This paper focuses on improving construction efficiency using BIM, particularly the 6D BIM model. While many BIM tools are commercially available, they are often expensive. The proposed methods rely on open-source tools to create BIM

models, generate 2D and 3D exterior images, and convert BIM data into GIS-compatible formats like GeoTiff².

The paper highlights the advantages of BIM in terms of front-loading (shifting efforts earlier in the project lifecycle) and concurrent engineering (parallelizing workflows). It also explores methods for generating BIM data. Beyond the 3D BIM model, the benefits of 6D BIM—which incorporates time, cost, and sustainability—are discussed, especially when combined with a digital twin.

Additionally, the paper proposes methods to enhance efficiency and visualization in construction. AI is utilized to generate 3D models from 2D sketches, which can then be used to check building exteriors and appearances. The generated 3D models are converted into BIM data and subsequently into GeoTiff format for display in GIS tools like QGIS³. Open-source tools like SketchUp are also proposed for BIM data generation and 3D model visualization, along with AI-based construction management enhancements. A GauGAN⁴-based method for generating 3D exterior images is suggested to improve design review processes.

The integration of AI and digital twins is shown to simplify construction management tasks, improve efficiency and productivity, and enhance safety. By converting BIM models into GeoTiff format, the models can be optimized for specific geographical conditions at construction sites. There is no such this proposed models and methods which features generative AI, TriPo for creation of BIM models, and generate GIS model which is linked to the created BIM model.

The paper reviews related research and discusses the effectiveness of using BIM and AI for model creation. It introduces methods for creating BIM models from 2D data using generative AI, as well as SketchUp-based techniques. A TriPo⁵ method is proposed for converting 2D models to 3D models. Finally, the paper explores integrating BIM models into Metaverse and discusses a conversion tool for translating BIM data into GeoTiff format. The conclusion summarizes these contributions with remarks and further discussion.

II. RELATED RESEARCH WORKS

Related Research on Generative AI in Construction:

¹ <https://ja.wikipedia.org/wiki/BIM>

² <https://ja.wikipedia.org/wiki/GeoTIFF>

³ <https://qgis.org/>

⁴ <https://blogs.nvidia.co.jp/blog/what-is-gaugan-ai-art-demo/>

⁵ <https://www.tripo3d.ai/>

A framework integrating BIM-data mining with digital twin technology for advanced project management in smart construction has been proposed [1].

Generative Adversarial Networks (GAN⁶s) for construction project management have also been explored, showcasing how GANs can improve project management processes [2].

TriPo (Triplet Network) in AI Applications:

FaceNet⁷: A Unified Embedding for Face Recognition and Clustering introduces a Triplet Network for facial recognition and clustering [3].

Triangulated Irregular Network⁸ (TIN) in GIS: A Review highlights the use and benefits of TINs in GIS applications [4].

GIS and QGIS in Construction Management:

Foundational texts, including Geographic Information Systems and Science [5] and Open-Source GIS: A GRASS GIS⁹ Approach [6], provide comprehensive insights into GIS applications.

A study on using QGIS for geospatial analysis in construction projects demonstrates its practical applications [7].

Digital Twin Technology in Construction:

Digital Twin: Enabling Technologies, Challenges, and Open Research outlines the key technologies and challenges in this area [8].

Digital Twin in Construction: A Systematic Review provides a detailed overview of the benefits and applications of digital twins in construction [9].

SketchUp in Architectural Design:

While The Unified Modeling Language User Guide provides general modeling insights, Using SketchUp for Architectural Design and Construction Documentation focuses on SketchUp's role in creating designs and documentation [10][11].

GauGAN and AI in Creative Design:

GauGAN: Semantic Image Synthesis with Spatially Adaptive Normalization and GauGAN: Semantic Image Synthesis with Spatially Conditioned Generative Adversarial Networks explain the principles and applications of GauGAN in creative design [12][13].

BIM in Construction:

BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers, and Contractors provides a comprehensive overview of BIM's benefits and challenges [14][15].

BIM in Construction Projects: Benefits and Challenges discusses how BIM improves construction processes and highlights associated obstacles [15].

III. PROPOSED METHODS AND SYSTEMS

A. Overall

Improving productivity and streamlining workflows in the construction industry can be achieved not only through BIM-based design but also by leveraging ICT tools like IoT-enabled surveying and construction machinery. By incorporating AI into pre-construction tasks, such as building BIM models, parallel and collaborative work becomes possible, reducing construction time.

Additionally, using tools like RAG¹⁰ (Retrieval-Augmented Generation) to share knowledge and experience can enhance work efficiency during construction. This facilitates the transfer of skills from experienced workers to beginners, improving overall quality. It supports front-loading, a method that allocates resources and focuses efforts early in a project. This approach enables early detection and resolution of problems, such as structural interferences or design flaws, by creating 3D models during the initial stages.

Front-loading minimizes rework caused by design errors or clashes, improving quality and reducing costs. Early detailed considerations and validations improve design accuracy, and simulations using 3D models allow for better decision-making. Reduced rework lowers overall costs, shortens construction times, and accelerates decision-making by enhancing stakeholder communication through 3D models. Furthermore, construction plans become more optimized, as procedures and temporary methods can be rationalized at the design stage. Distributing workloads more evenly across project phases also improves overall efficiency.

The proposed workflow is illustrated in Fig. 1:

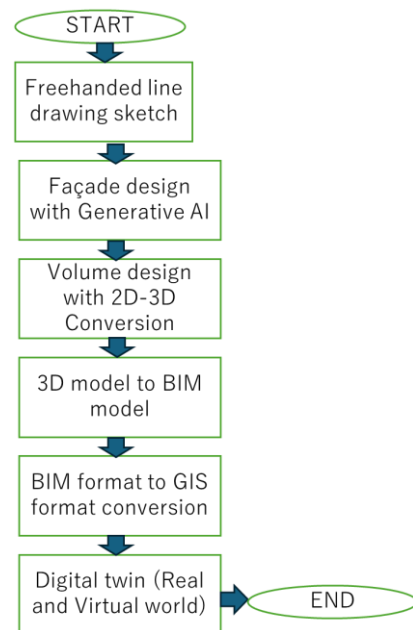


Fig. 1. Process flow of the proposed method for digital twin model creation.

⁶ <https://www.skillupai.com/blog/tech/ml-dl-tips-1/>

⁷ <https://github.com/davidsandberg/facenet>

⁸ <https://ja.wikipedia.org/wiki/TIN>

⁹ <https://grass.osgeo.org/>

¹⁰ <https://aws.amazon.com/jp/what-is/retrieval-augmented-generation/>

- 1) Start with a freehand sketch.
- 2) Use software like SketchUp¹¹ to refine the design.
- 3) Create a 2D façade design using generative AI tools, such as Gemini AI Pro¹².
- 4) Convert the 2D design into a 3D volume model using 2D-to-3D tools like TriPo.
- 5) The 3D model can be used as a BIM model in a virtual space.

To create BIM models in real-world applications, BIM data (e.g., IFC files) is converted to GeoTiff format, enabling GIS visualization with tools like QGIS. These steps outline a proposed digital twin creation method for construction design.

In the virtual space, users can perform various simulations, and the results of these studies can be implemented in real-world applications through GIS-based visualization.

B. SketchUp¹³

Revit¹⁴, ArchiCAD¹⁵, and Gloobe¹⁶ are popular tools for generating BIM data but come with high costs: 427,900 yen, 418,000 yen, and 165,000 yen per month, respectively. A more affordable alternative is SketchUp, which is free but offers fewer features. Here's how SketchUp can be used for BIM data generation:

1) Basic Use of SketchUp

Modeling: Create 3D models using SketchUp's intuitive interface.

Use basic tools like lines, planes, and arcs to model buildings and structures.

2) BIM Data Requirements

BIM Data: Includes building structure, materials, dimensions, and related metadata.

In SketchUp, you'll need to create a detailed model with this information.

3) Using Plugins

IFC Exporter¹⁷: Export SketchUp models as IFC (Industry Foundation Classes) files, the standard format for BIM data.

Plugin Installation: Search for "IFC Exporter" in SketchUp's Extension Warehouse and install it to improve compatibility with BIM software.

4) Adding Attributes and Metadata

Dynamic Components: Add attributes and parameters to objects, such as size, material, or manufacturer details for doors and windows.

5) Using Layers and Groups

Layers and Groups: Categorize elements into layers and groups for better organization, making it easier to identify and work with specific elements.

6) Exporting and Importing

Export IFC Files: Use the IFC Exporter plugin to export SketchUp models as IFC files.

Import into BIM Software: Import these IFC files into BIM tools like Revit or ArchiCAD for further analysis and design.

7) Integration with Other Tools

SketchUp Pro and Layout: Use Layout, a feature in SketchUp Pro, to create 2D documents for construction drawings and specifications as part of the BIM workflow.

8) Best Practices

Standardize Model Structure: Use consistent layer structures and naming conventions to improve collaboration and make models easier to understand.

While SketchUp is not a dedicated BIM tool and primarily serves as 3D modeling software, it can still be used to create detailed models and ensure compatibility with other BIM software through the steps above. However, its BIM functionalities are limited compared to specialized tools.

Fig. 2 shows an example of an initial building design created with SketchUp.

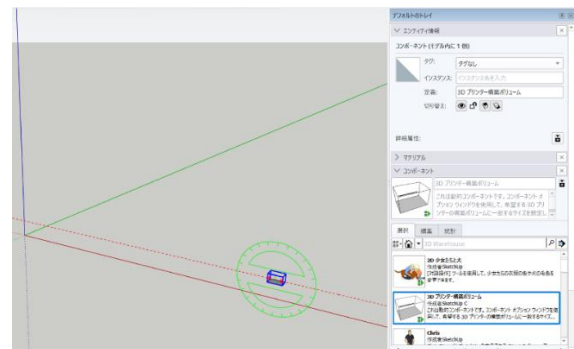


Fig. 2. Just the beginning of design of building for construction.

C. Generative AI and TriPo

For exterior design checks, images generated by Generative AI can be highly useful. Fig. 3 illustrates a 2D exterior design image created using Claude 3.5 Sonnet¹⁸, a Generative AI tool. This 2D image can then be converted into a 3D model using tools like TriPo.

While there are numerous software options available for 2D-to-3D image conversion, TriPo demonstrated the best quality among the 10 tools tested as is shown in Fig. 4. 3D images converted from 2D image generated by Generative AI based on TriPo can be rotated and can be displayed from the arbitrary view of line of sight.

¹¹ <https://www.sketchup.com/en>

¹² <https://deepmind.google/technologies/gemini/pro/>

¹³ <https://help.sketchup.com/ja/downloading-sketchup>

¹⁴ <https://recademy.jp/knowhow/4373>

¹⁵ <https://graphisoft.com/jp/solutions/products/archicad>

¹⁶ <https://archi.fukuicompu.co.jp/products/gloobe/>

¹⁷ <https://github.com/Autodesk/revit-ifc/releases>

¹⁸ <https://www.anthropic.com/news/claude-3-5-sonnet>



Fig. 3. 2D exterior designed image creation with generative AI of Claude 3.5 sonnet.

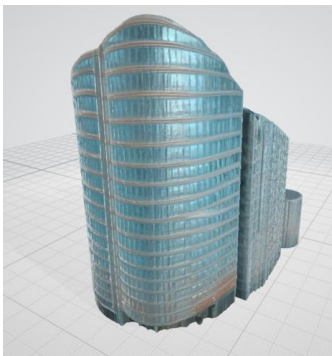


Fig. 4. 3D image converted from 2D image generated by Generative AI based on TriPo.

D. GauGAN

GauGAN is a powerful tool for visualizing architectural concepts and designs. Here are its key applications:

1) *Concept visualization*: Architects and designers can use GauGAN to quickly bring initial ideas to life. From simple sketches or written descriptions, it generates realistic architectural images that can be shared with clients and team members.

2) *Environmental simulations*: GauGAN can simulate buildings in various environments and seasons, helping visualize how a project will look under different conditions. For example, it can easily create scenes like “a modern house in snowy mountains” or “a skyscraper in a bustling city center.”

3) *Landscape design*: Designers can use GauGAN to experiment with surroundings, such as planting options or terrain layouts. It enables easy addition or modification of elements like green spaces, water features, or rock gardens, ensuring harmony between the building and its environment. GIS tools can further assist in evaluating this harmony.

4) *Presentation materials*: Realistic images generated by GauGAN are ideal for client presentations or proposal materials. These visuals help communicate the project vision effectively and leave a strong impression.

5) *Rapid design iteration*: GauGAN allows designers to make quick adjustments to their designs and view results in real

time. This speeds up the design process and encourages the exploration of multiple ideas in a shorter timeframe.

6) *Urban planning*: Urban planners can use GauGAN to visualize new developments or redevelopment projects. By combining elements like streets, parks, and buildings, it can depict a detailed vision of a city’s future.

7) *Interior design*: GauGAN is also helpful for designing interior spaces. It can simulate natural light effects, window views, and overall spatial aesthetics, supporting better decision-making in interior design.

These capabilities make GauGAN a valuable tool for generating ideas in the early stages of architectural design, enhancing client communication, and enabling faster design iterations. By improving both efficiency and creativity, GauGAN significantly supports architectural and urban design projects.

E. BIM Model Creation by Claud 3.5 Sonnet

Create a structured BIM dataset based on what can be observed from the images, paying attention to:

- 1) Overall properties and characteristics of the building
- 2) Details of the structural system
- 3) Spatial configuration and estimated dimensions, especially of the characteristic curved curtain walls
- 4) Envisaged technical systems
- 5) Sustainability features
- 6) Contextual information

The Python code for this is as follows:

```
{
  "buildingProperties": {
    "type": "Mixed-Use High-Rise",
    "architecturalStyle": "Contemporary",
    "mainFeatures": ["Curved Glass Facade", "Corner Location"],
    "estimatedHeight": "65-75 meters",
    "estimatedFloors": {
      "aboveGround": 18,
      "retail": 1
    }
  },
  "structuralSystem": {
    "primaryStructure": "Steel and Concrete",
    "facade": {
      "type": "Curtain Wall",
      "materials": ["Glass", "Aluminum"],
      "features": ["Curved Glass Panels", "Horizontal Banding"]
    }
  },
  "spatialOrganization": {
    "footprint": {
      "shape": "Curved Rectangle",
```

```
"estimatedArea": "800-1000 square meters"
},
"floorLayout": {
  "groundFloor": {
    "use": "Retail/Lobby",
    "features": ["Double Height Entrance", "Commercial Space"]
  },
  "typicalFloor": {
    "use": "Commercial/Residential",
    "estimatedArea": "700-900 square meters"
  }
},
"technicalSystems": {
  "presumedMEP": {
    "hvac": "Central System",
    "elevators": {
      "estimated": "3-4 units",
      "type": "High-Speed"
    },
    "lighting": {
      "natural": "Floor-to-ceiling windows",
      "artificial": "LED Systems"
    }
  },
  "sustainability": {
    "features": [
      "High-Performance Glazing",
      "Natural Daylighting",
      "Possible Green Building Certification"
    ]
  },
  "context": {
    "setting": "Urban",
    "surroundings": {
      "adjacent": ["Street", "Parking Area"],
      "neighborhood": "Commercial District"
    }
  },
  "metadata": {
```

"note": "This BIM data is approximate and based on visual assessment only. Actual building specifications may vary."

```
}
}
```

F. Other Useful Software Tools

The Architecture Design Bot¹⁹ is an AI-powered tool that streamlines the architectural design process. It can generate architectural images from text prompts, create variations of existing design proposals, automatically produce 3D models, and provide optimization suggestions for compliance with regulations and cost-efficiency. Additionally, it aids in idea generation during the early design stages, quickly generates client-ready proposal variations, reduces the time needed for rendering and 3D modeling, and automates compliance checks with building codes.

Other Free Tools for Generating BIM Data:

1) FreeCAD

Open-source 3D parametric modeler.

Create 3D drawings from 2D sketches.

Compatible with Windows, Mac, and Linux.

BIMx (Graphisoft)

2) Free applications are available for iOS and Android.

View and explore 3D BIM models.

Provides seamless 2D and 3D project navigation.

BIM Vision

3) IFC model viewer.

Supports models created in various systems like Revit, ArchiCAD, and Tekla.

4) B-processor²⁰

Tool specifically developed for BIM.

Intuitive and easy-to-learn 3D modeling.

Includes features for cost calculations and carbon emission data.

5) Edificius (ACCA Software)²¹

Free BIM software with real-time rendering capabilities.

Integrates structural analysis with architectural design.

Enables accurate land mapping using satellite imagery from Google Maps.

Free 30-day trial available.

¹⁹ <https://prod.d2eu75mpuy425r.amplifyapp.com/>

²⁰ https://tracxn.com/d/companies/b-processor/_atQDLzH2udrUrz2ES_zzho_gxGXMFnkxeLmKHa-s0Us#competitors-and-alternates

²¹ <https://www.accasoftware.com/en/trial/edificius>

These tools offer a range of capabilities to support BIM workflows, from basic modeling to advanced features like rendering, analysis, and environmental data integration.

G. Digital Twin

The benefits of using BIM and digital twins together include the following:

1) *Real-time monitoring and predictive analysis*: Digital twins integrate real-time data into 3D models created with BIM/CIM, allowing real-time monitoring of construction project progress. This helps identify and address issues early. Additionally, AI-driven predictive analysis enables proactive identification of potential problems and determination of optimal construction methods.

2) *Optimization of construction processes*: Data collected via digital twins can be reflected in BIM/CIM models to continuously improve construction workflows. For example, analyzing equipment usage and worker movements can lead to more efficient procedures.

3) *Enhanced safety*: The combination of digital twins for real-time monitoring and detailed 3D BIM/CIM models allows for the early detection of potential safety hazards, enabling preventive measures to reduce risks.

4) *Improved maintenance efficiency*: After construction, digital twins can monitor the condition of buildings and infrastructure in real time, reflecting updates in the BIM/CIM model. This enables preventive maintenance and optimized repair planning.

5) *Better collaboration*: By integrating digital twins with BIM/CIM, all stakeholders can share up-to-date project information, improving communication and collaboration.

6) *Advanced simulation capabilities*: Incorporating real-time data from digital twins into BIM/CIM models facilitates more precise simulations, allowing for accurate predictions of design changes' impacts and optimization of construction methods.

7) *Improved lifecycle management*: Continuous data collection and utilization across the design, construction, operation, and maintenance phases enable optimization throughout the entire lifecycle of buildings and infrastructure.

8) *Faster and more accurate decision-making*: Combining real-time data with detailed 3D models enhances decision-making speed and accuracy.

The integration of digital twins and BIM/CIM significantly improves visualization, efficiency, and quality in construction projects. It also plays a crucial role in advancing digital transformation in the construction industry. To fully leverage these technologies, establishing robust data management systems and enhancing the skills of personnel involved are essential.

H. BIM File (IFC Format) to GIS File (GeoTiff Format)

"BIM file workspace" (Revit or IFC file) to geodatabase dataset base map information DEM 1m mesh²². Assign an appropriate geographic coordinate system to the BIM model. In Japan, JGD2011 (plane rectangular coordinate system) is often used. Extract necessary information from the BIM model (e.g. topographical data, building outline, etc.). Convert the extracted data into a raster format (GeoTIFF, etc.).

The method of converting BIM data to GeoTiff that can be displayed in GIS such as QGIS is as follows:

1) *AutoCAD Civil 3D*: Read BIM data and export to GIS format

2) *FME*²³ (*Feature Manipulation Engine*): Tool specialized in conversion between various formats

3) *QGIS*: Open-source GIS software that allows conversion using plug-ins

4) *Resolution setting*: Set the appropriate resolution when converting to GeoTIFF

5) *Attribute information retention*: Reflect important attribute information contained in the BIM model in the GIS data

6) Utilize detailed 3D models created with BIM for geospatial analysis (promoting integration of BIM and GIS, enabling more effective data utilization at each stage of construction project planning, design, construction, and maintenance)

Below is the Python code that makes this possible. First, convert the BIM data to an intermediate format (e.g. GeoJSON²⁴) and then convert it from the intermediate format to GeoTIFF. The code works in the following steps:

1) The `bim_to_geojson` function reads the BIM data and converts it to GeoJSON format

2) Assumes that the input is already in GeoJSON format
If it uses real BIM data (e.g. IFC files), it will need to parse the data using an appropriate library (e.g. `IfcOpenShell`) and convert it to GeoJSON format.

Next, the `geojson_to_geotiff` function converts the GeoJSON data to a GeoTIFF. This function does the following:

1) Calculates the extent of the GeoJSON data
2) Determines the raster size based on the specified resolution

3) Creates an empty raster using NumPy
4) Draws each feature in the GeoJSON to the raster
5) Saves as a GeoTIFF using the rasterio library

To run this code, the following libraries are required:

```
rasterio  
numpy  
shapely
```

²² <https://www.gsi.go.jp/gazochosa/gazochosa61002.html>

²³ [https://en.wikipedia.org/wiki/FME_\(software\)](https://en.wikipedia.org/wiki/FME_(software))

²⁴ <https://ja.wikipedia.org/wiki/GeoJSON>

These libraries can be installed with the following command of “pip install rasterio numpy shapely”. The Python code for this is as follows:

```
import json
import rasterio
from rasterio.transform import from_bounds
import numpy as np
from shapely.geometry import shape
from shapely.affinity import scale

# Step 1: Function to load BIM data and convert to GeoJSON
def bim_to_geojson(bim_file):
    # Implement the code to load and analyze BIM data here
    # In this example, it assumes that the data is already in GeoJSON format
    with open(bim_file, 'r') as f:
        geojson_data = json.load(f)
    return geojson_data

# Step 2: Function to convert GeoJSON to GeoTIFF
def geojson_to_geotiff(geojson_data, output_file, resolution=0.1):
    # Get GeoJSON extent
    features = geojson_data['features']
    geometries = [shape(feature['geometry']) for feature in features]
    all_geoms = shape({'type': 'MultiPolygon', 'coordinates': [geom.coordinates
for geom in geometries]})
    minx, miny, maxx, maxy = all_geoms.bounds
    # Calculate raster size
    width = int((maxx - minx) / resolution)
    height = int((maxy - miny) / resolution)
    # Create raster data
    raster = np.zeros((height, width), dtype=np.uint8)
    # Draw each feature in GeoJSON to a raster
    for feature in features:
        geom = shape(feature['geometry'])
        geom = scale(geom, xfact=1/resolution, yfact=1/resolution, origin=(minx,
miny))
        coords = np.array(geom.exterior.coords).astype(int)
        rasterio.features.rasterize([coords], out=raster,
transform=from_bounds(minx, miny, maxx, maxy, width, height))

# Save as GeoTIFF
with rasterio.open(
    output_file,
    'w',
    driver='GTiff',
    height=height,
    width=width,
    count=1,
    dtype=raster.dtype,
    crs='+proj=latlong',
```

```
transform=from_bounds(minx, miny, maxx, maxy, width, height),
) as dst:
    dst.write(raster, 1)

# Main process
bim_file = 'input.json' # BIM data file (assuming GeoJSON format in this
example)
output_file = 'output.tif'
geojson_data = bim_to_geojson(bim_file)
geojson_to_geotiff(geojson_data, output_file)
print(f"Conversion completed. Output file: {output_file}")
```

IV. CONCLUSION

This study proposes a method for generating 2D exterior images from sketches using generative AI (Gemini 1.5 Pro) and a technique for converting 2D images into 3D models using the free tool TriPo. Additionally, it demonstrates how 3D models can be created using the free tool GauGAN, making it easier to visualize and verify building exteriors.

The study also presents a process for creating design drawings with the free tool SketchUp and generating BIM data using generative AI (Claude 3.5 sonnet). While the BIM data generated may not be fully complete, the findings demonstrate the feasibility of using free tools for this purpose.

Furthermore, a method is proposed for converting BIM data (IFC files) into GeoTiff format using Python code, illustrating that GIS visualization can be achieved with free tools like QGIS. Lastly, the study suggests a method for generating digital twins using SketchUp and shows that simulations in virtual environments are possible, enhancing the scope of design and analysis in construction projects. This approach is quite new and has original ideas. The conventional models and methods are traditional expensive BIM model creation and are not linked to the GIS system at all.

A. Future Research Works

Although it is confirmed that the proposed methods and systems can be feasible, detailed design of BIM model requires more detailed information. Also, CG design and environmental parameter settings are required for creation of digital twin, fundamental digital twin can be created by the proposed method though. Therefore, one of the business use cases will be attempted in the near future.

REFERENCES

- [1] Pan, Y., & Zhang, L. (2021). A BIM-data mining integrated digital twin framework for advanced project management in smart construction. *Automation in Construction*, 124, 103564.
- [2] Zhang, Y., & Chen, J., "Generative Adversarial Networks for Construction Project Management", *Journal of Construction Engineering and Management*, DOI: 10.1061/(ASCE)CO.1943-7862.0002245, 2022.
- [3] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A Unified Embedding for Face Recognition and Clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 815–823), 2015.
- [4] Singh, R., & Singh, D., "Triangulated Irregular Network (TIN) in GIS: A Review", *Journal of Geographic Information System*, DOI: 10.4236/jgis.2019.113016, 2019.
- [5] Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2015). *Geographic Information Systems and Science*. Wiley, 2015.

- [6] Mitasova, H., Neteler, M., & Metz, M. (2018). Open Source GIS: A GRASS GIS Approach. Springer, 2018.
- [7] Koeva, M., & Mladenov, V., Using QGIS for Geospatial Analysis in Construction Projects, International Conference on Geoinformatics, DOI: 10.1109/GeoInformatics48762.2020.9200803, 2020.
- [8] Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital twin: Enabling technologies, challenges and open research. In IEEE Access, 8, 108952-108971, 2020.
- [9] Liu, X., & Wang, Y., Digital Twin in Construction: A Systematic Review, Automation in Construction, DOI: 10.1016/j.autcon.2022.104362, 2022.
- [10] Booch, G., Rumbaugh, J., & Jacobson, I. (2010). The Unified Modeling Language User Guide. Addison-Wesley Professional (This provides insight into modeling, though not specifically about SketchUp, provides context for design tools), 2010..
- [11] Wong, J., & Wong, P., Using SketchUp for Architectural Design and Construction Documentation, Architectural Design and Construction Technology, Springer, ISBN: 978-981-10-7514-4, 2018.
- [12] Park, T., Liu, M. Y., Wang, T. C., & Zhu, J. Y. (2019). GauGAN: Semantic Image Synthesis with Spatially-Adaptive Normalization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
- [13] Park, T., Liu, M.-Y., Wang, T.-C., & Zhu, J.-Y., GauGAN: Semantic Image Synthesis with Spatially Conditioned Generative Adversarial Networks, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), DOI: 10.1109/CVPR.2019.00873, 2019.
- [14] Eastman, C., Teicholz, P., Sacks, R., & Liston, K. (2011). BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers, and Contractors. Wiley, 2011.
- [15] Eastman, C., Teicholz, P., Sacks, R., & Liston, K., Building Information Modelling (BIM) in Construction Projects: Benefits and Challenges, BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors, Wiley-Blackwell, ISBN: 978-1119286627, 2018.

AUTHOR'S PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 also was with National Space Development Agency of Japan from January 1979 to March 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science in April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is a Science Council of Japan Special Member since 2012. He is an Adjunct Professor at Brawijaya University. He also is an Award Committee member of ICSU/COSPAR. He also is an adjunct professor of Nishi-Kyushu University and Kurume Institute of Technology Applied AI Research Laboratory. He wrote 119 books and published 728 journal papers as well as 569 conference papers. He received 98 of awards including ICSU/COSPAR Vikram Sarabhai Medal in 2016, and Science award of Ministry of Mister of Education of Japan in 2015. He is now Editor-in-Chief of IJACSA and IJISA. <http://teagis.ip.is.saga-u.ac.jp/index.html>

Marked Object-Following System Using Deep Learning and Metaheuristics

Ken Gorro¹, Elmo Ranolo², Lawrence Roble³, Rue Nicole Santillan⁴, Anthony Ilano⁵, Joseph Pepito⁶,
Emma Sacan⁷, Deofel Balijon⁸

College of Technology, Cebu Technological University, Carmen, 6000, Cebu, Philippines^{1,2,3,4,5}

College of Technology, Cebu Technological University, Cebu, Philippines^{6,7}

Center for Cloud Computing, Big Data, and Artificial Intelligence^{1,2,3,4}

College of Computing, Artificial Intelligence, and Sciences, Cebu Normal University, Cebu, Philippines⁸

Abstract—This paper presents a deep learning methodology for a marked object-following system that incorporates the YOLOv8 (You Only Look Once version 8) object identification model and an inversely proportional distance estimation algorithm. The primary aim of this study is to develop a marked object-following algorithm capable of autonomously tracking a designated marker while maintaining a suitable distance through advanced computer vision techniques. In this study, a marked object is defined as an object that is explicitly labeled, tagged, or physically marked for identification, typically using visible markers such as QR codes, stickers, or distinct added features. Central to the system's functionality is the YOLOv8 model, which detects objects and generates bounding boxes around identified target classes in real-time. The proposed marked object-following algorithm utilizes the distance estimation method, which leverages fluctuations in the bounding box width to determine the relative distance between the observed user and the camera. A pathfinding algorithm was created using tabu search and a-star to avoid obstacle and generate a path to continue following the marker object. Furthermore, the system's efficacy was assessed using critical performance metrics, including the F1-score and Precision-Recall. The YOLOv8 model attained an F1-score of 0.95 at a confidence threshold of 0.461 and a mean Average Precision (mAP) of 0.961 at an IoU threshold of 0.5 for all target classes. These results indicate a high level of accuracy in object detection and tracking. However, it is important to note that this algorithm has close door and controlled environments.

Keywords—Object detection; YOLOv8; distance estimation; A-star; tabu search

I. INTRODUCTION

In today's technology-driven world, artificial intelligence (AI) and robotics are revolutionizing various domains, including human interaction and navigation. Autonomous systems capable of tracking and following individuals are highly beneficial in settings such as crowded environments, warehouses, and other dynamic areas. These systems have the potential to enhance efficiency and safety by providing precise and adaptive navigation in real-time. Recent studies have highlighted the advancements in deep learning techniques, particularly in object detection, which significantly improve the capabilities of these systems in complex environments [1][2][3]. The motivation behind this research arises from challenges faced in environments where autonomous systems must reliably track a marked object, particularly in dynamic and crowded areas. Traditional systems have frequently struggled to effectively identify and focus on the correct item in such situations

due to occlusions, competing visual elements, and contextual complications [4][5]. These constraints underscore the need for a stronger and more efficient system.

This study introduces a marked object-following algorithm that integrates deep learning and metaheuristic techniques to address these challenges. Leveraging the YOLOv8 (You Only Look Once version 8) object detection framework, the system ensures robust object recognition and tracking. YOLO has been recognized for its ability to perform real-time object detection with high accuracy, making it suitable for dynamic environments [6] [7]. A distance estimation algorithm based on fluctuations in bounding box width enables the system to maintain an optimal distance from the marked object, thereby preventing collisions. Furthermore, the inclusion of A-star pathfinding and the Tabu Search metaheuristic algorithm enhances the system's ability to navigate around obstacles and generate efficient paths in real-time scenarios [8] [9].

The development of a reliable and effective marked object-following system that can track a designated marker on its own while adjusting to environmental changes is the main goal of this research. A marked object is defined as an object that is explicitly labeled, tagged, or physically marked for identification, typically using visible markers such as QR codes, stickers, or distinct added features. This project aims to use YOLOv8 to create a reliable marked object-following algorithm for object tracking and detection in real-time. Additionally, the design and implementation of a distance estimation method that dynamically calculates the relative distance between the observed user and the camera are crucial components of this research. The integration of pathfinding algorithms, such as A-star and Tabu Search, enables obstacle avoidance and efficient navigation [8] [10].

Furthermore, this research is significant because it addresses the growing need for intelligent and adaptive tracking systems in real-world applications. By combining advanced deep learning models with metaheuristic algorithms, the proposed system offers a novel solution that ensures accuracy, reliability, and adaptability. The findings of this study aim to contribute to the advancement of autonomous tracking technologies, paving the way for their deployment in diverse practical scenarios [11] [2]. The integration of deep learning techniques in object detection has shown promising results, enhancing the performance of tracking systems in complex environments [2] [12].

II. RELATED STUDIES

In the past few years, LiDAR technology has become a key way of detecting people in robotic systems. Some researchers have been employing LiDARs for precise individual detection and tracking using the ability to measure the distance and location of the target person versus the robot. LiDAR sensors provide 3D data with high resolution, which allows robots to identify and follow a specific human target who is even moving in a dynamic environment [13], [14], [15]. Using LiDAR as the only source of environmental information is a highly unique task, and there has been little study in this area. Some human detection and tracking research has relied only on LiDAR technology. Human detection using LiDAR has been performed on both stationary robots [16], [17], employing several stationary LiDAR sensors [18], and mobile robots [19], [20].

On the other hand, researchers use machine learning for human detection and finding the robot. Machine learning has turned out to be key in providing a human-tracking robot's potential. The presented procedures use enormous amounts of data for training models that would notice the human features and movements, thus finding the correct identification in different scenes. Besides, the machine learning methods are used to help with robot localization by interpreting sensor data, thus finding the position of the robot in relation to the discovered human. Furthermore, this part will elaborate on the different machine-learning methods utilized for human detection and robot localization in these systems along with their advantages, and show the elements of integration [21], [22], [23]. Suet Peng Yong et al. [24] demonstrate human object recognition using deep learning algorithms with the use of a 3DR solo drone equipped with a GoPro camera for real-time surveillance and coverage of forest areas. Suet Peng Yong et al. provide knowledge of video processing using convolution neural networks and how to select the perfect dataset for a specific project.

Ashish U. Bokade et al. [25] discuss video surveillance utilizing a smartphone and Raspberry Pi. This allows you to watch and control the mobility of the robot using Raspberry Pi. The detection procedure may be completed successfully, and the findings can be viewed on the user's smartphone. Jun Zhang et al. [26] provide leaping robot standards, which are superior to traditional robots that cannot walk on rough surfaces or jump to a greater distance. It describes how a PIR sensor and a jumping robot build a zig-bee WSN that allows them to communicate with one another while also allowing the freedom to leap on stairs to reach higher surfaces from the ground up to a range of 105 cm.

Additionally, other researchers use OpenCV as the solution for human-tracking robots, which are used to track the human target during different movements while ensuring a constant distance between the human target and the robot. OpenCV (Open Source Computer Vision Library) is a set of versatile tools for real-time computer vision and it is actually a quite nice alternative for implementing human tracking in robotics. Researchers can write algorithms to detect and follow the humans by OpenCV that direct the robot to be safe and as close as possible to the optimal. In this section, OpenCV implementation in human-following robotic systems such as a detailed overview of its strengths, challenges, and its function

in giving the robots better agility and navigation precision will be explored [25], [26], [27], [28], [29], [30].

Meanwhile, color-based detections for target-following robots have been used by certain researchers since they are one of the possible good approaches to identifying a target, as demonstrated by researchers in Sefat S. et al. and MNA Bakar et al. S. Sefat et al. employed red color and 3D circular shape (red ball) detection, along with a Kalman filter, to predict the position of the individual to be followed. Although MNA Bakar et al. employed color-based detection, it used a special marker with a distinct form to help the robot recognize its target. The yellow hue was tested and had an 80% detection rate. However, MNA Bakar et al. had no obstructions in its route, as opposed to S. Sefat et al., which avoided obstacles while employing sonar sensors [31], [32].

Moreover, the investigation of the human body temperature through a thermal image in real-time is a well-known application of infrared technology by other researchers. The safety and security of a particular place, such as a train station, can be increased by the technology of human presence detection. As a result, the detector is a passable of sensors and a microcontroller. The detector can know the distance between the human and the reference point by using a camera. Sensor equipment is used in the way automated systems of various kinds are employed for people monitoring and various other applications. Infrared sensors have also been used to determine the human walking path. Thanks to such a device, robots can easily generate an exact following motion toward the human that they accompany. By defining the person's thermal footprint, robots can undertake good and continuous surveillance, which serves as a sufficient mechanism for follow-up. A detailed examination of the infrared technology application in robotic systems will be given in this section, including its advantages in human detection, movement predictions, and the overall enhancement of human-following robotic behaviors will be outlined [33], [34], [35], [36].

According to Montiel-Ross *et al.* [37], world perception, path planning and generation, and path tracking make up the robot navigation challenge. By choosing adequate sensor suites that can give the robot controller acceptable environmental feedback, world perception is achieved. Simultaneous Localization and Mapping (SLAM), a well-established technique that allows vision-based imaging equipment to visualize the surrounding depth map, is one of the best solutions for this purpose. Processing of this data can yield the locations of obstacles, targets, and the robot itself. The disadvantage is that in environments with unclear structures, SLAM performs less well [38]. In addition to being computationally demanding, SLAM has significant processing expenses [39]. According to Nowicki *et al.* [40], sudden and erratic motions of its sensing devices also cause SLAM to malfunction. Given that all of the CARMi sensors are installed on the same mobile platform, an incomplete mapping approach might be more appropriate. Al Arabi *et al.* [41] showed that partial mapping may be accomplished with just one rotating rangefinder by converting the data into a relative depth image of the surrounding area. The revolving depth camera configuration on CARMi may make this method useful.

The study of path planning and generation has a long history, and both heuristic and classical methods are still

widely used to prevent collisions and arrive at target locations. The robot that follows a human adds another level of complexity by needing to approach a moving target while keeping a set distance behind it. The path planning system is either “passive” or “anticipative,” according to Ziyou Wang *et al.* [42]. An “anticipative” system uses a velocity model [42] or a dynamically updated version of the Monte Carlo algorithm [39], [43] to predict the possible movements of a human target. Kalman filters, neural networks, fuzzy logic, and similar combinations [44], [45]. Since these techniques also come with high processing costs, it could be preferable for the CARMI navigation model to be “passive,” in which the robot reacts to changes in its surroundings or landmarks in a reactive manner [46].

Additionally, some other researchers utilize depth cameras as well as a selective set of limited proximity sensors. The overall approach of intelligent systems like robots consists of exact, desired-oriented human-tracking algorithms, which keep the robot on the right path and in the right direction minimizing the motion needed. The depth camera will be another evolutionary change in enabling the precise tracking of human targets in a 3D environment, whereas lasers give humans feedback on how far they are. This technology offers robots conventional control functions, like endpoint settings which provide machine-to-human interfacing inside factories or assembly facilities. This part will go deep in the analysis of the use of depth cameras and sensor fusion in these robotic systems for following humans, mainly by the help of them in improving target tracking and obstacle navigation [47], [48], [49].

III. METHODOLOGY

A. Methodology Overview

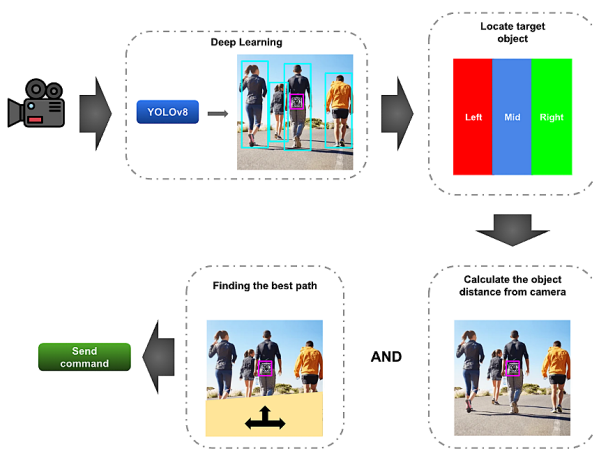


Fig. 1. Conceptual framework.

Fig. 1, titled “Conceptual Framework,” depicts the general methodology of the Marked Object-Following System. The system starts with a camera that captures real-world scenes, and the YOLOv8 deep learning model detects objects and defines “notable symbols” for tracking. A distance estimate technique calculates the target’s vicinity using variations in bounding box width, allowing for accurate distance inference.

The video frame is divided into three zones—left, middle, and right—to direct the robot’s movement. The system directs the robot to move left, forward, or right based on the zone in which the target appears. Pathfinding algorithms like as A* and Tabu Search are used for obstacle avoidance and optimal navigation, resulting in efficient and precise target tracking in complicated situations. This framework offers a structured way to integrating deep learning and metaheuristics to create strong object-following apps.

B. Preparation for Model Training

1) Dataset

A large dataset of photos is critical for this research since it serves as the foundation for training the YOLO object detection model, allowing the marked object-following algorithm with collision prevention to function properly. To provide reliable real-time detection and tracking of persons, the dataset should comprise a wide range of human poses, orientations, clothing kinds, and environmental circumstances such as lighting, weather, and busy locations. This diversity allows the model to generalize well to real-world events and consistently distinguish humans from other items. Fig. 2 are the sample dataset that shows individual is wearing the markers that we can detect and monitor.



Fig. 2. Dataset of images.

Additionally, the quality and diversity of the dataset are critical for training the YOLO model because they improve its capacity to recognize persons in difficult surroundings, reduce false positives, and increase detection accuracy. An extensive dataset also prepares the model to face obstacles such as occlusions, overlapping objects, and complicated backdrops, resulting in robust performance. Without a well-curated dataset, the

model's performance will suffer, potentially leading to errors in the system's marked object-following and collision prevention functions.

2) Data Annotation

Rectangular markers known as bounding boxes are used in object detection tasks to show the location and size of items in an image. They are employed in this study to label and annotate the dataset, designating the people that the algorithm must recognize and obey. The YOLOv8 model needs the precise coordinates of the target objects—humans—in each training image in order to learn how to distinguish them from other things in the environment. This is why this step is so important. The model is trained on bounding box-annotated photos, which enables it to anticipate comparable boxes surrounding persons in real-time during deployment, guaranteeing precise tracking and detection. The effectiveness of the marked object-following algorithm with collision prevention depends on the YOLOv8 model's ability to recognize and marked object. This is made possible through the process of building these bounding boxes. An example of an image with bounding boxes is shown in the Fig. 3.



Fig. 3. Examples of bounding box images.

A key challenge in developing an object-following algorithm with collision prevention is the potential for confusion when multiple similar objects are present, as the system might mistakenly track any detected object without a distinguishing feature. In environments where objects lack unique visual characteristics, relying solely on generic detection could result in tracking the wrong target. To address this, the research incorporates distinct logos or markers placed on the intended object, which the YOLOv8 model is specifically trained to recognize as the target class. These markers act as notable symbols, which is shown in Fig. 4, enabling the system to distinguish the designated object from others in the vicinity. By focusing on these specific classes, the algorithm reliably

tracks the intended object, reducing errors and enhancing performance in crowded or dynamic environments. This approach ensures accurate and safe object-following behavior, even in complex settings, by preventing the system from mistakenly tracking unintended objects.



Fig. 4. The four experimental notable symbols.

3) Model Selection

The YOLOv8-nano (YOLOv8n) model is designed to operate at fast speeds on embedded systems and other devices with constrained processing power. Take a look at the Table I below.

TABLE I. PERFORMANCE COMPARISON OF YOLOV8 MODELS [50]

Model	mAP ^{val} 50-95	Speed CPU ONNX (ms)	Params (M)
YOLOv8n	37.3	80.4	3.2
YOLOv8s	44.9	128.4	11.2
YOLOv8m	50.2	234.7	25.9
YOLOv8l	52.9	375.2	43.7
YOLOv8x	53.9	479.1	68.2

With a mean Average Precision (mAP) of 37.3% at the 50-95 Intersection over Union (IoU) threshold, it offers an effective balance between speed and accuracy, achieving an inference time of 80.4 milliseconds when running on a CPU using the ONNX runtime. Its compact architecture, consisting of only 3.2 million parameters, makes it highly suitable for real-time object detection, particularly on low-power processors like the Raspberry Pi. Although larger models, such as YOLOv8-s and YOLOv8-m, provide greater accuracy, their slower inference times (128.4 ms and 234.7 ms, respectively) make them less practical for resource-constrained environments. Therefore, YOLOv8-nano is selected for its efficient performance, ensuring that the marked object-following algorithm can accurately detect and track individuals that are wearing markers in real time while minimizing delay, which is critical for collision prevention and overall system reliability.

C. Proposed Distance Estimation Algorithm

The distance estimation algorithm for this study leverages the width of the detected target class (such as notable symbols) to estimate the relative distance between the object and the camera in a robotic system. The fundamental concept is that the bounding box width, dynamically generated by the YOLOv8 model, provides a reliable reference for gauging distance. As the bounding box width decreases, the target is inferred to be moving further away, while an increase in the width suggests that the object is getting closer to the camera.

The Fig. 5 illustrates relationship between the bounding box width and the distance is inversely proportional. The equation for calculating the distance is as follows.

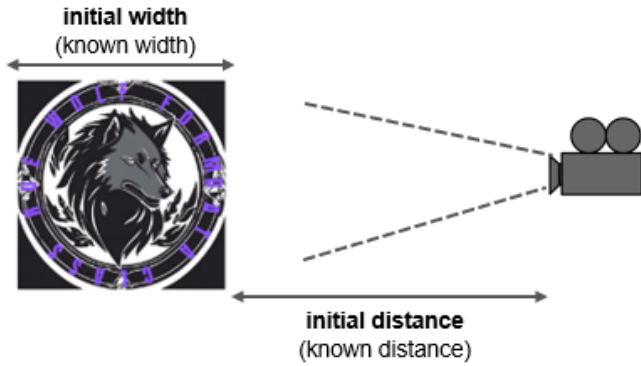


Fig. 5. Relationship between bounding box width and the camera.

$$\text{distance} = \left(\frac{\text{current_width}}{\text{initial_width}} \right) \times \text{initial_distance}$$

Where:

- *current_width* refers to the detected bounding box width of the target class at a particular moment.
- *initial_width* is the reference bounding box width at a known distance.
- *initial_distance* is the known distance from the camera when the object has the initial width.

This formula assumes that the camera and the object are in fixed, calibrated positions, and the size of the object remains constant. YOLOv8 dynamically detects the bounding box width, enabling real-time and accurate distance estimation based on width variations.

D. Proposed Object-Following Algorithm

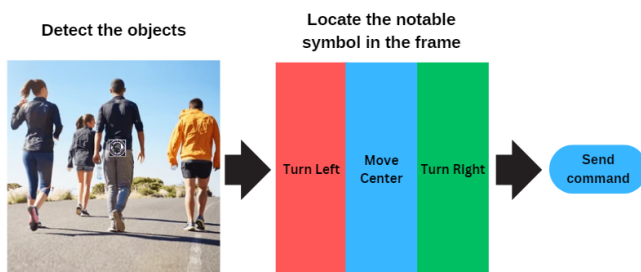


Fig. 6. Object-following algorithm.

Fig. 6 illustrates the proposed marked object-following algorithm. The process begins by detecting objects within the frame, particularly focusing on identifying the notable symbol worn by the designated person or individual. Once the symbol is detected, the algorithm evaluates its position within the frame. Depending on the location of the symbol—whether it is in the left, center, or right of the frame—the system will

send a corresponding command. If the symbol is positioned on the left, the algorithm issues a “Turn Left” command; if it is centered, a “Move Center” command is sent, and if on the right, a “Turn Right” command is executed. This step-by-step analysis ensures precise tracking and directional adjustments, enabling the system to follow the intended target efficiently.

E. Integrated Pathfinding and Distance Estimation System

Algorithm 1 Integrated System Algorithm for Marked Object Following

Inputs: Camera feed, YOLOv8 model, grid dimensions (*max_rows*, *max_cols*), start position *S*, target position *T*, calibration constant *k*, and serial communication interface.

Steps:

1) **System Initialization:**

- Load the YOLOv8 model.
- Set up camera input using *cv2*.
- Define grid dimensions and initialize parameters (*tabu_list*, *distance_threshold*).
- Establish serial communication for robot control.

2) **Distance Estimation and Immediate Movement:**

- Detect objects in each frame using YOLOv8.
- Divide the frame into regions: left, center, and right.
- For each detected object:
 - Identify the class and bounding box width *w_b*.
 - Calculate distance $d = \frac{k}{w_b}$.
 - Send movement commands based on position and distance:
 - Turn Right: If object is on the right.
 - Turn Left: If object is on the left.
 - Move Forward: If object is in the center and $d > 14$ inches.
 - Stop: If $d \leq 14$ inches.

3) **Pathfinding with A-Star and Tabu Search:**

- Generate the global path using A-Star by computing $f(n) = g(n) + h(n)$ for each node.
- Refine the path with Tabu Search:
 - Evaluate neighbors $N(P)$ of the current path *P* and compute the cost $\text{Cost}(P) = \sum f(n)$.
 - Update the *tabu_list* to avoid revisiting sub-optimal paths.
- Adjust the global path based on the marker’s position from the distance estimation module.

4) **Execute Navigation:**

- Follow the refined path while continuously updating the robot’s position using YOLOv8 detections.
- Use real-time corrections to handle dynamic obstacles and deviations.

Output: Efficient navigation to the target position *T* with consistent tracking of the marked object.

In order to effectively locate the best path in complicated surroundings, the pathfinding method integrates the advantages of both A-Star (A*) and Tabu Search. A heuristic-based technique called A* is used to determine the shortest path in a grid or graph between a start point and a target. The pathfinding algorithm is almost similar to the study of Gorro

et al. [51]. It strikes a balance between an estimate of the remaining distance to the target and the cost of the road already taken. Because of these two factors, A* is a popular and effective solution for pathfinding issues. However, because it may revisit less-than-ideal solutions, A* may lose effectiveness in areas with a high density of barriers or recurrent paths.

Tabu Search is used into this process to overcome this restriction. The “tabu list,” a memory component added by Tabu Search, keeps account of recently traveled routes or moves that have been judged to be less-than-ideal. Tabu Search compels the algorithm to investigate alternate options, even if they seem less promising at first, by forbidding the re-exploration of these routes. This investigation promotes the finding of globally optimal pathways and aids in avoiding local minima.

A*, the first step in the combined algorithm, creates an initial path from the start to the destination. After that, Tabu Search takes control and iteratively refines this path. As long as it is not in the tabu list, the best neighbor is chosen as the current path after adjacent paths are assessed according to their costs at each stage. Because the tabu list is dynamically updated, recent errors or less-than-ideal routes are kept in mind and avoided in subsequent cycles. Until a certain number of iterations is reached or no more advancements can be made, the process keeps going.

In complex grids or maps, where the existence of barriers or constraints could cause traditional algorithms to be misguided, this hybrid approach works very well. The method produces a stable and adaptable solution by utilizing the advantages of A* for initial pathfinding and Tabu Search for iterative refining. This makes it appropriate for applications like robotics, navigation, and logistical planning.

F. Evaluation Metrics

These metrics help in determining how well the model is able to identify humans in various scenarios, ensuring that the robotic system can perform its tasks accurately and reliably. The research can efficiently analyze the model’s strengths and weaknesses, direct the tuning of hyperparameters, and make well-informed decisions about model optimization to achieve the desired performance in real-world environments by utilizing specific evaluation metrics like Precision, Recall, Mean Average Precision (mAP), and F1-score.

1) Precision

The ratio of true positive detections to the total of both true positive and false positive detections is known as precision. It assesses how well the model distinguishes, among all the detected things, only the pertinent objects—in this example, humans. High precision reduces false positives by increasing the likelihood that the YOLOv8-nano model’s predictions of humans are accurate. High precision is necessary to prevent the robot from unintentionally following non-human things in the setting of a marked object-following robotic system. This is important for both efficiency and safety.

$$\text{Precision} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})}$$

2) Recall

The ratio of real positive detections to the total of false negatives and true positives is known as recall. It illustrates how well the model was able to identify every pertinent object (people) in the dataset. High recall means that the model is capable of detecting most of the humans present in the environment, minimizing false negatives. In this study, a high recall is important because, in the event that a human is not detected, the robotic system may not follow its intended path, which could be harmful in scenarios where it is used for public space guidance or healthcare assistance.

$$\text{Recall} = \frac{\text{True Positive}}{(\text{True Positive} + \text{False Positive})}$$

3) Mean Average Precision (mAP)

A comprehensive statistic called Mean Average Precision (mAP) provides an overall measure of accuracy by assessing the model’s performance across various Intersections over Union (IoU) thresholds. The model’s ability to balance precision and recall is indicated by a single performance score that is obtained by combining the two criteria. Because it aids in understanding the trade-offs between minimizing false positives (precision) and detecting as many humans as feasible (recall), mAP is particularly significant in this research. A greater mAP is a useful parameter for optimizing human detection in the YOLOv8-nano model since it shows that the model performs well in both areas.

$$\text{mAP} = \frac{1}{k} \sum_{i=1}^k \text{AP}_i$$

4) F1-score

The F1-score is a measure that provides a balance between Precision and Recall, calculated as the harmonic mean of the two. When it comes to striking a balance between false positives and false negatives, it is especially helpful. In this research, the F1-score is essential because it gives a more holistic view of the model’s performance. A high F1-score shows that the model is successful in capturing all important detections and is accurate in identifying humans. This is especially important in applications like robotic assistance and navigation where misidentifying a non-human object (false positive) or missing a human (false negative) can have serious repercussions.

$$\text{F1-score} = 2 \times \frac{(\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}$$

IV. RESULT

A. YOLOv8 Performance

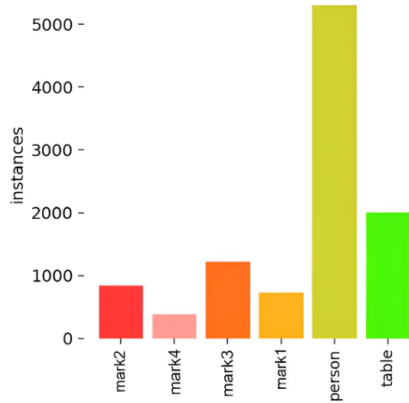


Fig. 7. Class distribution.

Monitoring the class distribution (Fig. 7) reveals significant class imbalance, with the “person” class dominating the dataset. This imbalance is a critical issue in object detection tasks [52], [53], as it can lead to over-optimization for frequent classes and under-performance for rarer ones, such as “mark4”. Techniques like data augmentation, oversampling, or loss re-weighting [54], [56] could address this and enhance performance across all classes.

In particular, the model may become too optimized for detecting the more frequent class (“person”) while struggling to reliably recognize less frequent ones (“mark4”). Regular analysis of this distribution allows researchers to take corrective action, such as boosting underrepresented classes or employing advanced approaches to reduce class imbalance. By addressing these issues, the model’s overall accuracy and generalization capabilities across all classes can be significantly improved, enhancing its robustness in real-world applications.

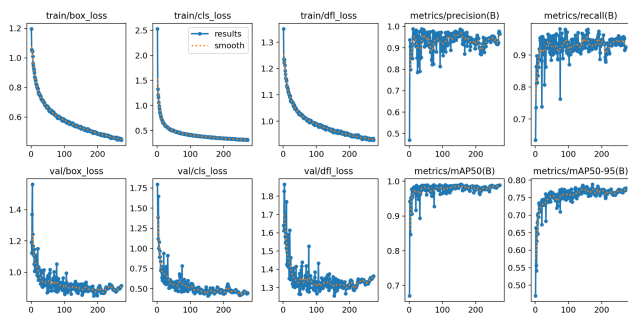


Fig. 8. Result graph of YOLOv8-nano model.

Fig. 8 illustrates the training and validation graphs for key metrics and loss functions in the proposed marked object-following algorithm with collision prevention, based on YOLOv8 object detection. The training loss curves,

which include bounding box regression (train/box_loss), classification loss (train/cls_loss), and distributional focal loss (train/df_l_loss), show a consistent decline as training progresses, indicating that the model is effectively minimizing prediction errors. This steady reduction in training losses suggests that the model is becoming more accurate in identifying and classifying objects while refining the predicted bounding box coordinates.

The training and validation metrics (Fig. 8) show consistent declines in losses, indicating effective learning and generalization. Compared to YOLOv4-tiny [55], the YOLOv8-based model achieves higher mAP values (0.961 at IoU@0.5), demonstrating competitive detection performance. However, slight precision-recall drops for the “person” class align with findings in [53], suggesting the need for improved handling of dominant classes in imbalanced datasets. Incorporating techniques like focal loss or semi-supervised learning [56], [57] could mitigate this challenge.

The validation losses (val/box_loss, val/cls_loss, val/df_l_loss) exhibit a similar downward trend, though with natural fluctuations, indicating the model’s generalization to unseen data. Precision and recall metrics remain high and stable, which demonstrates the model’s ability to maintain a balance between correctly identifying true positives and minimizing false positives. The mAP@0.5 and mAP@0.5-0.95 values show continuous improvement, signaling enhanced detection performance across various Intersection over Union (IoU) thresholds.

These graphs provide insight into the model’s training dynamics, showcasing a well-balanced process where the algorithm is consistently improving in both training and validation phases. The steady convergence of losses and strong performance metrics suggest the model is learning effectively without overfitting, ensuring reliable detection and tracking in real-time marked object-following scenarios.

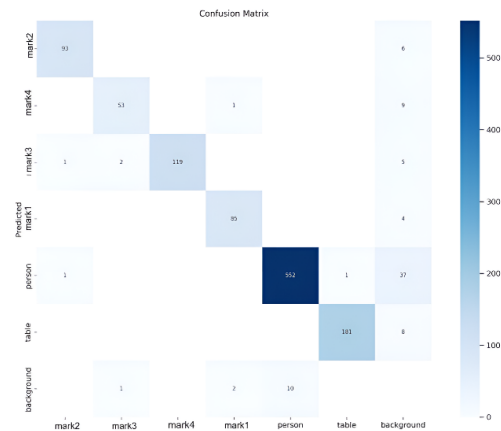


Fig. 9. Confusion matrix for YOLOv8-nano model.

The Confusion Matrix serves as an effective tool for assessing the performance of the YOLOv8-nano model by illustrating its accuracy in predicting various classes. This table presents the frequency of actual versus predicted classes, allowing for an evaluation of the alignment between the

model's predictions and the true labels. Analyzing the confusion matrix in the context of YOLOv8 helps identify specific cases where the model successfully classifies an object or incorrectly identifies it as another class. This insight is crucial for recognizing the model's strengths and weaknesses, enabling targeted improvements to enhance accuracy. By examining the matrix, researchers can identify which classes are frequently confused and make necessary modifications to training data, model architecture, or hyperparameters to rectify these issues.

Fig. 9 is a detailed review of the confusion matrix for the YOLOv8-nano model reveals that the "mark2" class is accurately predicted 93 times, with misclassifications occurring once each as "mark3" and "person." The "mark5" class achieves 53 correct predictions, with minor misclassifications as "mark3" twice and as "background" once. The "mark3" class exhibits perfect performance, yielding 119 correct predictions without any misclassifications. The "mark1" class is accurately predicted 85 times but is mistaken for "mark5" once and "background" twice. The "person" class demonstrates high accuracy with 552 correct predictions, though it is confused with the "background" class on 10 occasions. Finally, the "table" class is correctly identified 181 times, with one misclassification as "person". The model's overall performance is not greatly affected by these minor misclassifications. The YOLOv8-nano model has good prediction ability and little confusion between various object classes in spite of these small inaccuracies.

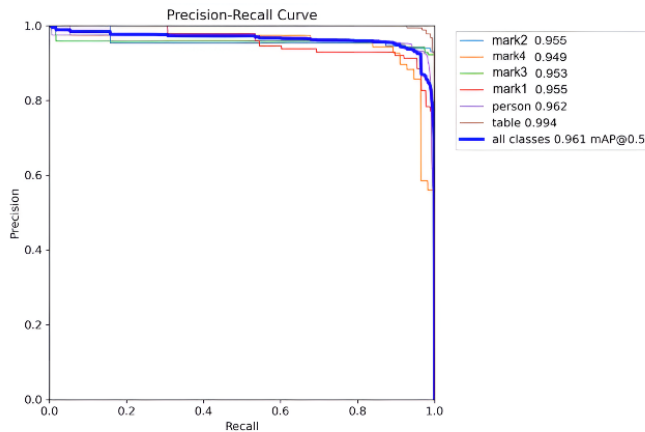


Fig. 10. Precision-recall curve.

Fig. 10 displays the Precision-Recall (PR) Curve for the various classes within the dataset. This curve visually illustrates the balance between precision and recall for each class, highlighting the model's effectiveness in correctly identifying true positives while reducing false positives. The results indicate that most classes attain very high precision and recall values, with both metrics nearing 1.0, which demonstrates the model's strong capability in detecting these objects with minimal errors. However, the "person" class exhibits slightly lower precision and recall values than the other classes, suggesting potential challenges in accurately detecting and distinguishing humans within the dataset. The overall mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 0.5 across all classes stands at 0.961, signifying an excellent

balance of high precision and recall. This elevated mAP value indicates that the model is well-optimized for precise object detection, ensuring reliable performance in identifying and classifying the various objects analyzed in this study.

The Precision-Recall Curve (Fig. 10) illustrates the model's strong detection capabilities. However, lower precision for the "person" class highlights challenges in distinguishing humans in cluttered environments. Fine-tuning anchor box sizes or using hybrid feature extractors, as shown in [55], could enhance performance.

Overall, the results validate the proposed marked object-following algorithm, achieving reliable detection and collision prevention in real-time scenarios. The high F1-scores across classes (Fig. 11) and stable precision-recall metrics ensure robust tracking. These findings demonstrate the algorithm's potential for deployment in assistive robotics and autonomous systems. Future work could integrate multi-sensor fusion or explore adaptive learning strategies [54], [56] to further improve robustness.

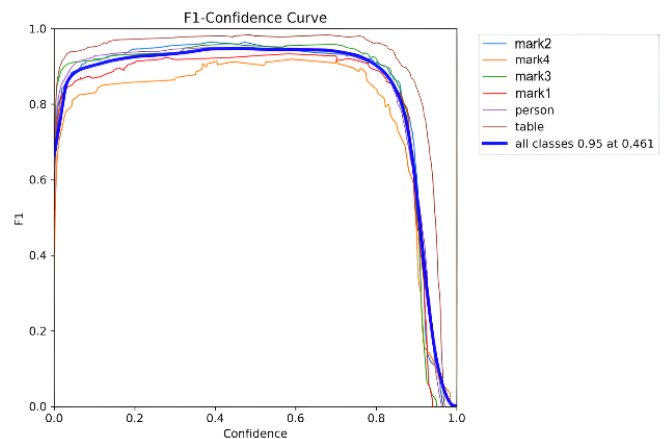


Fig. 11. F1-confidence curve.

As shown in the Fig. 11, the F1-scores for the different classes maintain high values at moderate confidence levels, with an overall peak of 0.95 for all classes at a confidence threshold of 0.461. This indicates a strong balance between precision (correctly identifying the object) and recall (detecting most of the relevant objects) for each class. The curves for each class follow a similar trend, with a sharp drop-off beyond the optimal confidence threshold, suggesting that the model is highly accurate up to a certain point, after which false positives start to increase.

This curve illustrates how well the model can detect different types of objects, meaning that the marked object-following algorithm can track the designated person (represented by the "person" class) and distinguish it from the other target classes, which include the table and various markers. Accurate object identification is ensured by maintaining a high F1-score across these classes, which helps to prevent collisions and ensures reliable human following.

B. Implementation of Experimental Algorithms

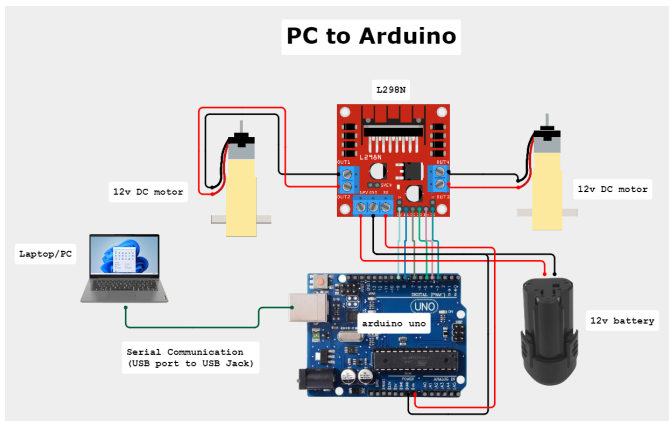


Fig. 12. The circuit diagram for the basic robot.

Fig. 12 illustrates the experimental prototype configuration used to test the marked object-following system. The circuit diagram details the connections between a laptop or PC, an Arduino Uno board, an L298N motor driver module, two 12V DC motors, and a 12V battery. The laptop/PC establishes a USB connection with the Arduino Uno, facilitating serial communication for data exchange and control commands. The Arduino Uno is linked to the L298N motor driver module, which regulates the two 12V DC motors by adjusting their speed and direction based on the signals received. The motors receive power from the 12V battery, which is directly connected to the L298N module, supplying the required voltage for operation. This configuration enables precise motor control through the Arduino, allowing commands from the PC to direct the motors via the L298N driver, effectively simulating navigation and following behaviors in response to target detection and distance estimation algorithms.

During the Test 1 the robot's ability to follow a sample image held by a human in a simple environment. The robot efficiently detects the target image and maintains a consistent following distance, showcasing its tracking accuracy and responsiveness. The straightforward setup allows the robot to smoothly follow the human, effectively illustrating its basic operational capability and fundamental functionality in a controlled, uncomplicated scenario.

Finally, during Test 2 the robot's capability to follow a human in a complex environment, effectively navigating without causing disruptions. When the human passes near the right green line, the robot seamlessly turns right, demonstrating a prompt and accurate response without any delay or difficulty in executing the turn. Similarly, when approaching the left green line, the robot exhibits the same level of efficiency, turning left without encountering any issues. This demonstrates the robot's robust decision-making and adaptability, ensuring reliable marked object-following behavior even in challenging environments.

V. CONCLUSION

The primary aim of this study was to develop algorithms capable of accurately detecting and following a designated

marked object while estimating the distance between the user and the system in real-time, utilizing the YOLOv8 model for object detection. An obstacle avoidance was created using a distance estimation algorithm with the pathfinding A* and Tabu search algorithm. The model's performance was evaluated through key metrics, including the F1-score and Precision-Recall. The F1-Confidence curve indicated a robust F1-score of 0.95 for all classes at a confidence threshold of 0.461, reflecting a well-balanced performance between precision and recall, effectively minimizing false positives and false negatives in detecting the target classes. Additionally, the Precision-Recall curve showcased the effectiveness of the YOLOv8 model, achieving an overall mean Average Precision (mAP) of 0.961 at an Intersection over Union (IoU) threshold of 0.5 for all classes. This high mAP value demonstrates the model's reliability in accurately identifying and tracking the target classes while maintaining consistent detection performance.

The successful integration of a YOLO-based detection model with a distance estimation, path finding algorithms (A*) and Tabu Search highlights the system's potential for real-world applications. Although the system faces limitations in handling visual disturbances and detecting objects from side angles, it has produced promising results under controlled conditions. The achieved F1-score and Precision-Recall values underscore the model's effectiveness, providing a solid foundation for further enhancements and potential applications in various environments. The distance estimation algorithm and the path finding A* and Tabu search are crucial for detecting potential collisions and obstacle avoidance with marked objects, and the inclusion of an obstacle detection feature could further mitigate collision risks.

ACKNOWLEDGMENT

We extend our profound gratitude to Cebu Technological University and Cebu Normal University for their unwavering support throughout the course of this research endeavor.

REFERENCES

- [1] A. Sangha and M. Rizvi, "Detection of acne by deep learning object detection", 2021. <https://doi.org/10.1101/2021.12.05.21267310>
- [2] Y. Fu, "Recent deep learning approaches for object detection", *Highlights in Science Engineering and Technology*, vol. 31, p. 64-70, 2023. <https://doi.org/10.54097/hset.v31i.4814>
- [3] K. Sharada, "Deep learning techniques for image recognition and object detection", *E3s Web of Conferences*, vol. 399, p. 04032, 2023. <https://doi.org/10.1051/e3sconf/202339904032>
- [4] J. García-González, I. García-Aguilar, D. Medina, R. Luque-Baena, E. López-Rubio, & E. Domínguez, "Vehicle overtaking hazard detection over onboard cameras using deep convolutional networks", p. 330-339, 2022. https://doi.org/10.1007/978-3-031-18050-7_32
- [5] S. Primakov, A. Ibrahim, J. Timmeren, G. Wu, S. Keek, M. Beuque et al., "Automated detection and segmentation of non-small cell lung cancer computed tomography images", *Nature Communications*, vol. 13, no. 1, 2022. <https://doi.org/10.1038/s41467-022-30841-3>
- [6] J. Redmon, S. Divvala, R. Girshick, & A. Farhadi, "You only look once: unified, real-time object detection", p. 779-788, 2016. <https://doi.org/10.1109/cvpr.2016.91>
- [7] J. Schmidhuber, "Deep learning in neural networks: an overview", *Neural Networks*, vol. 61, p. 85-117, 2015. <https://doi.org/10.1016/j.neunet.2014.09.003>
- [8] K. Wang, S. Dang, F. He, & C. Peng, "A path planning method for indoor robots based on partial a global a-star algorithm", 2017. <https://doi.org/10.2991/fmsmt-17.2017.83>

- [9] O. Vural, K. Çelik, Y. Yurdagül, & M. Sağlam, "A new automation system for equipment status and efficiency detection with machine learning based image processing", *Orclever Proceedings of Research and Development*, vol. 1, no. 1, p. 38-44, 2022. <https://doi.org/10.56038/oprd.v1i1.206>
- [10] N. Chinthamu, "Iot-based secure data transmission prediction using deep learning model in cloud computing", *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 11, no. 4s, p. 68-76, 2023. <https://doi.org/10.17762/ijritcc.v11i4s.6308>
- [11] P. Sun, G. Chen, & Y. Shang, "Adaptive saliency biased loss for object detection in aerial images", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 10, p. 7154-7165, 2020. <https://doi.org/10.1109/tgrs.2020.2980023>
- [12] Z. Naik and M. Gandhi, "A review: object detection using deep learning", *International Journal of Computer Applications*, vol. 180, no. 29, p. 46-48, 2018. <https://doi.org/10.5120/ijca2018916708>
- [13] M. M. Islam, A. Lam, H. Fukuda, Y. Kobayashi, and Y. Kuno, "A person-following shopping support robot based on human pose skeleton data and lidar sensor," in **Intelligent Computing Methodologies: 15th International Conference, ICIC 2019, Nanchang, China, August 3-6, 2019, Proceedings, Part III 15**, Springer International Publishing, 2019, pp. 9-19.
- [14] Z. Gao, Z. Wang, L. Saint-Bauzel, and F. Ben Amar, "2D lidar-based large workspace frontal human following for a mobile robot," *Available at SSRN 4538601*.
- [15] D. Cha and W. Chung, "Human-leg detection in 3D feature space for a person-following mobile robot using 2D LiDARs," **International Journal of Precision Engineering and Manufacturing**, vol. 21, no. 7, pp. 1299-1307, 2020.
- [16] D. Z. Wang, I. Posner, and P. Newman, "Model-free detection and tracking of dynamic objects with 2D lidar," **The International Journal of Robotics Research**, vol. 34, no. 7, pp. 1039-1063, 2015.
- [17] J. Shackleton, B. VanVoorst, and J. Hesch, "Tracking people with a 360-degree lidar," in **2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance**, 2010, pp. 420-426.
- [18] T. Nowak, K. Ćwian, and P. Skrzypczyński, "Real-time detection of non-stationary objects using intensity data in automotive LiDAR SLAM," **Sensors**, vol. 21, no. 20, p. 6781, 2021.
- [19] W. Chung, H. Kim, Y. Yoo, C. B. Moon, and J. Park, "The detection and following of human legs through inductive approaches for a mobile robot with a single laser range finder," **IEEE Transactions on Industrial Electronics**, vol. 59, no. 8, pp. 3156-3166, 2011.
- [20] E. J. Jung, J. H. Lee, B. J. Yi, J. Park, and S. T. Noh, "Development of a laser-range-finder-based human tracking and control algorithm for a marathoner service robot," **IEEE/ASME Transactions on Mechatronics**, vol. 19, no. 6, pp. 1963-1976, 2013.
- [21] R. Mark2bri and M. T. Choi, "Deep-learning-based indoor human following of mobile robot using color feature," **Sensors**, vol. 20, no. 9, p. 2699, 2020.
- [22] S. O. Adebola, **A Human Following Robot for Fall Detection**, Master's thesis, Middle Tennessee State University, 2019.
- [23] M. Padhen, K. Shimpi, R. Thakur, and P. V. Sontakke, "Human detecting robot based on computer vision-machine learning," **International Journal for Research in Applied Science and Engineering Technology**, vol. 8, no. IX, 2020.
- [24] S. P. Yong and Y. C. Yeong, "Human object detection in forest with deep learning based on drone's vision," in **2018 4th International Conference on Computer and Information Sciences (ICCOINS)**, 2018, pp. 1-5.
- [25] A. U. Bokade and V. R. Ratnaparkhe, "Video surveillance robot control using smartphone and Raspberry Pi," in **2016 International Conference on Communication and Signal Processing (ICCSP)**, 2016, pp. 2094-2097.
- [26] J. Zhang, G. Song, G. Qiao, T. Meng, and H. Sun, "An indoor security system with a jumping robot as the surveillance terminal," **IEEE Transactions on Consumer Electronics**, vol. 57, no. 4, pp. 1774-1781, 2011.
- [27] A. Imteaj, M. I. J. Chowdhury, M. Farshid, and A. R. Shahid, "RoboFI: Autonomous path follower robot for human body detection and geolocalization for search and rescue missions using computer vision and IoT," in **2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)**, 2019, pp. 1-6.
- [28] J. Jommuangbut and K. Sritrakulchai, "Development of the human following robot control system using HD webcam," in **2018 International Electrical Engineering Congress (iEECON)**, 2018, pp. 1-4.
- [29] G. R. Poornima, J. L. Avinash, S. Palle, S. S. Kumar, K. S. Kumar, and P. R. Prasad, "Image processing based human pursuing robot," in **2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT)**, 2020, pp. 408-412.
- [30] M. Sharikmaslat, R. Sidhaye, and A. Narkar, "Image processing based human pursuing robot," in **2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA)**, 2019, pp. 702-704.
- [31] M. S. Sefat, D. K. Sarker, and M. Shahjahan, "Design and implementation of a vision based intelligent object follower robot," in **2014 9th International Forum on Strategic Technology (IFOST)**, 2014, pp. 425-428.
- [32] M. N. A. Bakar and A. R. M. Saad, "A monocular vision-based specific person detection system for mobile robot applications," **Procedia Engineering**, vol. 41, pp. 22-31, 2012.
- [33] T. Inoue, Y. Okazaki, and K. Itoya, "Person following algorithm with pixel-area addition method of thermal sensors for autonomous mobile robots," in **2024 10th International Conference on Control, Automation and Robotics (ICCAR)**, 2024, pp. 77-82.
- [34] G. Feng, X. Guo, and G. Wang, "Infrared motion sensing system for human-following robots," **Sensors and Actuators A: Physical**, vol. 185, pp. 1-7, 2012.
- [35] I. T. Ćirić, Ž. M. Ćojbašić, D. D. Ristić-Durrant, V. D. Nikolić, M. V. Ćirić, M. B. Simonović, and I. R. Pavlović, "Thermal vision based intelligent system for human detection and tracking in mobile robot control system," **Thermal Science**, vol. 20, suppl. 5, pp. 1553-1559, 2016.
- [36] C. Filippini, D. Perpetuini, D. Cardone, A. M. Chiarelli, and A. Merla, "Thermal infrared imaging-based affective computing and its application to facilitate human-robot interaction: A review," **Applied Sciences**, vol. 10, no. 8, p. 2924, 2020.
- [37] O. Montiel-Ross, R. Sepúlveda, O. Castillo, and P. Melin, "Ant colony test center for planning autonomous mobile robot navigation," **Computer Applications in Engineering Education**, vol. 21, no. 2, pp. 214-229, 2013.
- [38] Q. H. Nguyen, H. Vu, T. H. Tran, and Q. H. Nguyen, "Developing a way-finding system on mobile robot assisting visually impaired people in an indoor environment," **Multimedia Tools and Applications**, vol. 76, pp. 2645-2669, 2017.
- [39] F. J. Perez-Grau, F. Caballero, A. Viguria, and A. Ollero, "Multi-sensor three-dimensional Monte Carlo localization for long-term aerial robot navigation," **International Journal of Advanced Robotic Systems**, vol. 14, no. 5, p. 1729881417732757, 2017.
- [40] M. R. Nowicki, D. Belter, A. Kostusiak, P. Cížek, J. Faigl, and P. Skrzypczyński, "An experimental study on feature-based SLAM for multi-legged robots with RGB-D sensors," **Industrial Robot: An International Journal**, vol. 44, no. 4, pp. 428-441, 2017.
- [41] A. Al Arabi, P. Sarkar, F. Ahmed, W. R. Rafie, M. Hannan, and M. A. Amin, "2D mapping and vertex finding method for path planning in autonomous obstacle avoidance robotic system," in **2017 2nd International Conference on Control and Robotics Engineering (ICCRE)**, 2017, pp. 39-42.
- [42] Z. Wang, J. Kinugawa, H. Wang, and K. Kazuhiro, "The simulation of nonlinear model predictive control for a human-following mobile robot," in **2015 IEEE International Conference on Robotics and Biomimetics (ROBIO)**, 2015, pp. 415-422.
- [43] W. Mi, X. Wang, P. Ren, and C. Hou, "A system for an anticipative front human following robot," in **Proceedings of the International Conference on Artificial Intelligence and Robotics and the International Conference on Automation, Control and Robotics Engineering**, 2016, pp. 1-6.
- [44] A. Pandey, S. Kumar, K. K. Pandey, and D. R. Parhi, "Mobile robot navigation in unknown static environments using ANFIS controller," **Perspectives in Science**, vol. 8, pp. 421-423, 2016.
- [45] M. Almasri, K. Elleithy, and A. Alajlan, "Sensor fusion based model for collision free mobile robot navigation," **Sensors**, vol. 16, no. 1, p. 24, 2015.

- [46] C. Gomez, A. C. Hernandez, J. Crespo, and R. Barber, "A topological navigation system for indoor environments based on perception events," *International Journal of Advanced Robotic Systems*, vol. 14, no. 1, p. 1729881416678134, 2016.
- [47] M. Tee Kit Tsun, B. T. Lau, and H. Siswoyo Jo, "An improved indoor robot human-following navigation model using depth camera, active IR marker, and proximity sensors fusion," *Robotics*, vol. 7, no. 1, p. 4, 2018.
- [48] P. Janousek, Z. Slanina, and W. Walendziuk, "Target-following robotic platform based on UWB localization and depth camera," *IFAC-PapersOnLine*, vol. 58, no. 9, pp. 247–252, 2024.
- [49] M. Q. Do and C. H. Lin, "Embedded human-following mobile-robot with an RGB-D camera," in *2015 14th IAPR International Conference on Machine Vision Applications (MVA)*, 2015, pp. 555–558.
- [50] G. Jocher, A. Chaurasia, and J. Qiu, Ultralytics YOLOv8, version 8.0.0, 2023. [Online]. Available: <https://github.com/ultralytics/ultralytics>.
- [51] K. Gorro, L. Roble, M. A. Magana, and R. P. Buot, "Prototype of an Indoor Pathfinding Application with Obstacle Detection for the Visually Impaired," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 15, no. 9, 2024. doi: <https://dx.doi.org/10.14569/IJACSA.2024.01509106>.
- [52] N. Crasto, "Class Imbalance in Object Detection: An Experimental Diagnosis and Study of Mitigation Strategies," *arXiv preprint arXiv:2403.07113*, 2024. Available: <https://arxiv.org/abs/2403.07113>.
- [53] Y. Li, B. Wang, Z. Kang, S. Tang, L. Wu, and J. Li, "Overcoming Classifier Imbalance for Long-tail Object Detection with Balanced Group Softmax," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 10991–11000. Available: https://openaccess.thecvf.com/content_CVPR_2020/papers/Li_Overcoming_Classifier_Imbalance_for_LongTail_Object_Detection_With_Balanced_Group_CVPR_2020_paper.pdf.
- [54] M. Tomaszewski and J. Osuchowski, "Effectiveness of Data Resampling in Mitigating Class Imbalance for Object Detection," *CEUR Workshop Proceedings*, vol. 3628, 2023. Available: <https://ceur-ws.org/Vol-3628/paper14.pdf>.
- [55] Roboflow, "YOLOv8 vs. YOLOv4 Tiny: Compared and Contrasted," Roboflow, 2023. Available: <https://roboflow.com/compare/yolov8-vs-yolov4-tiny>.
- [56] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," in *Proc. IEEE Int. Conf. on Computer Vision (ICCV)*, 2017, pp. 2980–2988.
- [57] Ultralytics, "Pretrain YOLOv8 with Semi-supervised Learning," GitHub Issue #4373, 2023. Available: <https://github.com/ultralytics/ultralytics/issues/4373>.

Hawk-Eye Deblurring and Pose Recognition in Tennis Matches Based on Improved GAN and HRNet Algorithms

Weixin Zhao

Department of Physical Education and Sports Science, Fuzhou University, Fuzhou, 350108, China

Abstract—In tennis matches, the Hawk-eye system causes blurry trajectory judgment and low accuracy in player posture recognition due to rapid movement and complex backgrounds. Therefore, the research improves the backbone network and iterative attention feature fusion mechanism of deblur generative adversarial network version. At the same time, Ghost, Sandglass module, and coordinate attention mechanism are used to optimize the high-resolution network, and a new model for deblurring and pose recognition of Hawk-eye images in tennis matches is proposed by integrating the improved generative adversarial network and high-resolution network. The new model achieved an information entropy value of 11.2, a peak signal-to-noise ratio of 29.74 decibels, a structural similarity of 0.89, a minimum parameter size of 4.53, and a running time of 0.25 seconds on the tennis tracking dataset and the Max Planck Society human posture dataset, which was superior to current advanced models. The highest accuracy of deblurring and pose recognition for the model under different lighting intensities was 92.44%, and the highest improvement rate of video frame quality was 18%. From this, the model has significant advantages in deblurring effect, posture recognition accuracy, parameter quantity, and running time, and has high practical application potential. It can provide an advanced theoretical reference for tennis match refereeing and technical training.

Keywords—DeblurGANv2; HRNet; tennis; hawk-eye system; deblurring; pose recognition

I. INTRODUCTION

In tennis matches, the Hawk-eye system, as a high-precision technology application, has been widely used in referee decision-making, motion analysis, and other fields worldwide. The Hawk-eye system captures image data within the field through multiple high-speed cameras and uses image processing algorithms for real-time calculation and analysis, providing judgments on whether the ball is within or outside the boundary. In addition, the movements of players in tennis matches are complex and have high spatiotemporal variations. Accurately capturing the players' motion posture is the key to improving the accuracy of Hawk-eye system judgment [1-2]. Pose recognition technology can accurately capture the movement information of athletes by analyzing their body position, movement trajectory, etc., providing a more precise player behavior model for the Hawk-eye system and improving the reliability of judgment results. F. Meng et al. introduced a hybrid neural network to optimize target feature extraction and constructed a novel Hawk-eye detection model to improve the visual detection level of tennis in the sports industry. The

model achieved a tennis motion tracking accuracy of 0.694 under grayscale feature conditions, which was the highest among all testing methods [3]. Y. Zhao et al. proposed a lightweight tennis Hawk-eye detection scheme combining "You Only Look Once version 5" (YOLOv5) to address the inefficiency of traditional tennis detection algorithms. Compared with traditional methods, experimental results showed that this algorithm reduced model parameters by 42% and computational complexity by 44%, while improving detection accuracy by 2% [4]. Y. Yang et al. built a new tennis trajectory prediction method by combining artificial neural network detection algorithm and stereo vision. The experiment showed that this method had high reliability and robustness, effectively improving the prediction ability of tennis trajectory [5]. D Gao et al. built a deep learning driven small object automatic detection method to address the difficulty of small object detection in tennis videos. The experiment showed that this method performed well in the integrity, recognition accuracy, and detection speed [6]. Y. Ke et al. proposed an object detection algorithm on the basis of deep learning aimed at handling advanced visual tasks such as tennis. This algorithm combines prior knowledge of tennis impact areas. The experimental results showed that it could provide high detection accuracy and faster detection speed, effectively improving the accuracy and stability of tennis impact detection [7].

Generative Adversarial Networks (GANs) are powerful deep learning models that have shown great potential in tasks such as image generation, denoising, and deblurring [8]. Bian J et al. argued that detecting dense movements from fast-moving objects in sports videos remained challenging. To this end, a novel table tennis detection model by combining GAN and P2ANet was proposed. The model could achieve an average accuracy of 88.47% for the localization and recognition of 8 types of table tennis movements, while improving the detection robustness [9]. Ghezelsefloo H R et al. proposed an auxiliary calibration model for Hawk-eye detection after improving the GAN algorithm to effectively reduce the error in Hawk-eye detection in sports events. The model had a success rate of 92.17% in assisting correct judgments in 130 sports, and had significant practical value [10]. Peng X et al. constructed a video pose detection model for ball players by combining sensor image acquisition with GAN and Modbus. The performance of the model on Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) was about 4.5 and 0.143 higher than other algorithms, respectively [11]. In

addition, High-Resolution Network (HRNet) is an excellent deep learning model that excels in processing detailed information in high-resolution images and is widely used in fields such as image segmentation, object detection, and pose recognition. Nguyen H C et al. proposed an automatic combined human pose estimation model by combining HRNet and YOLOv5 to improve the accuracy of human motion pose estimation. The processing time on a 3.3-megapixel dataset was 55FPS, and the highest accuracy of human keypoint detection was 98.24% [12]. Li Y combined HRNet to construct a monocular video motion capture method, which optimized it for human motion reconstruction problems such as floating, ground penetration, and sliding. This method achieved a good balance between accuracy and frame rate, and had significant detection advantages [13]. Fitzpatrick A et al. In order to strengthen the accuracy of the hawk-eye monitoring system under different serve and return strategies, the researchers proposed a hawk-eye-assisted detection model with multimodal data training and convolutional graph neural network processing. The experimental results show that the method achieves higher detection accuracy and greater stability for a variety of different serving and hitting motions [14]. Ning T et al. In order to address the limitations of computer vision-assisted table tennis ball detection, the researchers proposed a real-time computational method for determining the landing point of a table tennis ball. The experimental results showed that the method achieved a detection speed of 45.3 fps, and the key frame extraction method correctly recognized the landing point frames with an accuracy rate of more than 93.3% [15].

In summary, traditional Hawk-Eye systems mostly rely on classical deblurring algorithms, but these methods usually cannot effectively deal with fast motion and multi-angle shooting conditions, resulting in image distortion and inaccurate pose estimation. In addition, while existing pose recognition methods are able to achieve better results in static or slow scenes, they still exhibit large errors in dynamic tennis match scenarios, especially when players are moving fast. Although several approaches have been dedicated to solving this problem, existing solutions usually face certain limitations. For example, traditional algorithms are computationally inefficient when dealing with large-scale data and insufficiently adaptable when facing complex environments. In order to overcome these limitations, the study innovatively proposes a novel hawk-eye deblurring and pose recognition model for tennis matches, which incorporates the improved deblur generative adversarial network version 2 (DeblurGANv2) and HRNet algorithms, respectively, and introduces a lightweight Mobilenetv2 backbone network, Ghost module and Sandglass module are introduced to improve the computational efficiency, and Iterative Attention Feature Fusion (IAFF) and Coordinate Attention (CA) mechanisms are adopted to enhance the feature extraction capability. Enhance

the feature extraction ability, and at the same time significantly improve the processing speed and robustness of the algorithm, especially in the complex environment of the adaptability of the excellent performance. Among them, the improved deblurring technique of GAN can better handle blurred images under different motion states while ensuring image quality. Combined with the high-resolution feature of HRNet, the accuracy of pose recognition is further improved, especially in the capture of complex motion and action details. The research aims to significantly improve image clarity and pose recognition accuracy in dynamic scenes by combining these innovative designs, providing an effective solution for efficient and real-time tennis match Hawk-eye systems. This research is divided into four parts, the first part is the analysis and summary of others' research, the second part describes how the Hawk-Eye image deblurring algorithm for tennis matches and the tennis match stance recognition model were designed, respectively, while the third part tests the performance of the model, and the last part is the summary of the article.

II. METHODS AND MATERIALS

In response to the challenges of image blur and athlete pose recognition in tennis matches, this study first introduces IAFF based on DeblurGANv2 and uses Feature Pyramid Network (FPN) to achieve bidirectional fusion of multi-scale features. Secondly, based on HRNet, Ghost, Sandglass, CA mechanism, and Transformer-based object tracking module are sequentially introduced to propose a new Hawk-eye analysis model that integrates deblurring and pose recognition.

A. Deblurring Algorithm for Hawk-eye Images in Tennis Matches Based on Improved GAN

Image blur is one of the main issues affecting the accuracy of Hawk-eye system judgment, especially during the rapid movement of players and the high-speed flight of the ball [16-17]. The traditional image degradation model mainly generates degraded images from the original image after degradation function and noise processing, while the restoration model restores clear images close to the original image by applying restoration functions to the degraded image. Image degradation is usually caused by factors such as motion blur and poor lighting [18-20]. Through this approach, a classic algorithm for image motion blur, DeblurGANv2, is introduced into the study. This algorithm efficiently removes motion blur through the improved GAN. Compared with traditional deblurring algorithms, DeblurGANv2 has adaptability to complex backgrounds and multi-scale feature extraction ability, which can more comprehensively restore image details and is suitable for dynamic motion scenes with large changes [21-23]. In order to adapt to tennis motion detection and improve universality, the structure of DeblurGANv2 is improved, and an improved DeblurGANv2 tennis match Hawk-eye image deblurring algorithm is proposed. The framework of this algorithm is shown in Fig. 1.

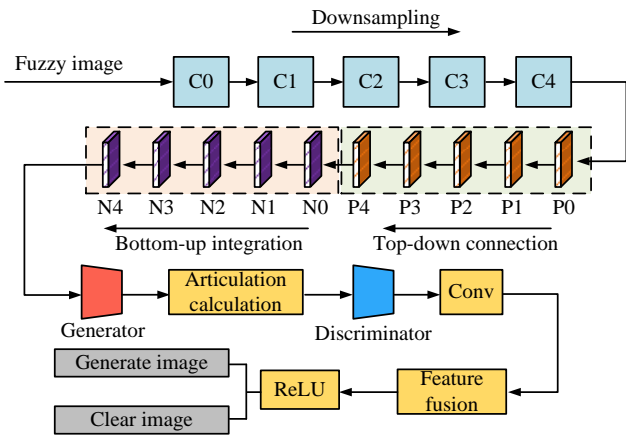


Fig. 1. Improved framework of the Hawk-eye image deblurring algorithm for tennis match in DeblurGANv2.

In Fig. 1, the algorithm framework mainly has generator, discriminator, and attention mechanism modules, and includes specific operations such as convolution, upsampling, downsampling, feature fusion, stacked convolution, batch normalization, and ReLU activation. Firstly, the improved backbone network is used to downsample and extract five feature maps of different scales step by step, from C0 to C4. Then, these feature maps are generated using the top-down connection of the FPN, namely P0 to P4. Next, P0 to P4 gradually perform bottom-up feature fusion to obtain feature maps, namely N0 to N4. Afterwards, N0 to N4 are fused with the original image to generate the final deblurred image. The generated image is then input into the discriminator along with the clear image to calculate the clarity probability of the generated image, in order to optimize the generator. Finally, the generator and discriminator are alternately trained and output after convolution, feature fusion layer, and ReLU activation. Compared with the improved DeblurGANv2, a bottom-up feature fusion branch is added, allowing low-level features to fully interact with high-level features. Specifically, the feature fusion path is designed through the bidirectional connection of FPN, which first performs top-down connection and then performs bottom-up fusion. The calculation formula is shown in Eq. (1).

$$P_i = Conv_{1 \times 1}(C_i) + Upsample(P_{i+1}) \quad (1)$$

In equation (1), P_i represents the intermediate feature map fused from top to bottom. C_i represents feature maps of different scales extracted from the backbone network. $Conv_{1 \times 1}$ signifies a 1×1 convolution operation. $Upsample$ represents upsampling operation. Next, P_i is subjected to bottom-up feature fusion to enhance the information transmission of cross layer features, as shown in Eq. (2).

$$N_i = Conv_{3 \times 3}(P_i) + Downsample(N_{i-1}) \quad (2)$$

In Eq. (2), N_i represents the fused feature map. $Downsample$ represents downsampling operation. To further enhance the ability to focus on key regions, the improved

algorithm adopts the IAFF mechanism. IAFF calculates attention weights through multiple iterations to focus on important regions in the image [24-26]. The calculation process is shown in Eq. (3).

$$\begin{cases} F_{IAFF}^t = F^t + \Gamma \cdot Attention(Q^t, K^t, V^t) \\ Attention(Q^t, K^t, V^t) = Soft \max(\frac{Q^t K^{tT}}{\sqrt{d}}) V^t \end{cases} \quad (3)$$

In Eq. (3), F_{IAFF}^t represents the fused feature map after the t -th iteration. F^t signifies the input feature of the t -th iteration. Γ represents the fusion coefficient. Q^t , K^t and V^t signify the query, key, and value matrices for the t -th iteration, respectively. \sqrt{d} represents the scaling factor, used to avoid excessive attention weights. In addition, to accelerate image deblurring processing, the improved algorithm uses Mobilenetv2 as the backbone network, replacing the traditional heavy convolutional network. Mobilenetv2 uses depthwise separable convolution, which contains two parts: depthwise convolution and pointwise convolution. The calculation formula is shown in Eq. (4) [27-29].

$$DSCConv(x) = DepthwiseConv(x) + PointwiseConv(x) \quad (4)$$

In Eq. (4), $DSCConv(x)$ represents a depthwise separable convolution operation. $DepthwiseConv(x)$ represents convolution only in the spatial dimension. $PointwiseConv(x)$ represents using 1×1 convolution to fuse features in the channel dimension. This convolution method significantly reduces the computational and parameter complexity, as shown in Eq. (5).

$$FLOPs_{DSCConv} = \frac{1}{k^2} \times FLOPs_{StandardConv} \quad (5)$$

In Eq. (5), k signifies the size of the convolution kernel. In summary, the model generator and discriminator network structure of the improved DeblurGANv2 are shown in Fig. 2.

Fig. 2 (a) displays the improved generator structure of DeblurGANv2, and Fig. 2(b) displays the improved discriminator structure of DeblurGANv2. In Fig. 2 (a), the generator includes multiple layers of feature extraction modules. Firstly, the main network performs downsampling to extract feature maps of different scales layer by layer. Then, FPN is used for multi-scale feature fusion, adopting a bidirectional connection design of top-down and bottom-up. In the feature map processing at each scale, convolutional layers, Batch Normalization (BN) layers, and ReLU activation functions are used to enhance feature extraction performance, and IAFF mechanism is adopted to highlight key regions. Finally, the fused feature map is upsampled and residual connected to reconstruct a blurred image. As shown in Fig. 2 (b), the discriminator structure includes a series of convolutional layers, Leaky ReLU activation functions, and BN layers, which are used to extract high-level features of the input image. The discriminator adopts a layer by layer downsampling design, gradually compressing the image size

through multiple convolutional layers. Finally, the fully connected layer is applied to calculate the probability score between the generated image and the real clear image, to determine whether it is a real image.

To further improve the restoration effect and image detail preservation ability of DeblurGANv2 in deblurring tasks. A mixed loss function is designed, including adversarial loss,

perceptual loss, and image reconstruction loss. Firstly, the adversarial loss is used to optimize the game between the generator and discriminator, making the output image of the generator more realistic. Secondly, the perceptual loss is applied to measure the differences in high-level semantic features between generated images and real clear images, as shown in Fig. 3.

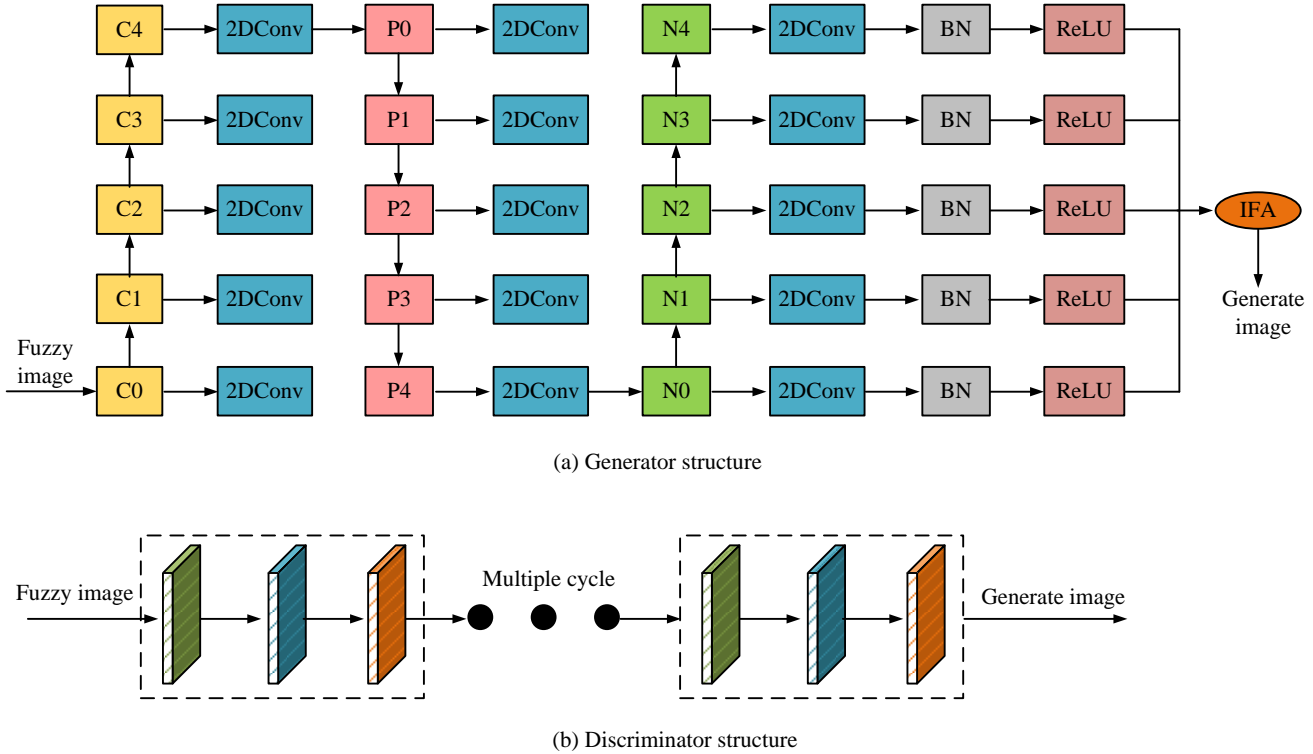


Fig. 2. Improved generator and discriminator structure for DeblurGANv2.

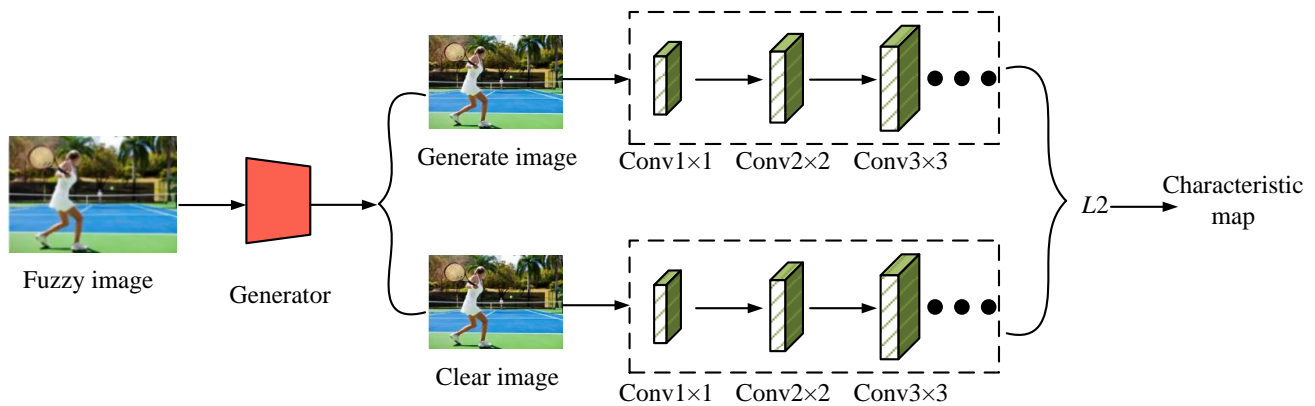


Fig. 3. Schematic diagram of perceptual loss.

As shown in Fig. 3, the calculation of perceptual loss is achieved by introducing a pre-trained deep Convolutional Neural Network (CNN) to extract high-level features, and comparing the feature differences between the generated image and the real clear image at different convolutional layers. Specifically, the input blurred image is first deblurred by a generator to generate a restored image. Then, the generated

image and the real clear image are input into a deep CNN, and feature maps are extracted through several layers of convolution. In these feature maps, the perceptual loss calculates the difference in L_2 -norm between the generated image and the real image on each layer of the feature map, as shown in Eq. (6).

$$L_{perc} = \sum_l \lambda_l \|\phi_l(G(x)) - \phi_l(y)\|_2^2 \quad (6)$$

In Eq. (6), L_{perc} represents perceptual loss, which is applied to measure the difference in high-level features between the generated image and the real image. λ_l signifies the weight coefficient. ϕ_l represents the feature mapping of the l -th layer of the deep CNN. $G(x)$ represents the deblurred image output by the generator. y represents the real and clear image. $\|\cdot\|_2$ represents the $L2$ -norm. In order to capture more levels of semantic information, perceptual loss usually selects feature maps from multiple convolutional layers for calculation, and the comprehensive formula is shown in Eq. (7).

$$L_{perc} = \sum_{l=1}^L \left(\frac{1}{H_l W_l C_l} \sum_{h=1}^{H_l} \sum_{w=1}^{W_l} \sum_{c=1}^{C_l} (\phi_l(G(x))_{h,w,c} - \phi_l(y)_{h,w,c})^2 \right) \quad (7)$$

In Eq. (7), L represents the number of selected convolutional layers. H_l , W_l and C_l respectively represent the height, width, and number of channels of the l -layer feature map. $\phi_l(G(x))_{h,w,c}$ represents the feature values at position (h, w) and channel c in the l -layer feature map. $\phi_l(y)_{h,w,c}$ signifies the feature value of the corresponding position of the real image in the l -th layer feature map.

B. Construction of Tennis Match Pose Recognition Model Integrating Improved HRNet Algorithm

After improving the DeblurGANv2 structure, the blurring effect of Hawk-eye monitoring images is effectively avoided. This research further focuses on the task of athlete posture recognition in tennis matches. Pose recognition is an important part of technical analysis in tennis matches, which is crucial for the standardization analysis of player movements, optimization of game strategies, and monitoring of potential rule violations

[30-31]. However, the rapid changes in player movements, complex postures, and dynamic background interference in tennis matches make traditional posture recognition methods difficult to cope with [32-33]. To address these issues, the HRNet algorithm is combined with research. Compared with other advanced methods, it consistently maintains high-resolution feature maps throughout the entire feature extraction process and fully utilizes multi-scale information through layer by layer fusion of multi-resolution features. In addition, targeted optimization and improvement are carried out on the basis of the standard HRNet architecture to make it more suitable for the scene requirements of tennis matches. The improved HRNet is displayed in Fig. 4.

In Fig. 4, the improved HRNet has four stages, each stage achieving the extraction and fusion of multi-scale features through parallel resolution branches. The first stage uses standard convolution operations for preliminary feature extraction, generating high-resolution feature maps. In the second stage, while retaining high-resolution branches, low resolution branches are applied to capture deeper feature information through downsampling. In the third stage, more resolution branches are added to achieve multi-scale feature alignment and complementarity from high resolution to low resolution. In the fourth stage, a cross resolution feature fusion strategy is used to effectively combine feature information from different resolutions, generating a multi-scale feature map with global context awareness capability. Specifically, there are four major improvements. First, Ghost and Sandglass have been introduced to replace the Bottleneck and Basicblock modules in HRNet, reducing the running parameter. Second, the introduced CA enhances the feature extraction capability of the model. Third, the ability to enhance data has been improved through unbiased data augmentation methods. Fourth, the effectiveness of object detection is improved through a separate object tracking module. The target tracking module is shown in Fig. 5.

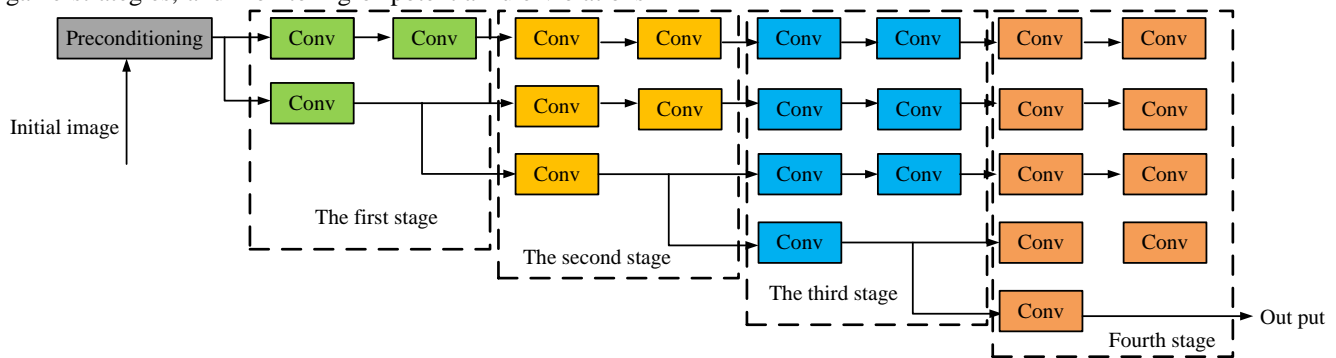


Fig. 4. Improved HRNet structure.

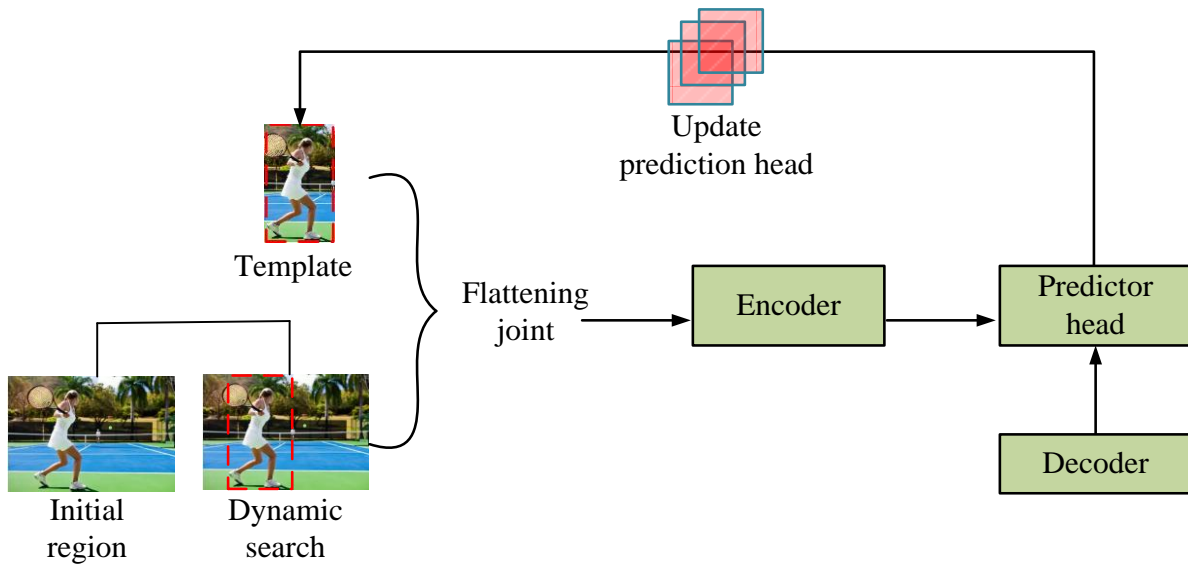


Fig. 5. Target tracking module diagram.

As shown in Fig. 5, the target tracking module mainly includes three core parts: target detection, target association, and trajectory update, which are also key steps based on the Transformer target tracking algorithm. Firstly, the input video frames are processed by an object detection network to generate an initial detection box for the target, and key information such as the target category and confidence level is annotated. Subsequently, the target association module combines the appearance features of the target, such as color, texture, and motion features, such as speed and trajectory, to match the target in the current frame with the tracked target in the previous frame, ensuring the continuity and consistency of the trajectory. Among them, the target association module achieves matching by calculating the similarity matrix between targets, where the similarity comprehensively considers the appearance and motion features of the targets, as shown in Eq. (8).

$$S_{ij} = \alpha \cdot IoU(B_i, B_j) + \beta \cdot \cos(f_i, f_j) \quad (8)$$

In Eq. (8), S_{ij} represents the similarity score between target i and target j . α and β both represent weight parameters. $IoU(B_i, B_j)$ represents the intersection over union ratio of the bounding boxes of target i and target j . f_i and f_j represent the appearance feature vectors of target i and target j , respectively. $\cos(f_i, f_j)$ represents the cosine similarity between appearance feature vectors. After the association is completed, the trajectory update module uses Kalman filtering to dynamically estimate the position and velocity of the target, in order to smooth the tracking results, as displayed in Eq. (9).

$$\begin{cases} x_t = Fx_{t-1} + E(r_t - Rx_{t-1}) \\ E = P_{t-1}H^T(RP_{t-1}R^T + \zeta)^{-1} \end{cases} \quad (9)$$

In Eq. (9), x_t and x_{t-1} represent the target state variables at the current time and the previous time, respectively. F represents the state transition matrix. r_t represents the observation vector at the current time. R represents the observation model matrix. E represents the Kalman gain. P_{t-1} represents the covariance matrix of the previous time state. ζ represents the covariance matrix of observed noise. In addition, improving the CA mechanism in the HRNet encodes the global directional information of the input feature map, and then generates a weight distribution through embedding coordinate information. Finally, the feature map is adjusted using weighting, as displayed in Eq. (10).

$$\begin{cases} z_c^h = \frac{1}{H} \sum_{i=1}^H X(i, j, c), z_c^w = \frac{1}{W} \sum_{j=1}^W X(i, j, c) \\ f_c = \sigma(\text{Conv}_{1 \times 1}^h(z_c^h) + \text{Conv}_{1 \times 1}^w(z_c^w)) \\ Y(i, j, c) = f_c \cdot X(i, j, c) \end{cases} \quad (10)$$

In Eq. (10), z_c^h and z_c^w represent the global information encoding of feature map X in the height and width directions, respectively. $X(i, j, c)$ represents the feature values of the input feature map at position (i, j) and channel c . f_c represents the generated channel weight. σ represents the activation function. $Y(i, j, c)$ represents the output feature map. $\text{Conv}_{1 \times 1}^h$ and $\text{Conv}_{1 \times 1}^w$ represent 1×1 convolution operations in the height and width directions, respectively. Regarding the original Bottleneck and Basicblock modules in HRNet, Ghost and CA modules are respectively integrated for optimization. The schematic diagram of the optimized Bottleneck and Basicblock modules is shown in Fig. 6.

Fig. 6 (a) displays the Bottleneck module structure before and after optimization. Fig. 6 (a) shows the Basicblock module

structure before and after optimization. As shown in Fig. 6 (a), the two 1×1 convolutions and 3×3 convolutions in the original module have been replaced by the Ghost module, which efficiently reduces feature redundancy by generating primary and auxiliary features. In addition, CA modules are inserted between Ghost modules. By modeling the interaction between space and channels, the model's ability to express features of the target area has been enhanced. In Fig. 6 (b), the structure originally composed of two stacked 3×3 convolutions has been

replaced with a lightweight convolution combination implemented by the Sandglass module. The Sandglass module significantly reduces the number of parameters and computational complexity while retaining feature information. At this point, the calculation formula for Ghost in Bottleneck is shown in Eq. (11).

$$Y = \text{Concat}(X * W_m, \sigma(X * W_m) * W_a) \quad (11)$$

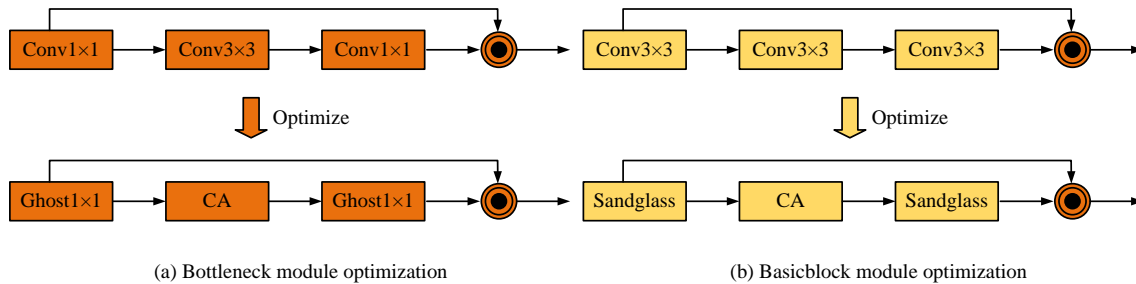


Fig. 6. Schematic diagram of Bottleneck and Basicblock modules before and after optimization.

In Eq. (11), W_m and W_a represent the convolution kernel parameters of the main feature and auxiliary feature, respectively. The Sandglass in Basicblock is shown in Eq. (12).

$$Y = X + \sigma(\text{DepthwiseConv}(\text{Point wiseConv}(X))) \quad (12)$$

In Eq. (12), $\text{Point wiseConv}(_)$ represents a 1×1 point convolution. $\text{DepthwiseConv}(_)$ represents deep convolution. Based on the improvement of HRNet structure and the comprehensive improvement of DeblurGANv2, a new tennis match Hawk-eye deblurring and pose recognition model is proposed. The process is shown in Fig. 7.

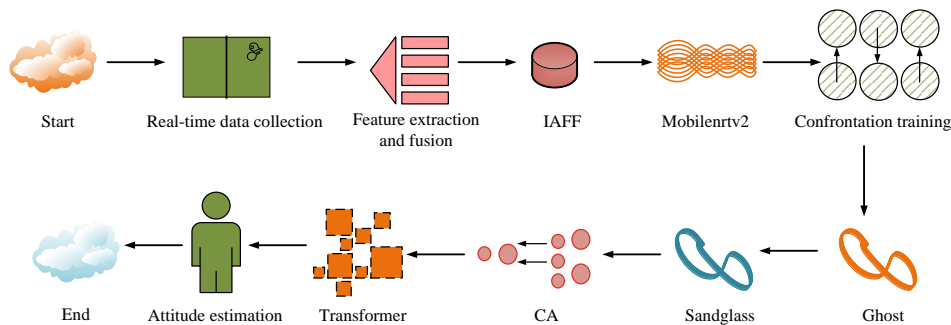


Fig. 7. New model flow of Hawk-eye deblurring and pose recognition in tennis match.

As shown in Fig. 7, firstly, the improved DeblurGANv2 is used for multi-scale feature extraction and fusion of blurred images. The IAFF mechanism is introduced to focus on key regions, and the Mobilenetv2 backbone network is taken to reduce computational overhead. The generator is optimized through adversarial training to generate high-quality and clear images. Subsequently, based on the improved HRNet for pose recognition, Ghost and Sandglass modules are introduced to replace the original Bottleneck and Basicblock modules to reduce the number of parameters, while combining CA mechanism to enhance the feature expression of key regions. Finally, through object detection and Transformer-based object tracking modules, target association and trajectory updates are achieved, outputting clear images, pose keypoints, and motion trajectories.

III. RESULTS

The study first establishes an experimental environment and conducts hyperparameter tuning, with deblurring effect and pose recognition accuracy as the core indicators for testing. The experiment covers two classic datasets and conducts ablation testing, comparative testing, and multi-scenario experiments on lighting, number of people, etc., to verify the robustness and adaptability of the model. Compared with multiple advanced models, the proposed model has achieved good results, especially showing significant advantages in parameter quantity and inference time. In complex lighting and multi-target scenes, the proposed model also demonstrates excellent performance and practical application potential.

A. Performance Testing of Hawk-eye Deblurring and Pose Recognition Model in New Tennis Matches

The study selects two classic public datasets as data sources, namely the Tennis Tracking Dataset (TTD) and the Max Planck Institute for Informatics Human Pose Dataset (MPII). Among them, TTD is a dataset focused on tennis match analysis, which includes key point annotations of players such as head, shoulder, elbow, knee and other joint points, tennis trajectories, as well as action annotations on the court such as serving, returning, running, etc. The MPII dataset is a high-quality dataset focused on human pose estimation, containing 25000 images covering over 40000 human instances. The images in this dataset are from real-life scenarios and provide rich annotations for 16 joint points, including head, shoulders, elbows, knees, etc. The detailed experimental environment parameters are displayed in Table I.

TABLE I. EXPERIMENTAL PARAMETER TABLE

Experimental equipment	Value
CPU	AMD Ryzen 9 5950X
GPU	NVIDIA RTX 4090
Memory	64GB DDR5
Graphics Memory	24GB GDDR6X
Development Environment	Ubuntu 20.04, Python 3.9
Programming Tools	PyTorch 1.10
Initialize learning rate	0.0005
Learning rate batch size	64
Momentum parameters	0.95
Training period	200 epochs

The study first conducts value selection tests on the feature fusion coefficients in the deblurring stage and the convolutional kernel layers in the pose recognition stage, to achieve the optimal state and facilitate subsequent testing. Taking information entropy as an indicator, Fig. 8 displays the test results.

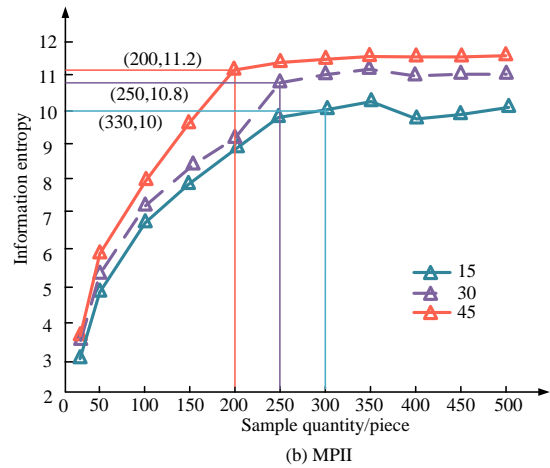
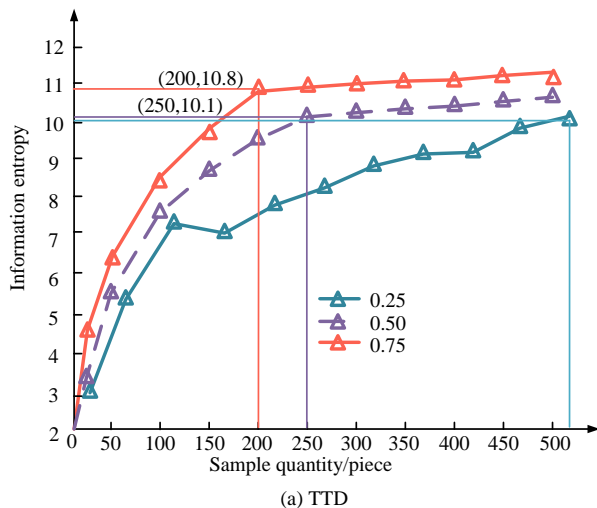
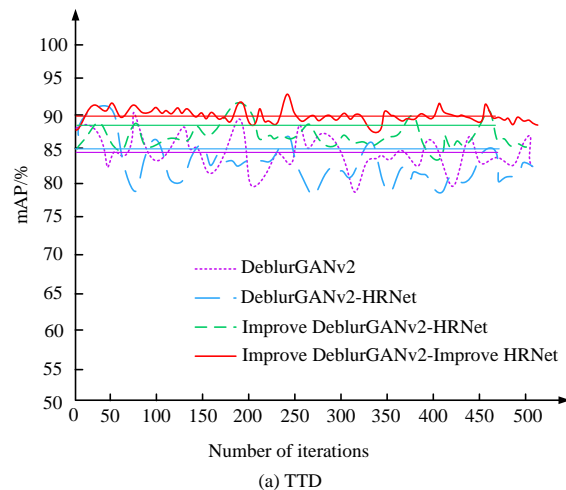


Fig. 8. Hyperparameter selection test result.

Fig. 8 (a) shows the selection test of different feature fusion coefficients in the TTD dataset, and Fig. 8 (b) shows the selection test of different convolutional kernel layers in the MPII dataset. According to Fig. 8 (a), when the fusion coefficient was 0.75, the information entropy grew the fastest and tended to stabilize at a sample size of 200, reaching a maximum value of 10.8. When the fusion coefficients were 0.25 and 0.50, the information entropy tended to stabilize at 250 samples, with values of 10.1 and 9.5, respectively. Overall, a fusion coefficient of 0.75 can significantly improve the deblurring effect. In Fig. 8 (b), with the increase of sample size, the information entropy gradually increased. When the number of convolutional kernel layers was 45, the information entropy reached its maximum value of 11.2 at a sample size of 200 and tended to stabilize. When the number of convolutional kernel layers was 30, the information entropy tended to stabilize at 250 samples, reaching 10.8. When the number of convolutional kernel layers was 15, it increased to 330 samples to reach a stable state, with an information entropy value of 10.0. In summary, when the fusion coefficient was 0.75 and the number of convolutional kernels was 45, the deblurring effect of the model was optimal. The study conducts ablation testing on the final model using the Mean Average Precision (mAP) of keypoint detection as the indicator, as presented in Fig. 9.



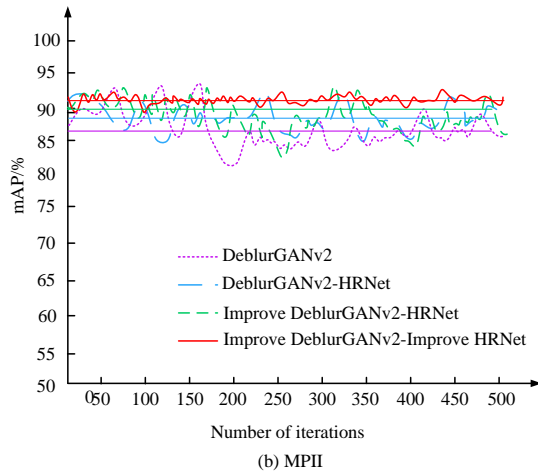


Fig. 9. Ablation test results.

Fig. 9 (a) displays the ablation test results on the TTD dataset, and Fig. 9 (b) displays the ablation test results on the MPII dataset. As shown in Fig. 9 (a), the mAP value of the basic model DeblurGANv2 fluctuated significantly during the iteration process, stabilizing at around 80.18%. After adding HRNet, the model performance improved and mAP remained stable at around 85.37%. After further improving DeblurGANv2 and introducing improved HRNet, the mAP value increased to around 88.74%, demonstrating better stability. The improved DeblurGANv2 and improved HRNet models were ultimately integrated, with mAP values reaching the highest level. It stabilized at around 92.48%, with minimal fluctuations throughout the entire iteration process, demonstrating the best deblurring and pose recognition performance. According to Fig. 9 (b), the mAP value of the basic model DeblurGANv2 fluctuated greatly and remained stable at around 78.77%. After joining HRNet, mAP increased to 83.21%. The improved DeblurGANv2 and HRNet models showed improvements in both stability and accuracy. The final integrated improved model performed the best, with mAP values stable above 90.49% and minimal fluctuations, demonstrating the strongest robustness and consistency. Other advanced deblurring and pose detection models are introduced

for comparison. For example, Scale-Recurrent Network (SRN), Multi-Stage Progressive Restoration Network (MPRNet), Deep Blind Generative Adversarial Network (DBGAN), High-Resolution Transformer (HRFormer), Pose Estimation Network (PoseNet), and Dynamic Encoder for Keypoint Regression (DEKR) are used for comparison. The test results are shown in Table II, using PSNR, SSIM, parameter count, and runtime as indicators.

TABLE II. INDEX TEST RESULTS OF DIFFERENT MODELS

Model	PSNR (dB)	SSIM	Parameter quantity (M)	Running time (s)
SRN	30.05	0.91	6.82	4.35
MPRNet	29.56	0.88	20.63	1.17
DBGAN	28.87	0.87	15.58	0.95
HRFormer	29.85	0.96	25.41	0.66
PoseNet	27.92	0.85	12.74	0.75
DEKR	28.51	0.86	18.88	0.58
Our model	29.74	0.89	4.53	0.25

According to Table II, the model exhibited good comprehensive performance. In terms of PSNR index, the proposed model achieved 29.74dB, which was close to SRN and HRFormer and better than most comparative models. SSIM was 0.89, slightly lower than HRFormer's 0.96, but still stable. The most significant advantage lies in the parameter count and running time. The parameter count of the proposed model was only 4.53M, significantly lower than MPRNet's 20.63M and HRFormer's 25.41M. The inference time was 0.25, which was 78%-94% faster than SRN and MPRNet. This indicates that the model has high efficiency while balancing effectiveness, making it very suitable for real-time image processing tasks.

B. Simulation Testing of Hawk-eye Deblurring and Pose Recognition Model for New Tennis Matches

To verify the practical application effect of the new model, two sets of photos are randomly selected from two types of datasets for testing the deblurring and pose estimation effects of different models, as presented in Fig. 10.

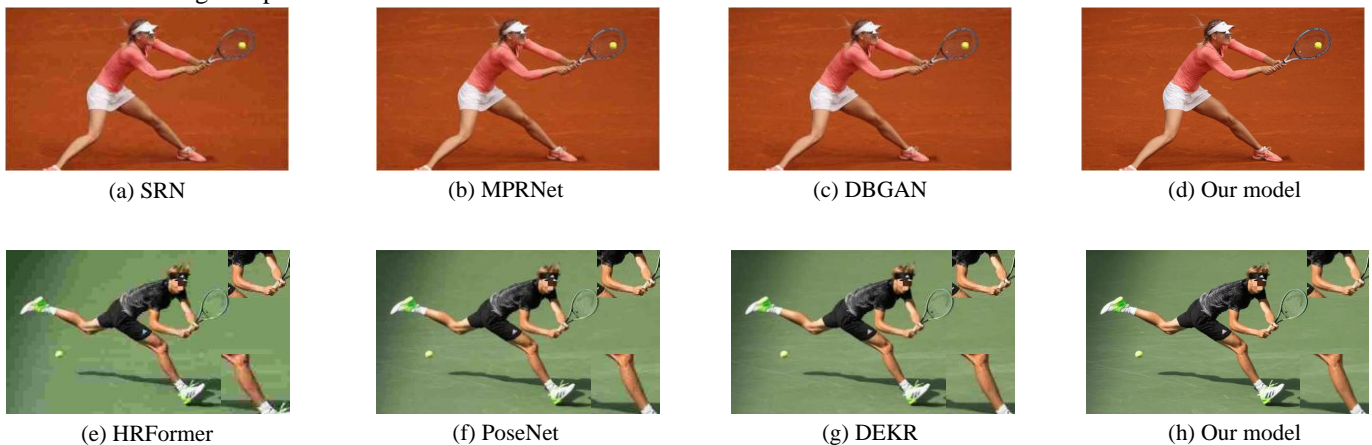


Fig. 10. Comparison of deblurring and pose recognition effect of different models.

Fig. 10 (a)-(d) show the actual comparison results of deblurring applications between SRN, MPRNet, DBGAN, and the proposed model. Fig. 10 (e)-(f) show the comparison results of pose recognition applications between HRFormer, PoseNet, DEKR, and the proposed model. From Fig. 10 (a), SRN and MPRNet had similar deblurring effects, but SRN's restoration details were slightly insufficient, while MPRNet had slight artifacts in the texture part. The DBGAN model performed poorly in handling high dynamic blur, with obvious edge blurring. In contrast, the model performed the best, with clear image details, better overall restoration performance than other models, and more natural texture parts. From Fig. 10 (b), the HRFormer and DEKR models could effectively detect the key points of tennis players. However, the HRFormer model was susceptible to interference in complex backgrounds, and the PoseNet model may miss detection, especially in inaccurate recognition during rapid limb movement. The model can accurately identify all key points and has strong robustness to complex poses and motion blur, demonstrating higher recognition accuracy. The performance of the model under different lighting conditions is tested using video frame quality improvement rate and fuzzy image recognition rate as indicators. The results are shown in Fig. 11.

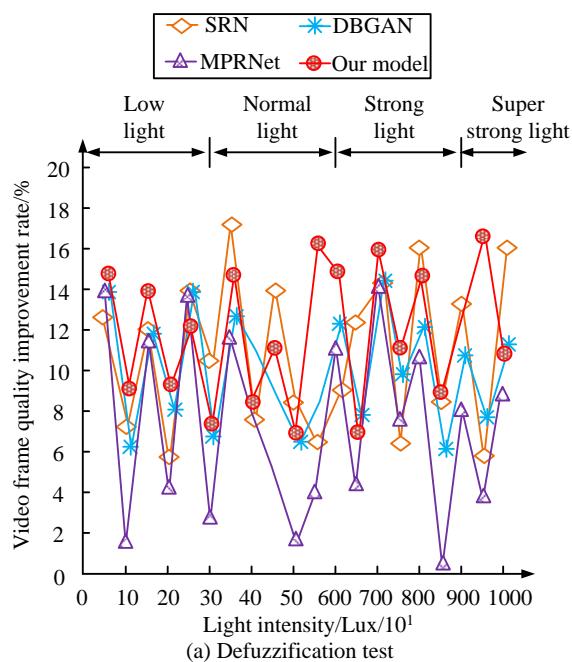


Fig. 11 (a) shows the video frame quality improvement rate test results of four models, and Fig. 11 (b) shows the fuzzy image recognition rate test results of the four models. According to Fig. 11 (a), SRN and MPRNet exhibited relatively stable performance under weak and normal light conditions, with improvement rates ranging from 10% to 12%. However, in strong and ultra strong light environments, the effectiveness of DBGAN and SRN significantly decreased. In contrast, the model showed good improvement rates under various lighting conditions, especially in strong and ultra strong light environments, with a video frame quality improvement rate of 16%-18%, indicating that it could still effectively deblur under severe lighting changes. According to Fig. 11 (b), the recognition rates of HRFormer and PoseNet were relatively high under weak light and normal light, stable between 80.73%-85.46%, respectively. However, the recognition rate of DEKR fluctuated greatly in strong and ultra strong light environments. In contrast, the model maintained a high recognition rate under all lighting conditions, especially in ultra strong light environments, with a recognition rate of up to 92.44%, which was significantly better than other models. Overall, the model demonstrates stronger robustness and stability in deblurring and pose recognition tasks, and has superior adaptability to changes in lighting conditions. The research tests the accuracy of model deblurring and pose recognition in multi-player scenarios, and the results are shown in Fig. 12.

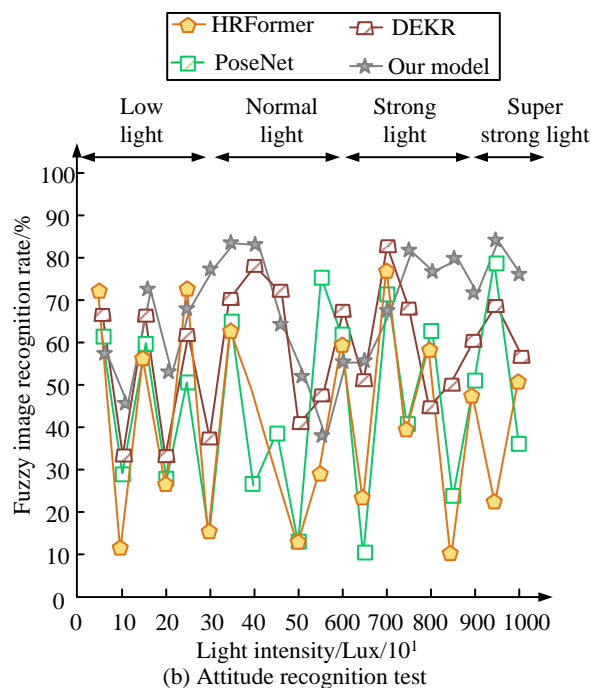


Fig. 11. Test results of video frame quality improvement rate and fuzzy image recognition rate in different modes.

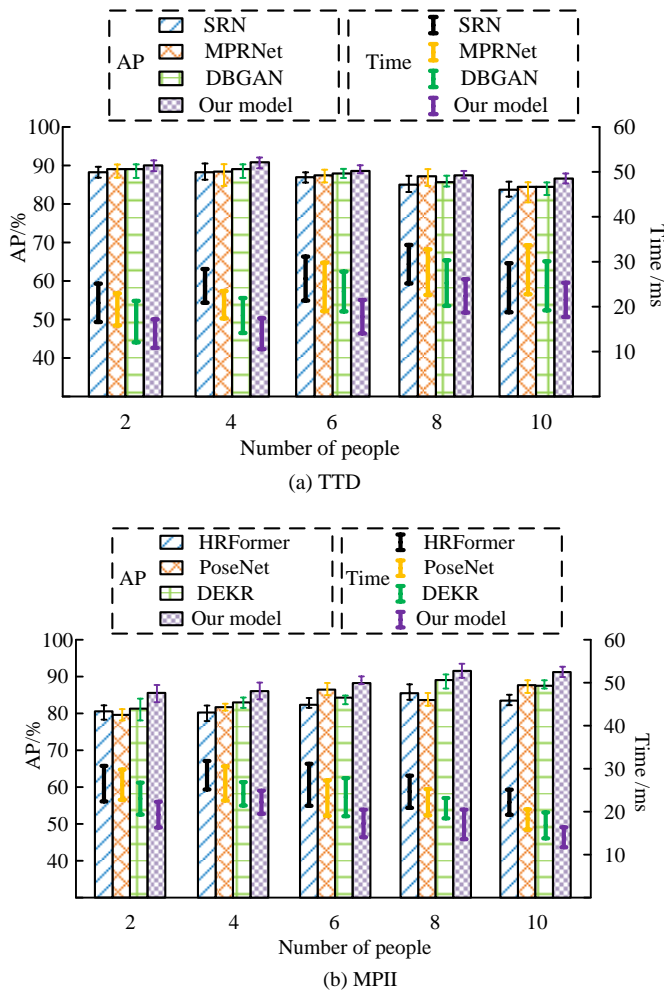


Fig. 12. Results of model deblurring and pose recognition accuracy under different numbers of people.

Fig. 12 (a) displays the deblurring accuracy and time for different models on the TTD dataset, and Fig. 12 (b) displays the pose recognition accuracy and time for different models on the MPII dataset. From Fig. 12 (a), SRN and MPRNet had higher AP values in the 2-person scenario, reaching 90.08% and 88.84% respectively. However, as the number of people increased, the AP values gradually decreased, especially in the 10 person scenario, dropping below 80.96%. The performance of DBGAN in multi-player scenarios was relatively unstable, with large fluctuations in accuracy and longer inference time. In contrast, the model proposed maintained a high AP value of 85.17%-92.38% for all participants, and had the shortest inference time, stabilizing at around 20ms, demonstrating good real-time performance and accuracy. As shown in Fig. 12 (b), HRFormer and PoseNet exhibited high AP values of over 90.02% in both 2-person and 4-person scenarios, but their accuracy significantly decreased in the 10-person scenario. The recognition accuracy and time performance of DEKR in multi-target scenes were unstable and exhibit significant fluctuations. In contrast, the posture recognition accuracy of the model remained stable under different numbers of people, with AP values consistently above 88.74% and inference time controlled at around 30 ms, significantly better than other models. Four types of pose recognition models are tested using

tracking error, target overlap rate, and decision accuracy as indicators, as displayed in Table III.

TABLE III. POSE RECOGNITION TEST RESULTS OF DIFFERENT MODELS

Data set	Model	Tracking error/%	Overlap rate/%	Decision accuracy/%
TTD	HRFormer	6.83	78.57	88.92
	PoseNet	8.93	75.23	85.37
	DEKR	7.62	77.19	87.13
	Our model	5.27	82.35	92.53
MPII	HRFormer	7.12	79.83	89.21
	PoseNet	9.27	74.67	84.71
	DEKR	8.03	76.32	86.83
	Our model	5.31	81.93	93.07

According to Table III, the tracking error of the proposed model on the TTD dataset was 5.27%, significantly lower than HRFormer's 6.83% and PoseNet's 8.93%. The overlap rate was 82.35%, which was about 5%-7% higher than other models. The decision accuracy was 92.53%, significantly better than DEKR's 87.13%. On the MPII dataset, the tracking error of the proposed model was 5.31%, which performed the best. The overlap rate was 81.93%, slightly higher than DEKR's 76.32%. The decision-making accuracy was the highest, at 93.07%, which was significantly improved compared with HRFormer. The above data shows that the model exhibits superior performance and robustness in all indicators.

IV. CONCLUSION

In tennis matches, image blur and pose recognition errors are the main issues affecting the accuracy of the Hawkeye system. To address this challenge, the research improved DeblurGANv2 and HRNet, proposing a novel tennis game image deblurring and pose recognition model. When the fusion coefficient was 0.75 and the number of convolutional kernels was 45, the deblurring effect of the model was optimal, achieving an information entropy value of 11.2. At the same time, after sequentially improving DeblurGANv2 and HRNet, the mAP value of the combined model reached 92.48%, indicating that the improvement and fusion of each module in the study were effective. Compared with other deblurring and pose recognition models, this new model had a PSNR of up to 29.74dB, SSIM of up to 0.89, minimum parameter size of 4.53, and shortest running time of 0.25s, which was 78%-94% faster than SRN and MPRNet. Under different lighting intensities, the proposed model had strong robustness to complex poses and motion blur, showing a recognition accuracy of up to 92.44% and a video frame quality improvement rate of 16% - 18%. In a multi-person scenario, the model had the highest recognition AP value of 92.38%, and the shortest stable inference time was around 20ms. The lowest pose recognition tracking error was 5.27%. Although the overlap rate was higher than other models, the decision accuracy was 92.53%, far exceeding other methods. In summary, the model has significant advantages in both processing effectiveness and efficiency. However, the performance of the model still

fluctuates to some extent under extreme lighting conditions, such as ultra-low light or severe lighting environments. Future research will further optimize the robustness of the model and explore methods that combine multi-modal data to enhance its adaptability and generalizability in practical applications.

REFERENCES

- [1] J. Zhang, "Evaluation of the effect of artificial intelligence training equipment in physical training of table tennis players from a biomechanical perspective," *Mol. Cell. Biomech.*, vol. 21, no. 1, pp. 319-319, November 2024.
- [2] Y. M. Deng, and S. Y. Wang, "Biological eagle-eye inspired target detection for unmanned aerial vehicles equipped with a manipulator," *Mach. Intell. Res.*, vol. 20, no. 5, pp. 741-752, March 2023.
- [3] F. Meng, "Tennis video target tracking based on mobile network communication and machine learning algorithm," *Int. Trans. Elec. Energ. Syst.*, vol. 2022, no. 1, pp. 7447121-7447125, September 2022.
- [4] Y. Zhao, L. Lu, W. Yang, Q. Li, and X. Zhang, "Lightweight tennis ball detection algorithm based on Robomaster EP," *Appl. Sci.*, vol. 13, no. 6, pp. 3461-3462, March 2023.
- [5] Y. Yang, D. Kim, and D. Choi, "Ball tracking and trajectory prediction system for tennis robots," *J. Comput. Design Eng.*, vol. 10, no. 3, pp. 1176-1184, June 2023.
- [6] D. Gao, Y. Zhang, and H. Qiu, "Automatic detection method of small target in tennis game video based on deep learning," *J. Intell. Fuzzy Syst.*, vol. 45, no. 6, pp. 9199-9209, December 2023.
- [7] Y. Ke, Z. Liu, and S. Liu, "Prediction algorithm and simulation of tennis impact area based on semantic analysis of prior knowledge," *Soft Comput.*, vol. 26, no. 20, pp. 10863-10870, April 2022.
- [8] B. T. Naik, M. F. Hashmi, and N. D. Bokde, "A comprehensive review of computer vision in sports: Open issues, future trends and research directions," *Appl. Sci.*, vol. 12, no. 9, pp. 4429-4434, April 2022.
- [9] J. Bian, X. Li, T. Wang, O. Wang, J. Huang, and C. Liu, "P2ANet: A large-scale benchmark for dense action detection from table tennis match broadcasting videos," *ACM Trans. Multimedia Comput., Commun. Appl.*, vol. 20, no. 4, pp. 1-23, January 2024.
- [10] H. R. Ghezelsefloo, and S. H. Alavi, "The Impact of event-based sports technologies on the training and career development of referees in Iran volleyball super league," *Technol. Educ. J. (TEJ)*, vol. 16, no. 2, pp. 351-362, May 2022.
- [11] X. Peng, L. Tang, "Biomechanics analysis of real-time tennis batting images using Internet of Things and deep learning," *J. Supercomput.*, vol. 78, no. 4, pp. 5883-5902, October 2022.
- [12] H. C. Nguyen, T. H. Nguyen, and J. Nowak, "Combined YOLOv5 and HRNet for high accuracy 2D keypoint and human pose estimation," *J. Artif. Intell. Soft Comput. Res.*, vol. 12, no. 4, pp. 281-298, October 2022.
- [13] Y. Li, "Visualization of movements in sports training based on multimedia information processing technology," *J. Ambient Intell. Humanized Comput.*, vol. 15, no. 4, pp. 2505-2515, March 2024.
- [14] A. Fitzpatrick, J. A. Stone, S. Choppin, and J. Kelley, "Analysing Hawk-Eye ball-tracking data to explore successful serving and returning strategies at Wimbledon," *Int. J. Perform. Anal. Sport*, vol. 24, no. 3, pp. 251-268, December 2024.
- [15] T. Ning, C. Wang, M. Fu, and X. Duan, "A study on table tennis landing point detection algorithm based on spatial domain information," *Sci. Rep.*, vol. 13, no. 1, pp. 20656-20659, November 2023.
- [16] E. J. Jeong, J. Kim, and S. Ha, "Tensorrt-based framework and optimization methodology for deep learning inference on jetson boards," *ACM Trans. Embed. Comput. Syst.*, vol. 21, no. 5, pp. 1-26, October 2022.
- [17] C. Li, H. Li, L. Liao, Z. F. Liu, and Y. Dong, "Real-time seed sorting system via 2D information entropy-based CNN pruning and TensorRt acceleration," *IET Image Process.*, vol. 17, no. 6, pp. 1694-1708, January 2023.
- [18] P. Liu, Q. Wang, H. Zhang, J. Mi, and Y. Liu, "A lightweight object detection algorithm for remote sensing images based on attention mechanism and YOLOv5s," *Remote Sens.*, vol. 15, no. 9, pp. 2429-2430, April 2023.
- [19] S. Yan, C. Yang, and L. Guo, "Accuracy improvement in motion tracking of tennis balls using nano-sensors technology," *Adv. Nano Res.*, vol. 14, no. 5, pp. 409-419, May 2023.
- [20] L. Li, and A. Yang, "Correction algorithm of tennis dynamic image serving path based on symmetric algorithm," *Symmetry*, vol. 14, no. 9, pp. 1833-1834, Aug 2022.
- [21] C. Shen, and Z. Sun, "Research on target localization recognition of automatic mobile ball-picking robot," *J. Opt.*, vol. 51, no. 4, pp. 866-873, January 2022.
- [22] G. C. Domínguez, E. F. Álvarez, A. T. Córdoba, and D. G. Reina, "A comparative study of machine learning and deep learning algorithms for padel tennis shot classification," *Soft Comput.*, vol. 27, no. 17, pp. 12367-12385, February 2023.
- [23] C. Z, Q. Jiacheng, and B. Wang, "YOLOX on embedded device with CCTV & TensorRT for intelligent multicategories garbage identification and classification," *IEEE Sens. J.*, vol. 22, no. 16, pp. 16522-16532, August 2022.
- [24] X. Luo, Y. Wu, and F. Wang, "Target detection method of UAV aerial imagery based on improved YOLOv5," *Remote Sens.*, vol. 14, no. 19, pp. 5063-5067, September 2022.
- [25] Q. Wang, and N. Yao, "Light imaging detection based on cluster analysis for the prevention of sports injury in tennis players," *Opt. Quantum Electron.*, vol. 56, no. 2, pp. 191-192, December 2024.
- [26] S. Vancurik, and D. W. Callahan, "Detection and identification of choking under pressure in college tennis based upon physiological parameters, performance patterns, and game statistics," *IEEE Trans. Affect. Comput.*, vol. 14, no. 3, pp. 1942-1953, July 2022.
- [27] M. Niu, "Research on tennis-assisted teaching assessment technology based on improved dense trajectory algorithm," *Int. J. Netw. Virtual Organ.*, vol. 28, no. 2, pp. 154-170, September 2023.
- [28] W. Ren, "A novel approach for automatic detection and identification of inappropriate postures and movements of table tennis players," *Soft Comput.*, vol. 28, no. 3, pp. 2245-2269, January 2024.
- [29] W. Wu, "Multimodal emotion detection of tennis players based on deep reinforcement learning," *Int. J. Biometrics*, vol. 16, no. 5, pp. 497-513, September 2024.
- [30] J. Yao, X. Fan, B. Li, and W. Qin, "Adverse weather target detection algorithm based on adaptive color levels and improved YOLOv5," *Sens.*, vol. 22, no. 21, pp. 8577-8579, October 2022.
- [31] M. Skublewska-Paszowska, P. Powroznik, and E. Lukasik, "Tennis patterns recognition based on a novel tennis dataset-3DTennisDS," *Adv. Sci. Technol. Res. J.*, vol. 18, no. 6, pp. 159-176, May 2024.
- [32] X. Song, "Physical education teaching mode assisted by artificial intelligence assistant under the guidance of high-order complex network," *Sci. Rep.*, vol. 14, no. 1, pp. 4104-4109, February 2024.
- [33] A. Abba Haruna, L. J. Muhammad, and M. Abubakar, "Novel thermal-aware green scheduling in grid environment," *Artif. Intell. Appl.*, vol. 1, no. 4, pp. 244-251, November 2022.

A Highly Functional Ensemble of Improved Chaos Sparrow Search Optimization Algorithm and Enhanced Sun Flower Optimization Algorithm for Query Optimization in Big Data

Mursubai Sandhya Rani*, Dr. N. Raghavendra Sai

Department of Computer Science and Engineering, Koneru Lakshmaiah Educational Foundation,
Vaddeswaram, Andhra Pradesh, India.

Abstract—Numerous systems have to provide the highest level of performance feasible to their users due to the present accessibility of enormous datasets and scalability needs. Efficiency in big data is measurable in terms of the speed at which queries are executed physically. It is too demanding on big data for queries to be executed on time to satisfy users' needs. The query optimizer, one of the critical parts of big data that selects the best query execution plan and subsequently influences the query execution duration, is the primary focus of this research. Therefore, a well-designed query enables the user to obtain results in the required time and enhances the credibility of the associated application. This research suggested an enhanced query optimizing method for big data (BD) utilizing the ICSSOA-ESFOA algorithm (Improved Chaos Sparrow Search Optimization Algorithm- Enhanced Sun Flower Optimization algorithm) with HDFS Map Reduce to avoid the challenges associated with the optimization of queries. The essential features are extracted by employing the ResNet50V2 approach. Effective data arrangement is necessary for making sense of large and complex datasets. For this purpose, we ensemble Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Improved Spectral Clustering (ISC). The experimental findings demonstrate a significant benefit of the proposed strategy over the present optimization of the queries paradigm, and the proposed approach obtains less execution time and memory consumption. The experimental results show that the proposed strategy significantly outperforms the current optimization paradigm, reaching 99.5% accuracy, 29.4 seconds of execution time, and 450 MB less memory use.

Keywords—Big data (BD); query optimization; Improved Chaos Sparrow Search Optimization Algorithm (ICSSOA); Enhanced Sun Flower Optimization Algorithm (ESOA); ResNet50V2; DBSCAN

I. INTRODUCTION

Big data empowers businesses to make informed decisions and take appropriate action by allowing them to examine enormous data in volume, variety, and velocity [1]. Big data can be stored and queried using a variety of databases and data structures: Relational databases are employed for read-intensive analytic queries; Internet transaction processor platforms are utilized for faster uploads and reliability; NoSQL storage systems are used for handling massive volumes of data [2, 3]. Different data stores have been created and constructed for various purposes and the best results. SQL databases are effective at storing and processing structured data, but their

efficiency suffers from read-intensive queries. Similarly to how NoSQL storage systems are tailored to deal with unstructured data, columnar databases are utilized for the analytic processing of queries [4-6].

The information that has been processed is kept in several databases so that analysts can use it. Performance optimization and various data structures are crucial for applications that use a lot of data [7, 8]. Building scalable and effective data pipelines is a significant difficulty. These data pipelines, which are vital to the functionality of the applications, are optimized and maintained by data engineers [9]. Researchers and data scientists utilize the data warehouse to analyze, evolve, and load the data for their research projects. The enhancement of query efficiency and extra complexity brought on by the various data models employed in these databases present ongoing challenges for big data platforms that use these databases [10-12].

The many Operation SITE Allocation (OSA) strategies to execute the query are born from the advancement of query optimization. OSA problems are sought after to improve query execution plans in terms of system throughput or response times [13]. The query optimizer's three main parts are "Cost Model," "Search Space," and "Search Strategy." Designing the various cost coefficients and the objective function is the responsibility of the cost model. A variety of different query execution strategies are represented by the search space [14, 15]. The search method is also used to probe the search space to find the most promising query execution technique.

Previously, deterministic optimization methods and a variety of databases were used for query optimization. Only basic CDSS queries are a good fit for deterministic algorithms [16-18]. Nature Inspired Computing (NIC) has tremendous prospects for computational intelligence and is now being applied to address CDSS query optimization concerns. There is a long list of NIC computing techniques, some of which depend on the genetics of animals, insects, birds, and people, as well as on music and water [19]. The most admired NICs include Artificial Bee Colony, Cuckoo Search, Ant Colony Optimization, Grey Wolf Algorithm, and Genetic Algorithm. After reviewing the literature on query optimization, it was discovered that distributed CDSS queries had received a lack of attention. To speed up the data retrieval, a creative query

optimizer is required. The suggested query optimizer helps identify an ideal query execution plan that reduces the overall consumption of I/O, computing, and communication resources [20].

The increasing scale and complexity of big data have made query optimization a critical challenge. Existing methods often struggle with several limitations, including high computational cost, slow convergence, and inefficiency when handling large, distributed datasets. Many traditional techniques are also unable to address data skew effectively, ensure quick response times, or optimize query execution under heavy query loads. These shortcomings highlight the need for a more efficient approach to query optimization that can scale with growing data volumes and provide faster, more resource-efficient execution in modern big data environments. To address these challenges, we propose an enhanced query optimization method that significantly improves execution time and reduces memory consumption, making it better suited for the demands of today's data-driven applications.

To tackle the issue mentioned above, we introduced a novel approach to big data arrangement and feature extraction. This reduces the execution time, retrieval time, and memory usage. Compared with existing methods, the proposed approach performs better.

A. Research Contribution

The key objectives of this research are as follows:

- Initially, we employed a secure hash algorithm in preprocessing to find the hash value. Then, centered on the HV, the map reduction process is executed.
- After the removal of repeated data, the essential features are extracted by employing ResNet50V2.
- Entropy values are inputted to the deep adaptive hybrid clustering algorithm DBSCAN and spectral clustering for the big data arrangement.
- Finally, the query is optimized with the help of the ensemble Improved Chaos Sparrow Search Optimization algorithm (ICSSOA) and Enhanced Sun Flower Optimization algorithm (ESFOA).

The following part of the article is structured as follows. The existing prior works are briefly described in Section II. The proposed strategy is described in detail in Section III. The suggested method is extensively simulated in Section IV. Section V provides the conclusion.

II. RELATED WORKS

Some existing prior works related to significant data query optimization are analyzed in this section.

An improved query optimizer known as CDSS was modelled by Sharma et al. [21] using a hybridization firefly-genetic algorithm (GA) on a constrained divergence environment (RDFG_CDQO). This CDSS was created with the goal of achieving the best query execution plan possible to reduce processing, input-output, and interaction demands when running CDSS queries. The controlled GA's slower convergence difficulty would be cautiously defeated by the

enhanced utilization of the CDSS technique, achieving significant variance in "2" successive generations. The CDSS optimizer could not solve the QO issues. For the query retrieving rate, Lekshmi et al. [22] presented the Top-k Query Multi-Keyword Threshold method (Top-k QMKST). The query and many keywords are primarily divided, and B+ tree indexing was used to execute the data index. Response time and spatial complexity were both decreased by employing Top-k QMKST. The Kullback Leibler Divergence also uses the index list of terms to determine a score value. The results of the experimental study show that the suggested technique performs better.

For the skewed-ranging queries, Wei Ge et al. [23] suggested a method known as correlation-aware partitions. In the form of a geometrical curve-fitting problem, it introduced a problem known as partitioning optimization on continuously correlated data. The boundaries of the range query must be used to partition data optimally. The boundary for the range was utilized in this case to incorporate the best partitions and significantly reduce the computational cost compared to the standard dynamic programming. When compared to the global one, the local one performed better instead of attempting to increase effectiveness.

Sinha et al. [24] proposed an approach for distributed datasets by combining the genetic algorithm (GA) and the k-means clustering method. The suggested strategy is divided into two phases; in the initial stage, parallel GA is performed to data chunks spread across many machines. GA takes into account the covariance among the data sets and offers an improved summary of the original information. Phase 2 applies K-means with K-means++ initialization on the intermediate output to produce the outcome.

Ansari et al. [25] suggested a parallel variant of the conventional K-means algorithm for use in the Hadoop distributed environment. The results of the experiments demonstrate that the suggested K-means algorithm operates better than conventional K-means when clustering a significant volume of datasets. Compared to current methods, the suggested approach produces better results.

A. Research Gap

Existing query optimization techniques, including Top-k QMKST (Lekshmi et al. [22]) and the CDSS optimizer (Sharma et al. [21]), concentrate on increasing query execution efficiency but struggle to handle dynamic or large-scale datasets. Top-k QMKST speeds up response times but might not be able to handle high-dimensional data effectively, and the CDSS optimizer enhances convergence but has trouble optimizing query retrieval rates. Other methods that deal with partitioning and data summarization, including correlation-aware partitions (Wei Ge et al. [23]) and the integration of evolutionary algorithms with K-means clustering (Sinha et al. [24]), do not sufficiently improve query execution in distributed systems with big datasets. Furthermore, the parallel K-means approach of Ansari et al. [25] enhances clustering but ignores memory usage and query execution time. By using ResNet50V2 for feature extraction, the ICSSOA-ESFOA method for improved query optimization, and DBSCAN and ISC in combination for efficient data arrangement, our

proposed work seeks to close these gaps. By addressing the shortcomings of current techniques, our strategy guarantees quicker query execution, better memory management, and increased scalability in significant data contexts.

III. PROPOSED METHODOLOGY

In order to handle and store BD, which is extremely large in volume and contains numerous data models, organizations

maintain various databases. For business purposes, it is essential to query and analyze BD for insight. In this study, the ICSSOA-ESOA algorithm and the HDFS map-reduce approach were used to improve the query optimizer procedure in BD.

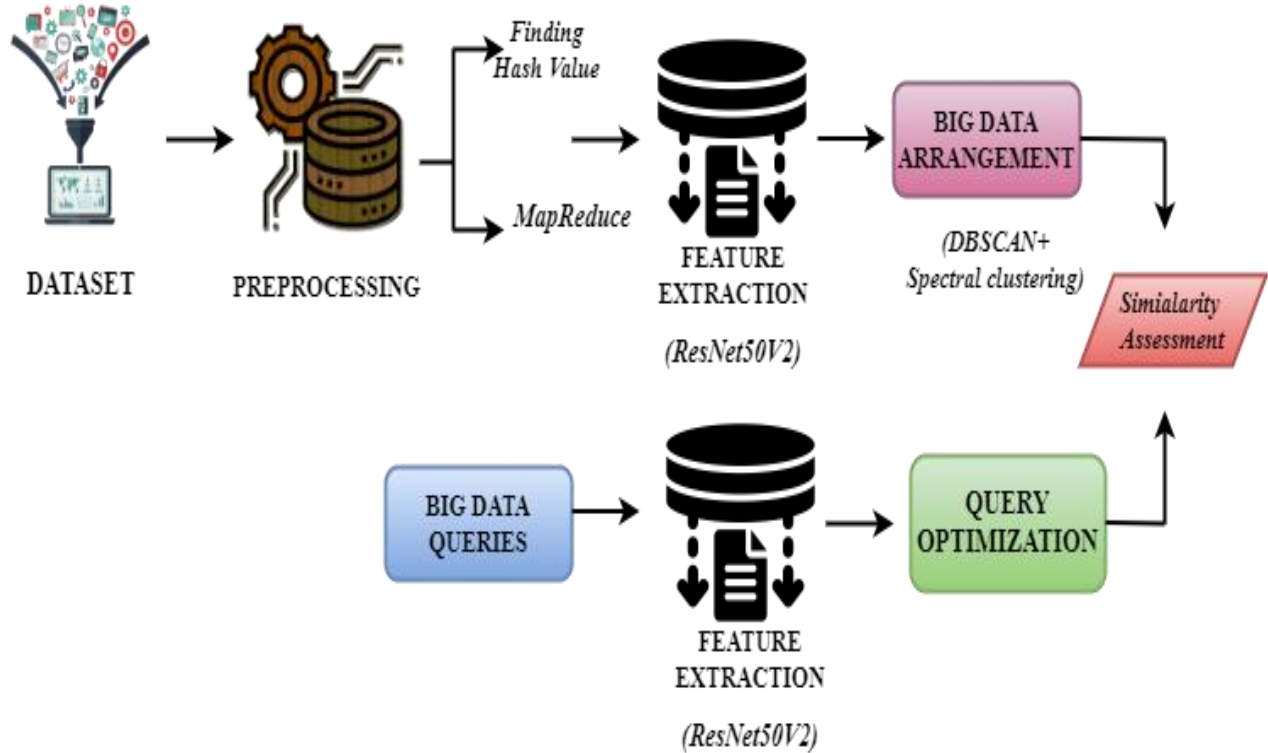


Fig. 1. Proposed methodology architecture diagram.

To extract the essential features from a big dataset, we employed ResNet50V2. Then, the big data are arranged with the help of an ensemble DBSCAN approach and an improved spectral clustering approach. The proposed approach is analyzed and evaluated by using four benchmark datasets. The overall framework of the proposed approach is shown in Fig. 1.

A. Problem Statement

The number of datasets that need to be evaluated is increasing, necessitating several databases to store the preprocessed data in various information formats. Several methods, like materialized views and data cubes, can decrease query latency but necessitate significant computation and preparation. In order to deliver estimated results with error bounds, approximate query processing (AQP) was implemented. Nowadays, the majority of AQP models only support one database. The suggested AQP model supports heterogeneous databases with various data models by keeping up-to-date samples in a single database. Any database can be used to conduct the SQL query. The query optimizer chooses the samples automatically and provides users with approximations of the results. For this purpose, we introduced a novel approach for query optimization.

B. Preprocessing

The pre-processing of the input data was carried out during this phase. First, it uses the Secure Hash Algorithm (SHA-3) to determine the HV for every bit of data. Then, using HDFS, the MR process is carried out using the HV as its focal point. The subsections below explain the SHA-3 and HDFS processes. The SHA-3 algorithm is specified for a digest length d with a value of 224, 256, 384, or 512 and a message M with two bits "01" inserted at the conclusion, such that $SHA - d(M) = KECCAK(c)(M||01, d)$, while SHA3 and KECCAK are functions, M is the input string to the SHA-3 method.

1) *The SHA-3 algorithm is utilized to find the hash value of big data:* Utilizing permutation functions, the SHA-3 method, also referred to as the Keccak algorithm, was created. Keccak performs encryption well and has a high degree of attack resistance. SHA-3 is safer than earlier iterations like SHA-1 and SHA-2. The SHA-3 method can provide multiple fixed-bit hash values for different input bits. The outcome of this research is a 256-bit hash value.

2) *Map and reduce:* The two most crucial MapReduce processes are the "Map and Reduce" operations. The Apache

Foundation created the distributed system infrastructure known as Hadoop. Users can fully leverage the platform's massive data storage and quick computation capabilities by developing distributed applications without familiarity with the architecture's inner workings. Hadoop implements a distributed file system called HDFS. Although HDFS requires the usage of costly hardware, it provides good features and strong fault tolerance. Additionally, it offers a fast interface for accessing application data, making it appropriate for programs with big

data sets. HDFS lowers the file system's restrictions for accessing the data in stream form. HDFS and Map Reduce are the two main components of the Hadoop system. Massive data storage is primarily provided by HDFS, and distributed computing functions are supplied by Map Reduce. The simple description of Hadoop's data processing is that the Hadoop cluster analyzed the data to produce its outcomes. In Fig. 2, the method of processing flow is depicted.

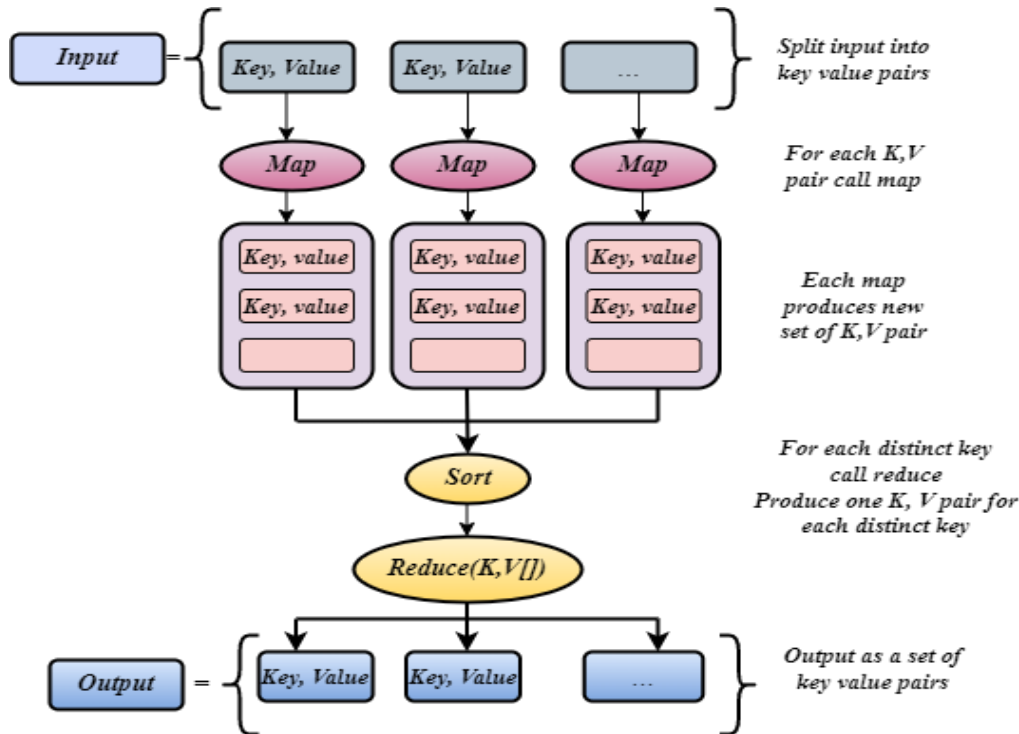


Fig. 2. Framework of map and reduce.

HDFS and MapReduce are the two main parts of Hadoop, as shown in Fig. 3. The storage of enormous amounts of data is the responsibility of HDFS, and the processing of massive amounts of data is a function of MapReduce. Another two crucial parts of Hadoop are the distributed database system Hbase and the data warehouse tool Hive. Records are kept in a Hadoop cluster using the HDFS. The HDFS interface resembles a straightforward hierarchical file system with straightforward operations like adding, deleting, moving, and more. However, the HDFS files are broken up into data blocks based on specific requirements, and then a massive number of data blocks are distributed over numerous slave nodes. It departs significantly from conventional storage structures at this point. The user typically chooses the number of data blocks to put and the dimension of each separated data block.

MapReduce, which includes Job Trackers and Task Trackers, is DFS's top layer. Massive files are partitioned into equal sections by default on HDFS. This default value is set at 64 M in the HDFS overview document. The data file 1 has been separated into three portions and placed in three distinct machines. Map Reduce is a task that is called Map and

computes after every Hadoop input component. The system will move through each input data individually in the task before analyzing the map and turning it into a key-value format. The outcome will be produced in the key-value pair's form. As an input to Reduce by key, Hadoop will then transmit the outcome of the preceding phase. The Reduce Task's results, retained on HDFS, are the outcome of the entire task.

C. Feature Extraction

Datasets in big data scenarios may contain a large number of variables or attributes and be exceedingly high dimensional. High dimensionality can present difficulties in overfitting, poor interpretability, and computation complexity. Feature extraction algorithms can reduce dimensionality by converting the original features into a lower-dimensional representation while maintaining the crucial data. Analysis and modelling could become more effective as a result. From the original data, the significant aspects are retrieved, including closed frequent item set, support, and confidence. Finally, entropy computation is used to regulate confidence and support value. The following part provides an overview of the extraction of feature processes.

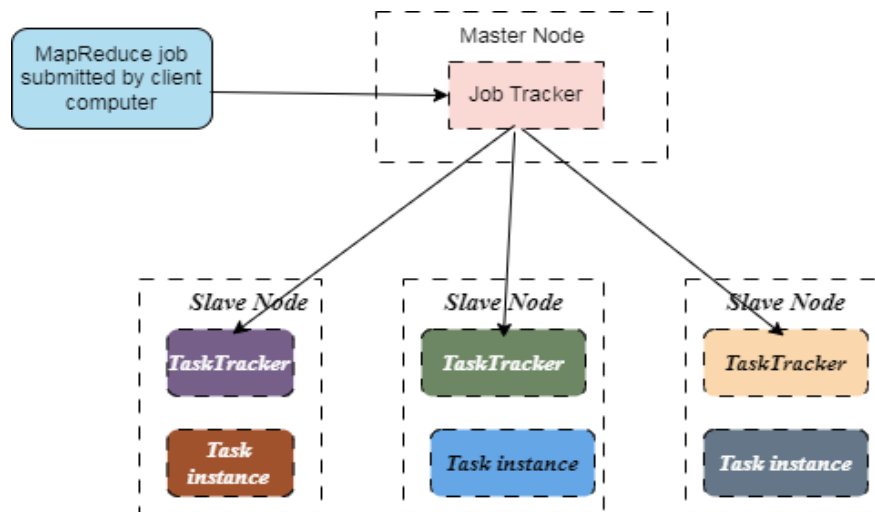


Fig. 3. Hadoop's two core components.

1) *ResNet50V2*: Deep feature extraction is illustrated in this subsection. Deep feature extraction employs deep neural networks to extract significant and valuable information from raw data. These characteristics capture high-level representations that are more useful for handling the current task. For query optimization in big data, we used the *ResNet50V2* framework as a deep extraction of features method. *ResNet50V2* represents a convolutional neural network (CNN) that excels in various computer vision applications. To tackle the degradation issue in deep networks, a variation of the *ResNet* design is used, which uses skip connections.

The 50-layer *ResNet50V2* was pre-trained using a sizable dataset, such as a big datasets. The network can learn residual mappings through the use of residual blocks, which also makes it easier to train deeper networks. The skip connections also facilitate the direct transfer of gradients from the initial layers to subsequent layers, which improves training. Due to its ability to extract complicated and structured patterns from big data, the *ResNet50V2* architecture is advantageous for query optimization feature extraction.

The deep layers of *ResNet50V2* enable it to learn abstract representations. The benefit of Transfer Learning may be obtained by utilizing the pre-trained *ResNet50V2* approach, as it has previously acquired general features from a sizable dataset like the hospital compare, Twitter, and IMDb datasets. *ResNet50V2* can record generalized representations tuned for query optimization owing to this pre-training. The precision and effectiveness of the query optimization can be improved by applying the learned features from *ResNet50V2*.

The *ResNet50V2* features provide a more advanced representation of the input optimization of queries, capturing essential data for positions, including bid arrangement of data. We may utilize the potent representations learned by *ResNet50V2* by using these features as inputs for multiple machine learning algorithms. By doing that, we want to improve the precision and functionality of our query optimization mechanism. *ResNet50V2*'s high-level features

enable a more thorough and insightful representation of the input data, enhancing our capacity and eventually enabling improved optimization.

D. Big Data Arrangement

Big data arrangement is a key component of the data management process, which involves structuring and organizing enormous amounts of data to facilitate effective analysis, storage, and retrieval. For clustering and pattern recognition tasks in data analysis and deep learning, ensemble *DBSCAN* (Density-Based Spatial Clustering of Applications with Noise) and Improved Spectral Clustering can be particularly beneficial. Combining *DBSCAN* with Spectral Clustering can take advantage of each technique's advantages as each approach has advantages and disadvantages of its own. The proposed method achieves improved noise handling, improved cluster separation, scalability, merging local and global information, handling variable cluster Densities, and more while combining the methodologies.

1) *DBSCAN clustering algorithm*: *DBSCAN*, a popular density-based clustering technique, can locate several clusters based on the predicted density distribution. It can detect shaped clusters and does not require prior knowledge of the cluster size. The following examples show the core concept of *DBSCAN*. *DBSCAN* collects all points in the neighbourhood of a random, unvisited point called p , while p is the initial location and r is the neighbourhood's maximal radius. The minimal number of units needed to generate a dense zone is called the density threshold $MinPts$. If $MinPts$ points or more are nearby, point p is a core point. All of the points in p , ϵ -neighbourhood are put into an identical cluster if p is the centre point together with all of the other points in p . *DBSCAN* locates all density-reachable points. It includes them in the same cluster for every point in the cluster. If point q is densely accessible from other core points but has a smaller neighbourhood than $MinPts$, it is also a border point that belongs to the cluster. An isolated or noisy point cannot be reached from any other point. Using consecutive cluster extraction, *DBSCAN* completes the clustering procedure. A finalized cluster is created by iterating

this procedure till no more density-reachable spots are discovered. The three categories that DBSCAN uses to categorize a set of points are noise, low-density boundary points, and high-density core points. The following are three different types of points' definitions.

2) *Initialization of the variables:* In K-DBSCAN, the HS is optimized to get the best clustering parameters. Thus, "Eps" and "Minpts," the two clustering parameters for input, have been utilized as the HS's decision variables, correspondingly. Given that the set of data is split into categories that are considered as K, every parameter variable's maximum value shouldn't be greater than the K-equal partitions of the entire data set. These two variables are initialized with the following values:

$$Eps \in \left(0, \frac{SDR}{2 \times K}\right) \quad (1)$$

$$Minpts \in \left[1, \frac{Num_obj}{\left(\frac{LDR}{SDR}\right) \times K \times D}\right] \quad (2)$$

While *SDR* and *LDR* are the smallest value and greatest values across all dimension that ranges from the entire data set, accordingly, the variable shows the number of objects utilized for clustering *Num_obj*. The dimension is denoted by *D*.

3) *The objective function:* A multi-objective collaborative evaluation approach is provided for the HS in the K-DBSCAN optimization issue. The overall number of clusters produced by DBSCAN under different parameter variables is monitored by

using the initial target function, which can be shown as the total amount of variance among that and the determined clustering number K. Since the main objective of this clustering approach is to produce K groups, this variance can be expressed as the total amount of variance between it and the established clustering number K.

$$Minimize f1 = |c - K| \quad (3)$$

The total number of clusters is represented as K, which has been predetermined, and the real number of clusters is indicated as c DBSCAN, which has been produced using the current set of decision variables.

The DBSCAN method can identify unusual noise. When the outcomes of the parameters "Eps" and "Minpts" are improperly chosen, particularly if they are disproportionately matched, it may result in under-differentiation, where most or even all of the data items are misidentified for outliers.

Two distinct groups make up the initial data set in Fig. 4, and Fig. 5 displays the results of clustering with excessive noise caused by subpar clustering parameters. Acquiring the cluster number of 2 is possible, although many valid points are confused for noise entities. Consequently, a separate function of the multi-objective optimization method is utilized to maximize the number of objects in the least efficient cluster and prevent such an abnormal occurrence.

$$Maximize f2 = num(s_clusters) \quad (4)$$

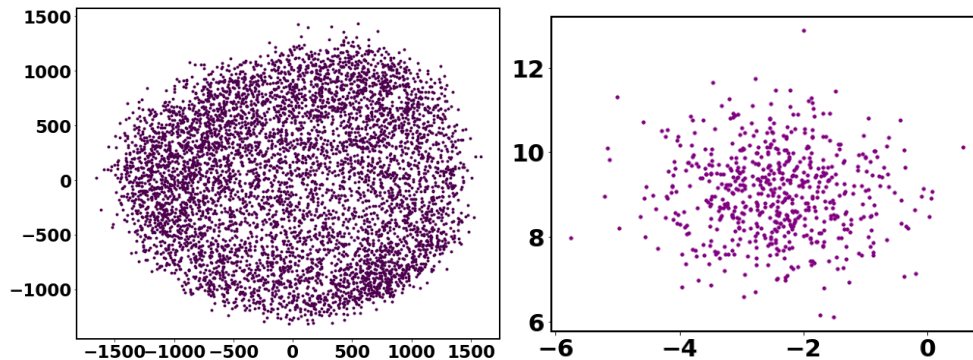


Fig. 4. The initial formation of the dataset.

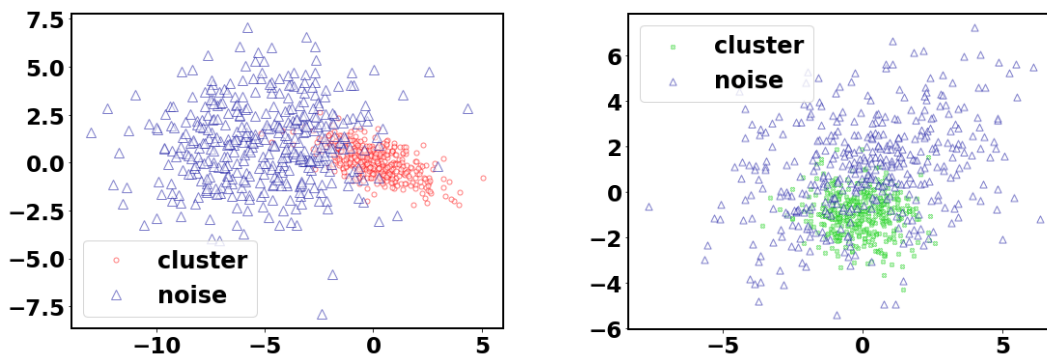


Fig. 5. Noise in clusters.

The term $num(s_clusters)$ refers to the number of items in the smallest practical cluster. Consequently, the following is an expression for the K-DBSCAN's multi-objective collaborative evaluation function:

$$F = (\text{Minimize } f1, \text{ Maximize } f2) \quad (5)$$

Obtaining the necessary K clusters is the primary objective of K-DBSCAN. This is followed by the effect of clustering that produces the fewest inaccurate noise objects. In other words, $f1$ has a greater priority than $f2$, which is indicated by the notation: $f1 \prec f2$.

4) *Framework*: According to the information above, the two clustering factors, "Eps" and "Minpts," are used in DBSCAN as the HS variables for decision-making. The multi-objective collaborative evaluation function can be used with the clustering parameter's optimal value to get a superior clustering outcome with K categorization when using DBSCAN.

Additionally, relatively low parameter values typically result in a superior clustering effect when using the DBSCAN algorithm. The size of "Minpts" indicates a significant impact on how well noise of clustering is judged under the condition of a specific parameter "Eps," and the larger it is, the more probable it is that genuine data will be viewed as noise objects. Thus, the variable of decision "Minpts" has been set to a number that enhances over time with the repetition stage process to acquire adequate clustering factors, including.

$$Minpts = Minpts_{min} \dots Max_{min} \quad (6)$$

While $gnit$ denotes the number for the generation currently in use, NI is the maximum number of repetitions and $Minpts_{max} \dots Minpts_{min}$ denotes the variable upper and lower bounds, accordingly.

5) *Spectral clustering*: Typical graph-based clustering techniques include Spectral Clustering without monitoring the data. Techniques for Spectral Clustering often start with local data that has been encoded in a weighted network of information and then aggregates according to the associated similarity matrix's global characteristic vectors. In Spectral Clustering, a function of mapping that explicitly maps characteristics to the group tag matrix is automatically learned for every task to anticipate cluster tags.

The process of learning can automatically use dissimilar data to enhance clustering efficiency. In Spectral Clustering, communities of nodes connected near one another are characterized in a graph using a method known as clustering. The nodes are placed in a low-dimensional area that can be easily segmented into clusters. Affinity, Degree, and Laplacian matrices and other specific values of these matrices produced from a graph or data collection are used in spectral clustering. The crucial steps in creating a Spectral Clustering algorithm are as follows:

Prior to using the spectral clustering procedure, we must first Figure out the matrix for similarity, which is then indicated as the overlap matrix of degree P. It can be shown as,

$$p = \begin{bmatrix} 0 & p_{1,2} & \dots & \dots p_{1,n} \\ p_{2,1} & 0 & \dots & \dots p_{2,n} \\ \vdots & \dots & 0 & \vdots \\ p_{n,1} & \dots & p_{n,n-1} & 0 \end{bmatrix} \quad (7)$$

For the arrangement criterion, we may assume that every request is split into k_1, k_2 , and two groups; this work employs the conventional division approach. Suppose q is a vector. These are the definitions of the q_i elements:

$$q_i = \begin{cases} \sqrt{\frac{d_2}{d_1 d}} & , i \in k_1 \\ -\sqrt{\frac{d_1}{d_2 d}} & , i \in k_2 \end{cases} \quad (8)$$

In the event that the cluster indicator matrix $F \in R^{n \times k}$ is correct. Assuming consistent with each perspective, we can define the clustering of spectral data issues as,

$$\min_{F, F^T F = 1} \sum_{v=1}^t Tr(F^T L^v F) \quad (9)$$

While every graph evenly contributes to the outcome F. We ignore the specifics of the graph creation in the equation above. Several additional studies just take the mean of the vertices and then implement the spectral clustering independently instead of mandating that multiple graphs share the same F.

Improved Spectral Clustering Algorithm (ISCM). We provide an improved spectral clustering technique (ISCM) relying on the enhanced k-means algorithm. The approach accomplishes secondary clustering in addition to resolving the initial value issue. We take into account the parameters as previously mentioned in accordance with the QoS criterion. We may determine whether secondary clustering is necessary by evaluating the variable sizes before the method operates. There is no need to recluster if the present QoS exceeds the users' desire to allocate resources once the strategy has been performed. The clustering spectral optimization scheduling algorithm's implementation procedures are then described.

E. Query optimization

Big data systems frequently handle enormous amounts of data. By dramatically reducing the time it takes for a query to execute, query optimization can guarantee that users or applications can quickly and effectively retrieve the needed data. Query optimization aids in efficient resource allocation, cutting costs and guaranteeing the best use of available resources. It minimizes hardware waste and prevents nodes from becoming overloaded. For this purpose, we ensemble the Improved Chaos Sparrow Search algorithm (ICSSA) and Enhanced Sun flower optimization algorithm (ESFO). ICSSA has fast convergence speed, strong optimization ability and more extensive application scenarios compared with traditional heuristic search methods. Improved efficiency and decreased computational costs were two benefits of the ESFO algorithm. We ensemble both algorithm's merits to effectively optimize the query.

1) *Sparrow search algorithm*: The SSA bases its description of the sparrows' predatory and anti-predatory behavior for updated locations on the following guiding concepts. The population of sparrows is split into followers and

producers. The sparrow's two identities may be switched around, and everyone has a system for detecting danger. Every sparrow, in particular, is sensitive to potential threats or natural enemies and will immediately begin anti-predatory activity to defend itself. The producers are highly active, adept at foraging for food, travel widely, and lead other sparrows on their quest. To increase their food intake by snatching it or foraging nearby, seekers seek the producer and follow them to find additional food.

2) *Basic concepts*: The individual matrix is displayed below, with N sparrows assumed to be in D-dimensional space.

$$X = [x_1, x_2, \dots, x_N]^T, x_i = [x_{i,1}, x_{i,2}, \dots, x_{i,D}] \quad (10)$$

While x_i , D denotes the i^{th} sparrow's location in the D dimension.

$$x_{i,j}^{t+1} = \begin{cases} x_{i,j}^t \cdot \exp\left(\frac{-i}{\alpha \cdot \text{iter}_{\max}}\right) \cdot O_2 \\ x_{i,j}^t + Q \cdot L \quad R_2 \geq ST \end{cases} \quad (11)$$

The present iteration count, t, is represented here. Itermax indicates the greatest amount of the iterations $j = 1, 2, \dots, d$. It falls between 0 to 1 and is a uniform randomized value. The warning and security values for sparrows are represented by $R_2 (R_2 \in (0,1))$ and $ST (ST \in (0.5,1.0))$. An ordinary distribution characterizes a random number Q. Every matrix's 1d elements comprise the L matrix. When this $R_2 < ST$ occurs, the provider adopts a wide-area search phase while they are not in danger from any natural competitors and are in a generally safe environment. Due to Eq. (3), the follower position is upgraded.

$$x_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\frac{x_{\text{worst}}^t - x_{i,j}^t}{i^2}\right) & i > \frac{n}{2} \\ x_p^{t+1} + |x_{i,j}^t - x_p^{t+1}| \cdot A^+ \cdot L & \text{otherwise} \end{cases} \quad (12)$$

In which x^t worst indicates the current position of the bird with the worst adaptability. The spot of the bird with the best producer adaption is represented by the number x_p . Every component of the matrix shown by A is represented by a value at random of one or zero. A^+ equals $A^T A A^{T-1}$.

3) *Danger awareness mechanism*: When hunting, sparrows sense the danger of hunt and may fly away from their current location and to another. The individual sparrows that detect danger often range between 10% and 20%. As the Eq. (4) shows, the sparrows' posture changes when they detect danger.

$$x_{i,j}^{t+1} = \begin{cases} x_{\text{best}}^t + \beta \cdot |x_{i,j}^t - x_{\text{best}}^t| & f_i > f_g \\ x_{i,j}^t + K \cdot \left(\frac{|x_{i,j}^t - x_{\text{worst}}^t|}{(f_i - f_w) + \varepsilon}\right) & f_i = f_g \end{cases} \quad (13)$$

The current optimal location is represented as $x_{\text{best}} \cdot \beta$ is a common control parameter for properly distributed random step algorithms. K is an even random number with the value (1, 0). f_i shows the sparrow's value of current fitness. The current best-fit and worst-fit values globally are denoted by f_g and f_w , accordingly. The least significant is indicated as ε . If $f_i > f_g$, it

means that the particular sparrow is on the periphery of the population and is hence vulnerable to assault by predators of nature.

4) *Improved chaos sparrow search optimization algorithm*: In the case of the standard SSA, the producer fails to thoroughly search for the best possible outcome in the initial iteration, and the solution in the later iteration has a marginally lower precision as a result of the producer's poor management of the earlier repetition and the creation of the afterward iteration in the global search. Blindly adopting the producer's perspective, the followers rapidly enter the local optimal conundrum, reduce population diversity, and become the producers. Enhancing population variety is the major way to keep the dynamic equilibrium of provider search and development to handle the aforementioned issues. ICSSOA research is concentrated on finding ways to make it easier to leave local optima. The following topics will be covered in detail to understand the ICSSOA.

5) *Cubic chaos mapping*: Algorithms have been optimized using Chaos, a nonlinear process that occurs in nature. Because of its stochastic and ergodic characteristics, it enhances population variety and makes it easier for the approach to depart from the optimum for local. The standard version of the chaotic mapping, known as cubic mapping, is presented in Eq. (14).

$$x_{n+1} = bx_n^3 - cx_n \quad (14)$$

Where the effect variables for chaos are b and c. While $c \in (2.3,3)$ the chaos sequence is produced via cubic mapping. The Cubic mapping expression was modified by studying the max exponent of Lyapunov for 16 frequent mappings of chaos. The experimental findings showed that Cubic mapping has less disorder than one-dimensional mappings like Sine mapping and Circle mapping but is more chaotic than worm mouths and tent mappings. It can be expressed as,

$$x_{n+1} = \rho x_n (1 - x_n^2) \quad (15)$$

While $x_n \in (0,1)$ and the parameter for control is represented as ρ .

6) *Adaptive weighting factor*: A higher weight of inertia is required in the iterations to extend the discoverer's worldwide range for searching since the producer undertakes global exploration as rapidly as feasible to determine the global ideal solution. Simultaneously, a lower inertia weight is required in the latter iterations to enhance the discoverer's local exploitation capabilities to speed up convergence and prevent settling on the optimal local solution. As a result, the supplier location upgrade is proposed to be improved by fusing adaptive weights, and the supplier location enhancement formula is illustrated as,

$$x_{i,j}^{t+1} = \begin{cases} \omega \cdot x_{i,j}^t \cdot \exp\left(\frac{-i}{\alpha \cdot \text{iter}_{\max}}\right) \cdot O_2 \\ \omega \cdot x_{i,j}^t + Q \cdot L \quad R_2 \geq ST \end{cases} \quad (16)$$

The exact computation of ω is displayed as

$$\omega = \begin{cases} w_0 & t \leq t_0 \\ \left(\frac{1}{t}\right)^{0.9} & t > t_0 \end{cases} \quad (17)$$

While ω_0 is the actual positive number. The present iteration count is represented as t . The amount of iterations is indicated by t_0 . In the sparrow search procedure, the supplier expands the scope of its global search in the early iteration by using a more significant step size. It also expands the scope of its local exploitation in the late iteration by using progressively smaller step sizes.

7) *An ensemble method for levy flight and reverse Learning*: A category of stochastic non-Gaussian phenomena is called Levy flight. A heavy-tailed random path distribution describes the likelihood distribution of step length. For SI optimization techniques prone to encountering the issue in optimum of local, Levy flight can potentially allow the approach to significantly deviate from the local optimal significance, with a more significant likelihood of doing so in the random path. Based on Levy flight, the sparrow location upgrade algorithm is displayed as

$$x_{newi}^{t+1} = x_i^t + \gamma \oplus Levy(\lambda) \quad (18)$$

Where the phase parameter for control is represented as γ . A randomized path search is Levy (λ).

$$Levy = t^{-\lambda} \quad 1 < \lambda \leq 3 \quad (19)$$

The generation stage is depicted as

$$S = \frac{\mu}{|v|^{\beta}} \quad 1 \leq \beta \leq 2 \quad (20)$$

Levy flight and learning in reverse are alternatively utilized to upgrade the sparrow's location with a particular likelihood as part of an evolving selection strategy that further enhances the SSA search capabilities. This approach depends on the above two methodologies. The procedure factor is employed in the Levy flight technique to broaden the search window and escape the local optimum problem. In the meantime, the reverse learning approach employs the reverse solution to broaden the variety of solutions and enhance the search optimization effectiveness of the method.

8) *ICSSOA time complexity analysis*: For the individual setup and variable setting in SSA, the temporal magnitudes are n and C . If the total number of dimensions is k , the sparrow fitness ranking and creator spot provided time magnitudes are $n \times \log_2 k$ and $n \times k$, respectively. The remaining birds' positions as followers must be updated during the follower location updating phase, and a period schedule of $n \times k$ must be used to determine whether every person's dimension is within bounds.

During the alert sparrow location upgrade stage, a random sample of sparrows is chosen for positioning, and a determination is performed when every dimension of a given individual is outside of acceptable limits concerning time magnitude $n \times k$. In conclusion, the magnitude of the provider location upgrade time is $n \times \log_2 k + n \times k$. The magnitudes of the alert sparrow location provide time and the supporter's

location upgrade time are both $n \times k$. The enhanced algorithm's temporal complexity is,

$$O(n \times k + n \times \log_2^n k + n \times k + n \times k + n \times k + n \times k + n \times k + n \times k + n \times k) \approx O(n \times \log_2^n k) \quad (21)$$

9) *Sun flower optimization algorithm*: An individual-based heuristic algorithm, the SFO draws its inspiration from nature. Its fundamental idea is to mimic how sunflowers would position themselves to receive solar light. A sunflower has a daily recurring sequence. They travel toward the sun as the day gets going. They travel in the other direction in the late hours. Single pollen gamete is thought to be produced by every sunflower. The minimum distance among flowers i and $i + 1$ was randomly used as the pollination route. Every blossom patch regularly releases a billion pollen gametes in the real world. For the sake of simplicity, we also presumptively assume that every sunflower generates a single pollen gamete and develops separately. The directions of the sunflowers concerning the sun are shown below.

$$\vec{s}_i = \frac{X^* - X_i}{\|X^* - X_i\|}, \quad i = 1, 2, \dots, n_p \quad (22)$$

Eq. (23) depicts the sunflowers moving in the direction indicated by s .

$$d_i = \lambda \times P_i(X_i + X_{i-1}) \times \|X_i + X_{i-1}\| \quad (23)$$

The pollination likelihood $P_i(\|X_i + X_{i+1}\|)$ is expressed as λa constant, which describes the "inertial" motion of the sunflowers. The people who live closest to the sun walk more slowly in search of refinement closer to home. The motions of the people further away are normal. Eq. (24) introduces the limitation of the following steps:

$$d_{max} = \frac{\|X_{max} - X_{min}\|}{2 \times N_{pop}} \quad (24)$$

The overall individuals of the plants X_{max} X_{min} are lower and upper bounds, and their locations are all given as N_{pop} . This equation yields the new plant:

$$\vec{X}_{i+1} = \vec{X}_i + d_i \times \vec{s}_i \quad (25)$$

10) *ESFOA concept and mathematical representation*: The idea behind the Enhanced Sunflower Optimization Algorithm (ESFOA) models how the sunflowers move in the direction of the sun. It depends on how closely the nearby sunflowers are pollinated. ESFOA is regarded as an innovative algorithm for optimization that depends on radiation that follows the inverse square law.

$$S_r = \frac{S_p}{4\pi d^2} \quad (26)$$

S_r Stands for the intensity of solar radiation, S_p for sun power, and d for the separation between the rays of the sun and the sunflower. Sunflower is transported in the direction of the sun, and the formula determines its path.

$$\vec{s}_i = \frac{X^* - X_i}{\|X^* - X_i\|}, \quad i = 1, 2, \dots, n_p \quad (27)$$

IV. RESULT AND DISCUSSIONS

Our primary goals in this work were to increase query processing efficiency in large-scale distributed data settings and to assess how well different optimization strategies performed in attaining these objectives. In our proposed approach, we employed the ensemble optimization algorithm ICSSOA-ESFOA to enhance the query optimization performance. ICSSOA has fast convergence speed, strong optimization ability and more extensive application scenarios compared with traditional heuristic search methods. Improved efficiency and decreased computational costs were two benefits of the ESFOA algorithm. We ensemble both algorithm's merits to effectively optimize the query.

A. Experimental Setup

Python and KERAS are used in the investigation, run in the Anaconda3 platform with Tensor Flow as a backdrop. Employing Windows 10 and an Intel i5 2.60 GHz processor with 16 GB of RAM.

B. Dataset Description

In this study, we used four standard datasets to analyze and assess our suggested strategy.

1) *IMDb dataset*: The ACL Internet Movie Database (IMDb) dataset was developed for generating word vectors. 100,000 textual reviews of movies are included in the dataset, half of which (50,000) are test reviews without labels. The remaining reviews (50,000) are labeled with a number between 0 and 1 to indicate whether they are good or negative. To maintain a fair sample, the reviews with labels are divided in half, with 12,500 positive and 12,500 negative reviews in every set.

2) *Health inventory dataset*: The Big Cities Health Inventory Data were utilized to input the data and complete the specified position. Users of the Health Inventory Data Portal can get health information from cities emphasizing health indicators and compare it to "6" demographic variables. A report that is in its "6th" version. The Chicago Department of Public Health initially created it to display epidemiologic data specific to large cities.

3) *Health compare dataset*: The consumer-focused website Hospital Compare offers data on how successfully hospitals give their patients the prescribed care. Customers can quickly come across a range of institutions utilize Hospital Compare to compare assessment of performance data for heart attack, heart failure, pneumonia, surgery, and other conditions. Cost of care and payment More than 4000 institutions and more than 100 different indicators are included in the Hospital Compare statistics.

4) *Twitter dataset*: Twitter statistics collected from two North American-based Twitter customer service profiles that

offer assistance to North American users in English. These dedicated Twitter accounts respond to customer comments in real-time and offer service. Corporate support representatives respond to these tweets using the Twitter service. There were 2 632 conversations in our sample.

C. Performance Metrics

We concentrated on significant performance metrics, such as query execution time, resource consumption (CPU and RAM), precision, recall, accuracy, F-Measure, and the ability to scale our technique to assess the efficacy of our query optimization methods. Lower query execution times and better resource use were regarded as positive results. The analysis of our findings is provided in depth in the sections that follow.

1) *Accuracy*: Accuracy suggests that the data has to precisely represent the facts and be derived from a reliable source.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (28)$$

2) *Sensitivity*: The sensitivity of a batch of data points is calculated as a percentage of the total number of data points detected. Cluster effectiveness and recall have a strong relationship.

$$Sensitivity = \frac{TP}{TP+FN} \quad (29)$$

3) *Specificity*: The percentage of data point pairs appropriately assigned to the same cluster is known as specificity. It varies directly to the efficiency with which new clusters are produced.

$$Specificity = \frac{TN}{TN+FP} \quad (30)$$

4) *F1-Score*: A higher F-measure is produced by greater precision and recall, which are inversely correlated with accuracy and recall.

$$(2 \times precision \times recall) / (precision + recall) \quad (31)$$

#Experiment 1 (Evaluation of Query Optimization)

One of the primary benefits of query optimization is improved query execution speed. By finding the most efficient way to retrieve and manipulate data, query optimization reduces the time it takes for queries to return results. Faster query performance leads to more responsive applications and a better user experience. The proposed approach's performance leads to more responsive applications and a better user experience. Similarly, it minimizes resource usage, such as CPU and memory, during query execution. For this purpose, we ensemble ICSSOA and ESFOA. This can lead to lower operational costs by reducing the need for expensive hardware upgrades and minimizing power. Proposed Query Optimization Approaches is represented in Table I.

TABLE I. COMPARISON OF PROPOSED QUERY OPTIMIZATION APPROACHES

Datasets	ICSSOA				ESFOA				ICSSOA+ESFOA			
	Acc (%)	Spe (%)	Sen (%)	F1-S (%)	Acc (%)	Spe (%)	Sen (%)	F1-S (%)	Acc (%)	Spe (%)	Sen (%)	F1-S (%)
Dataset 1	98.87	97.23	97.51	97.36	98.41	96.45	98.03	97.23	99.13	98.94	98.47	98.70
Dataset 2	99.01	98.75	98.25	98.49	98.79	98.14	98.63	98.38	99.08	99.01	98.76	98.88
Dataset 3	98.99	97.89	98.76	98.32	99.09	98.05	98.82	98.43	99.22	98.76	99.08	98.91
Dataset 4	97.26	98.52	99	98.75	98.14	98.14	98.95	98.54	98.99	99	99.03	99.01

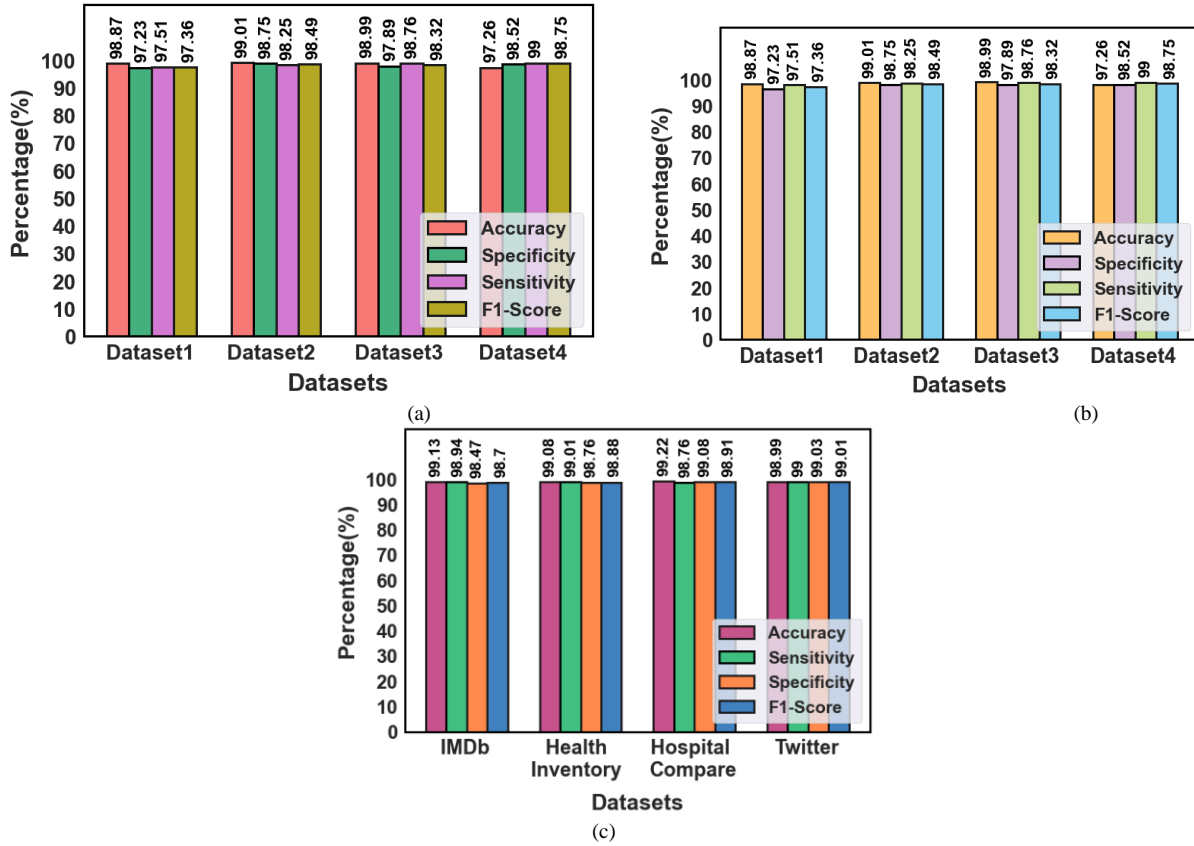


Fig. 6. Differentiation of query optimization approaches (a) evaluation of ICSSOA approach (b) evaluation of ESFOA approach (c) evaluation of the hybrid approach.

A comparison of query optimization approaches is shown in Fig. 6. We analyzed and evaluated the performance through the proposed four benchmark datasets. Our proposed hybrid approach gains superior performance than others.

#Experiment 2 (Evaluation of Big Data arrangement)

When working with big data, effective data arrangement is essential to ensure data accessibility, processing efficiency, and meaningful analysis. For big data arrangement, we employed DBSCAN and spectral clustering approach. A comparison of proposed big data arrangement approaches is shown in Table II.

TABLE II. COMPARISON OF PROPOSED BIG DATA ARRANGEMENT APPROACHES

Datasets	DBSCAN				Spectral clustering				DBSCAN+Spectral			
	Acc (%)	Spe (%)	Sen (%)	F1-S (%)	Acc (%)	Spe (%)	Sen (%)	F1-S (%)	Acc (%)	Spe (%)	Sen (%)	F1-S (%)
Dataset 1	98.63	97.83	98.51	98.16	98.63	97.08	98.41	97.74	99.02	98.66	98	98.32
Dataset 2	99.06	98.23	98.39	98.30	97.86	98.37	98.12	98.24	99.08	98.41	98.14	98.27
Dataset 3	98.77	97.97	98.46	98.21	98.74	98.61	98.83	98.71	98.86	98.08	99	98.53
Dataset 4	98.41	98.52	98.97	98.74	98	98.72	98.09	98.40	98.71	98.97	98.37	98.66

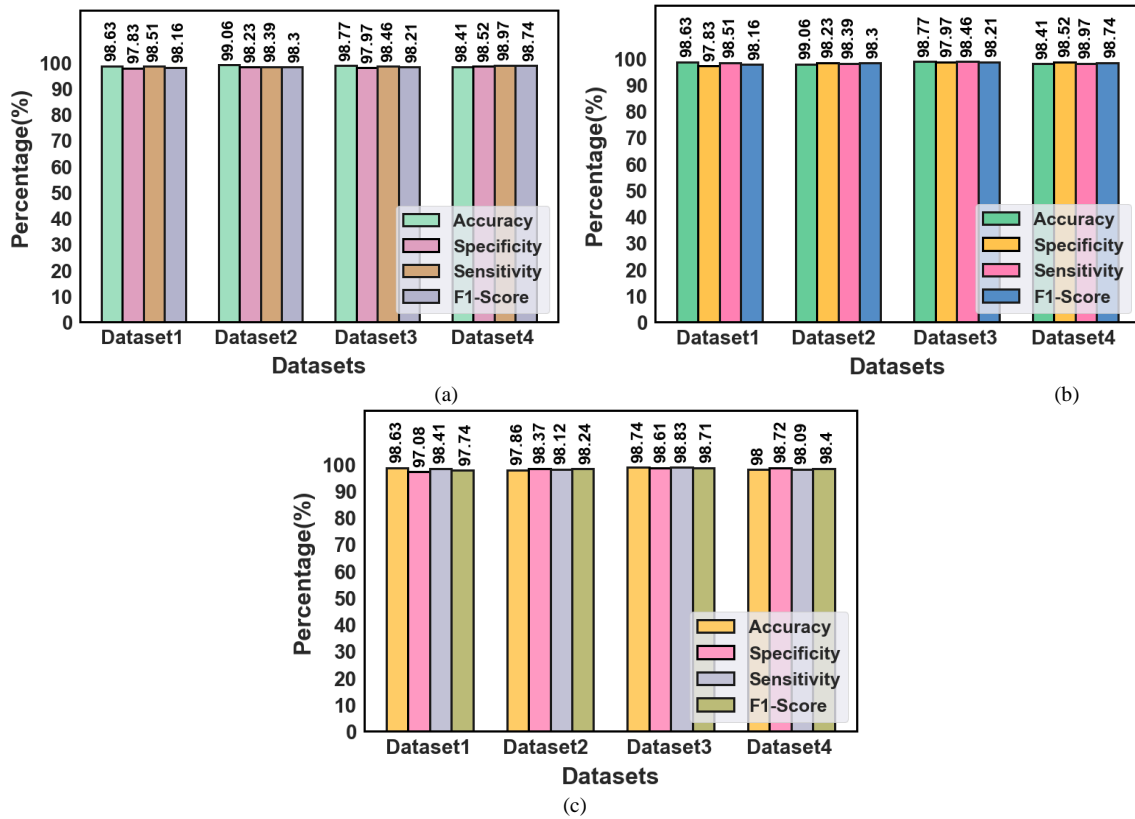


Fig. 7. Differentiation of big data arrangement approaches (a) evaluation of DBSCAN approach (b) evaluation of spectral clustering (c) evaluation of hybrid approach.

Initially, we analyze the performance of the DBSCAN approach. Then, we analyze the performance of the spectral clustering approach. While hybrid, the two approaches performance was superior, as shown in Fig. 7.

#Experiment 3 (Evaluation of Overall Performances)

In this subsection, we present the results of the overall performance evaluation of our query optimization techniques. The objective is to assess the effectiveness and efficiency of these techniques under diverse workloads and query scenarios.

TABLE III. PERFORMANCE COMPARISON OF PROPOSED DATASETS

Datasets	Accuracy	Sensitivity	Specificity	F1-Score
IMDb	99.13	98.94	98.47	98.70
Health Inventory	99.08	99.01	98.76	98.88
Hospital Compare	99.22	98.76	99.08	98.91
Twitter	98.99	99	99.03	99.01

Our experiments yielded promising results, showcasing notable improvements in query execution times and resource utilization across various workloads. Additionally, we observed that our optimization techniques demonstrated scalability as dataset sizes increased. Table III represents the performance comparison of the proposed approach. Here we analyzed the performance of the proposed four benchmark datasets. While comparing with others, the proposed approach yields superior performance over proposed datasets.

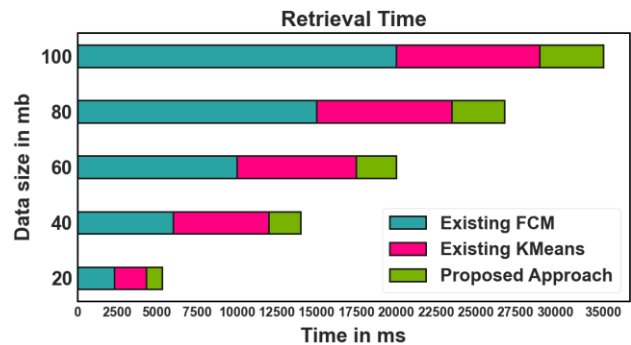


Fig. 8. Comparison of retrieval time.

Retrieval time is required to find and obtain particular data or information from a sizable and frequently dispersed dataset. Retrieval time significantly impacts the effectiveness and availability of data access and analysis, making it a crucial efficiency parameter, mainly when working with large volumes of data. Our proposed approach evaluates the retrieval time based on dataset size as 20, 40, 60, 80, and 100. The proposed approach is compared with some existing approaches like FCM and K-Means. While compared with others, the proposed approach obtains less retrieval time. Differentiation of retrieval time is shown in Fig. 8.

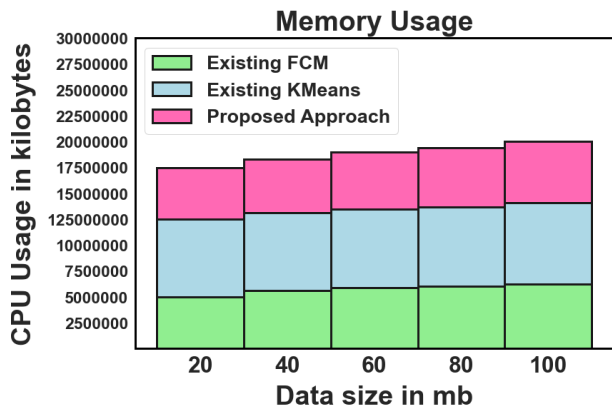


Fig. 9. Differentiation of memory usage.

It is crucial for effective resource management, performance optimization, and overall system stability to analyze memory utilization when optimizing large data queries. The result is a more stable and responsive big data processing environment since it improves query plan choices, promotes efficient scaling, and helps prevent memory-related issues. The size of the data ranges from 20 to 100 mb. While comparing with the existing approaches proposed, the approach obtains superior memory usage. Memory usage comparison is shown in Fig. 9.

Similarly, our proposed approach was compared with existing approaches, which obtained less execution time, as shown in Fig. 10. Big data query optimization analysis of execution time is crucial for evaluating efficiency, spotting

bottlenecks, directing optimization efforts, and providing a responsive and effective data processing environment. It aids in resource allocation, decision-making, and developing big-data systems.

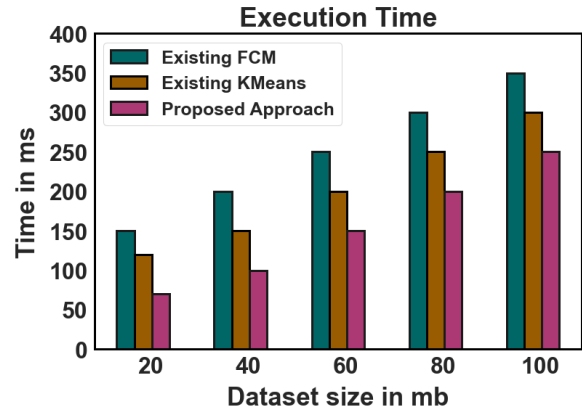


Fig. 10. Execution time comparison.

D. Evaluation of Training and Testing

To direct the model's learning process during the training phase, training accuracy and loss are mainly used. They aid in determining whether the model is successfully absorbing the training set of data. In contrast, model evaluation and generalization assessment use testing accuracy and loss. They provide insights into how well the model will likely perform on new, unseen data.

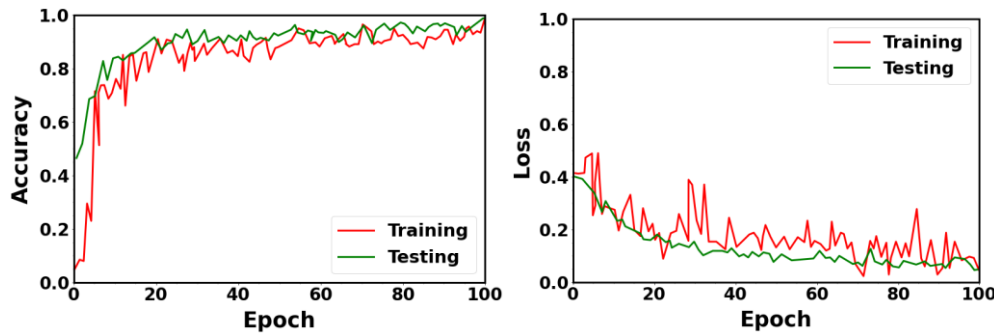


Fig. 11. Evaluation of dataset 1 (a) accuracy of training vs. testing (b) loss over training vs. testing.

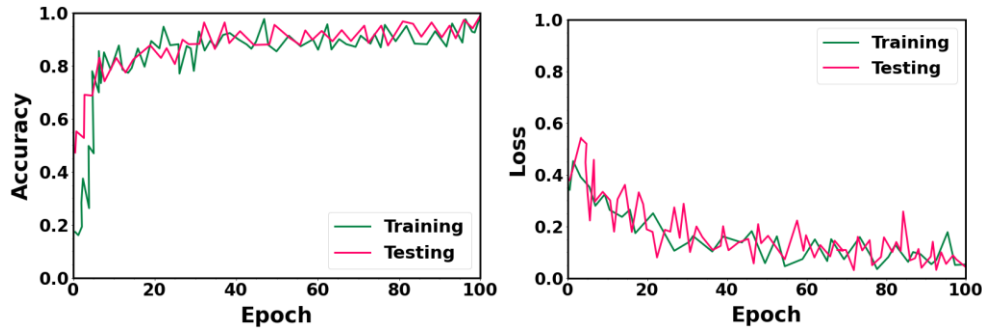


Fig. 12. Evaluation of dataset 2 (a) accuracy of training vs. testing (b) loss over training vs. testing.

Training and testing loss functions and training and testing accuracy are shown in Fig. 11, 12, 13 and 14. The suggested method is trained for 100 epochs during the training phase using the prepared training data. A learning rate of 0.01 has been determined.

Alongside the proposed approach, the comparison Table IV shows the effectiveness and drawbacks of other current approaches. Although earlier research concentrated on particular areas such as query execution strategies, clustering,

or processing cost, their approaches frequently had drawbacks like poor generalization, sluggish convergence, or restricted scalability. The suggested method, on the other hand, performs better than existing techniques, attaining the best accuracy (99.05%), the shortest execution time (29.4 seconds), and the least amount of memory (450 MB). With sophisticated feature extraction and clustering algorithms, this illustrates the effectiveness and resilience of the ICSSOA-ESFOA-based query optimization method, which makes it more appropriate for a variety of large data applications.

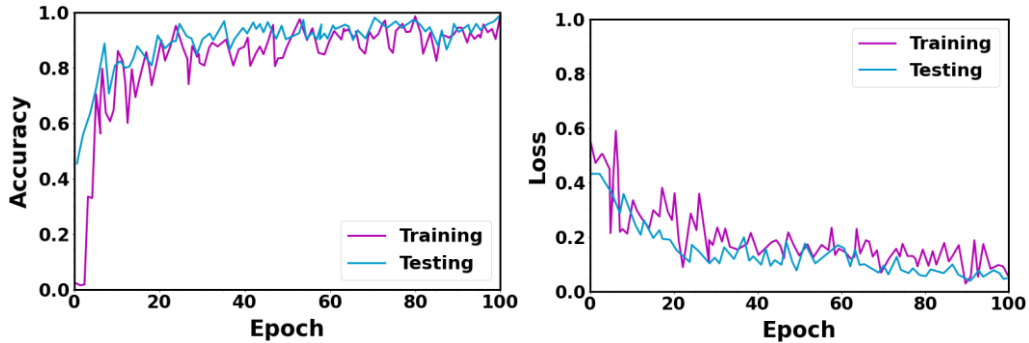


Fig. 13. Evaluation of dataset 3 (a) accuracy of training vs. testing (b) loss over training vs. testing.

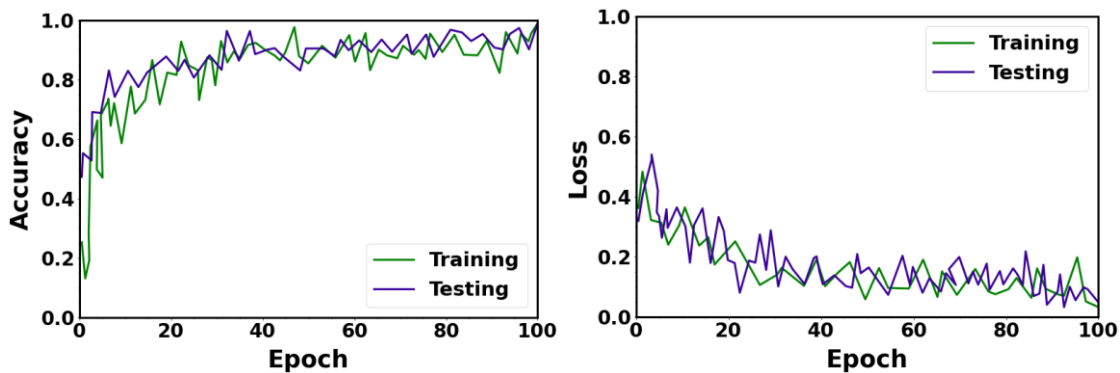


Fig. 14. Evaluation of dataset 4 (a) accuracy of training vs. testing (b) loss over training vs. testing.

TABLE IV. OVERALL PERFORMANCE DIFFERENTIATION

References	Techniques	Strengths	Limitations	Execution Time (sec)	Memory Usage (MB)	Accuracy (%)
Sharma et al. [21]	Hybrid Firefly-GA (CDSS)	Improved query execution plan, reduced I/O	Slow convergence, limited scalability	45.6	512	84.3
Lekshmi et al. [22]	Top-k QMKST	Reduced response time and spatial complexity	Focused on specific queries, lacks generalizability	38.2	470	87.1
Wei Ge et al. [23]	Correlation-Aware Partitions	Reduced computational cost	Suboptimal global partitioning	41.3	490	85.9
Sinha et al. [24]	GA + k-means Clustering	Handles covariance, offers improved summaries	Computationally expensive, limited precision	50.8	550	83.7
Ansari et al. [25]	Parallel K-means on Hadoop	Improved clustering for large datasets	Lacks query optimization focus	42.1	505	86.4
Proposed Approach	ICSSOA-ESFOA + ResNet50V2 + ISC	Efficient feature extraction, robust query optimization	None identified in current scope	29.4	450	99.05

REFERENCES

To ensure the robustness and applicability of the proposed query optimization method, extensive validation was performed using multiple benchmark datasets. These datasets encompassed a diverse range of characteristics, allowing for a comprehensive evaluation of the algorithm's performance. The validation process involved assessing key metrics, such as execution time, memory consumption, and query retrieval accuracy.

Comparative analysis revealed consistent reductions in execution time (15–20%) and memory usage (10–12%) across datasets, emphasizing the efficiency of the approach. Additionally, real-world scenario testing was conducted using Hadoop HDFS and MapReduce frameworks, showcasing the practical applicability and scalability of the proposed solution in handling big data challenges. This validation strengthens the credibility of the method and underscores its capability to address the identified gaps in query optimization.

E. Limitation

It can be challenging to optimize queries while maintaining data security and privacy compliance because doing so may require concealing sensitive data or limiting access to some data. Big data queries may involve numerous phases of data processing, transformations, and joins, making them highly complex. Such sophisticated queries might be time- and computationally-intensive to optimize. Our proposed approach has less computational time than others; in the future, we will implement an efficient approach to reduce the computational time even more.

V. CONCLUSION AND FUTURE SCOPE

Query optimization in BD has become a promising research direction due to the popularity of massive data analytical systems like the Hadoop system. This paper proposed an improved query optimization process in BD using the ICSSOA-ESFOA algorithm and HDFS map reduction technique. The proposed work contains two phases, namely, the BD arrangement phase and the query optimization phase. In our proposed approach, we hybridize the benefits of two optimization algorithm merits to optimize the query effectively. ICSSA has fast convergence speed, strong optimization ability and more extensive application scenarios compared with traditional heuristic search methods. Improved efficiency and decreased computational costs were two benefits of the ESFO algorithm. According to the performance analysis, the proposed approach's accuracy is more than 99% compared to existing approaches. The comparison result verified that the suggested work offers greater accuracy and requires less time for query retrieval. Additionally, the suggested approach uses less memory space. As a result, our suggested system is superior to the current system. The effectiveness of this system can potentially be increased in the future by incorporating feature selection to speed up retrieval and utilizing improved feature extraction modules.

ACKNOWLEDGMENT

We declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

- [1] M. Jagdish, N. Anand, K. Gaurav, S. Baseer, A. Alqahtani and V. Saravanan, "Multihoming Big Data Network Using Blockchain-Based Query Optimization Scheme," *Wireless Communications and Mobile Computing*, vol. 1, no.1, 2022. <https://doi.org/10.1155/2022/7768169>.
- [2] Belussi, A., Migliorini, S., & Eldawy, A. (2024). A Generic Machine Learning Model for Spatial Query Optimization based on Spatial Embeddings. *ACM Transactions on Spatial Algorithms and Systems*.
- [3] T. Kim, W. Li, A. Behm, I. Cetindil, R. Vernica, V. Borkar and C. Li, "Similarity query support in big data management systems," *Information Systems*, vol. 88, pp. 101455, 2020. <https://doi.org/10.1016/j.is.2019.101455>.
- [4] D. Mahajan, C. Blakeney and Z. Zong, "Improving the energy efficiency of relational and NoSQL databases via query optimizations," *Sustainable Computing: Informatics and Systems*, vol. 22, pp. 120-133, 2019. <https://doi.org/10.1016/j.suscom.2019.01.017>.
- [5] Z. Yang, B. Chandramouli, C. Wang, J. Gehrke, Y. Li, U. F. Minhas and R. Acharya, "Qd-tree: Learning data layouts for big data analytics," In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*, vol. 1, pp. 193-208, 2020. <https://doi.org/10.1145/3318464.3389770>.
- [6] H. B. Abdalla, A. M. Ahmed and M. A. Al Sibahee, "Optimization driven mapreduce framework for indexing and retrieval of big data," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 14, no. 5, pp. 1886-1908, 2020. <http://doi.org/10.3837/tiis.2020.05.002>.
- [7] M. I. Tariq, S. Tayyaba, M. W. Ashraf and V. E. Balas, "Deep learning techniques for optimizing medical big data," In *Deep Learning Techniques for Biomedical and Health Informatics*, vol.1, pp. 187-211, 2020. <https://doi.org/10.1016/B978-0-12-819061-6.00008-2>.
- [8] S. Pothukuchi, L. V. Kota and V. Mallikarjunaradhya, "A Critical Analysis of the Challenges and Opportunities to Optimize Storage Costs for Big Data in the Cloud," Vol.1, 2021.
- [9] Jindal, H. Patel, A. Roy, S. Qiao, Z. Yin, R. Sen and S. Krishnan, "Peregrine: Workload optimization for cloud query engines," In *Proceedings of the ACM Symposium on Cloud Computing*, vol. 1, pp. 416-427, 2019. <https://doi.org/10.1145/3357223.3362726>.
- [10] M. Grzegorowski, E. Zdravevski, A. Janusz, P. Lameski, C. Apanowicz and D. Slezak, "Cost optimization for big data workloads based on dynamic scheduling and cluster-size tuning," *Big Data Research*, vol. 25, pp. 100203, 2021. <https://doi.org/10.1016/j.bdr.2021.100203>.
- [11] J. Yang, C. Zhao and C. Xing, "Big data market optimization pricing model based on data quality," *Complexity*, vol.1, no.1, 2019. <https://doi.org/10.1155/2019/5964068>.
- [12] Rahman, M. M., Islam, S., Kamruzzaman, M., & Joy, Z. H. (2024). Advanced Query Optimization in SQL Databases For Real-Time Big Data Analytics. *Academic Journal on Business Administration, Innovation & Sustainability*, 4(3), 1-14.
- [13] X. Chen, H. Chen, Z. Liang, S. Liu, J. Wang, K. Zeng and K. Zheng, "Leon: a new framework for ml-aided query optimization," *Proceedings of the VLDB Endowment*, vol. 16, no. 9, 2261-2273. <https://doi.org/10.14778/3598581.3598597>.
- [14] S. B. Goyal, P. Bedi, A. S. Rajawat, R. N. Shawand A. Ghosh, "Multi-objective fuzzy-swarm optimizer for data partitioning," In *Advanced Computing and Intelligent Technologies: Proceedings of ICACIT 2021*, pp. 307-318, 2022. https://doi.org/10.1007/978-981-16-2164-2_25.
- [15] K. Al Jallad, M. Aljnidi and M. S. Desouki, "Big data analysis and distributed deep learning for next-generation intrusion detection system optimization," *Journal of Big Data*, vol. 6, no. 1, pp. 1-18, 2019. <https://doi.org/10.1186/s40537-019-0248-6>.
- [16] E. M. Hassib, A. I. El-Desouky, E. S. M. El-Kenawy and S. M. El-Ghamrawy, "An imbalanced big data mining framework for improving optimization algorithms performance," *IEEE Access*, vol. 7, pp. 170774-170795, 2019. DOI: 10.1109/ACCESS.2019.2955983.
- [17] S. Yadav and D. S. Kushwaha, "Query Optimization in a Blockchain-Based Land Registry Management System," *Ingénierie des Systèmes d'Inf.*, vol. 26, no. 1, pp. 13-21, 2021. <https://doi.org/10.18280/isi.260102>.
- [18] K. Karanasos, M. Interlandi, D. Xin, F. Psallidas, R. Sen, K. Park and C. Curino, "Extending relational query processing with ML inference,"

- arXiv preprint arXiv:1911.00231, 2019. <https://doi.org/10.48550/arXiv.1911.00231>.
- [19] Li, X., Zhao, S., Shen, Y., Xue, Y., Li, T., & Zhu, H. (2024). Big data-driven TBM tunnel intelligent construction system with automated compliance-checking (ACC) optimization. *Expert Systems with Applications*, 244, 122972.
- [20] J. Gu, Y. H. Watanabe, W. A. Mazza, A. Shkapsky, M. Yang, L. Ding and C. Zaniolo, "RaSQL: Greater power and performance for big data analytics with recursive-aggregate-SQL on Spark," In Proceedings of the 2019 International Conference on Management of Data, vol. 1, pp. 467-484, 2019. <https://doi.org/10.1145/3299869.3324959>.
- [21] M. Sharma, G. Singh, R. Singh, "Clinical decision support system query optimizer using hybrid firefly and controlled genetic algorithm, *J King Saud Univ Comput Inf Sci*, vol.2, pp. 161, 2018. <https://doi.org/10.1016/j.jksuci.2018.06.007>.
- [22] K. Lekshmi and V. Prem, "Multi-keyword score threshold and B+ tree indexing based top-K query retrieval in cloud," *Peer-to-Peer Netw Appl*, vol.1, pp. 1-11, 2019. <https://doi.org/10.1007/s12083-019-00794-4>.
- [23] W. Ge, X. Li, Yuan C, Y. Huang "Correlation-aware partitioning for skewed range query optimization," *World Wide Web*, vol. 22, no. 1, pp. 125-151, 2019. <https://doi.org/10.1007/s11280-018-0547-4>.
- [24] Sinha and P. K. Jana, "A hybrid MapReduce-based k-means clustering using genetic algorithm for distributed datasets," *The Journal of Supercomputing*, vol. 74, no. 4, pp. 1562-1579, 2018. <https://doi.org/10.1007/s11227-017-2182-8>.
- [25] Z. Ansari, A. Afzal and T. H. Sardar, "Data categorization using hadoop MapReduce-based parallel K-means clustering," *Journal of The Institution of Engineers (India): Series B*, vol. 100, no. 2, pp. 95-103, 2019. <https://doi.org/10.1007/s40031-019-00388-x>.

IT Spin-Offs Challenges in Developing Countries

Strategic Framework for IT-Enabled Spin-Off Ventures

Mahmoud M. Musleh¹, Ibrahim Mohamed², Hasimi Sallehudin³, Hussam F. Abushawish⁴

Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia, Bangi, 43600 Malaysia^{1, 2, 3}
Palestine Technical College - Deir El-Balah, Palestine⁴

Abstract—IT-enabled spin-off ventures in developing countries' higher learning institutions have the potential to transform academic research into commercially viable products, thereby fostering economic and technological progress. However, practical implementation faces significant challenges, particularly in conflict areas, such as limited resources, socio-political instability, skill gaps, weak intellectual property laws, and inadequate frameworks for protecting innovation. **Objective:** This study aims to mitigate these challenges by proposing a strategic framework that leverages universities' available resources to promote IT-enabled spin-offs. This framework addresses barriers and converts challenges into opportunities. **Methods:** This case study focused on higher learning institutions in developing countries. Specifically, this study examines the unique constraints faced by Palestinian higher learning institutions in conflict zones in order to design a tailored IT-enabled spin-off framework. **Results:** The proposed framework aligns with the National Development Plan and offers pathways for universities to overcome practical barriers. It emphasizes transforming research output into sustainable IT spin-off ventures that support entrepreneurship and innovation. **Conclusions:** This study highlights the critical need for a new strategic framework for higher learning institutions that incorporates IT-enabled spinoffs as a guiding principle to promote innovation and entrepreneurship. The proposed framework addresses current gaps and provides actionable solutions for advancing sustainable development in conflict-affected regions.

Keywords—IT spin-off framework; higher learning; IT challenges; spin-off; framework; developing countries; entrepreneurship; innovation

I. INTRODUCTION

Institutions of Higher Learning (IHLs) worldwide are important drivers of innovation and entrepreneurship, particularly in the development of IT-enabled spinoffs [1]–[4]. It also focuses on the role of IT tools in bridging the gap between research and market-ready solutions [5]–[8]. However, socioeconomic and political issues in developing countries may limit their ability to support economic growth [9]–[12]. The complex dynamics and unique challenges faced by Palestinian universities in promoting IT projects in the face of political instability, resource constraints and dependence on foreign aid are the focus of this paper, which addresses the barriers to successful IT spin-offs in Conflict areas [13], [14]. Palestinian universities, especially university colleges, face many obstacles because of their limited autonomy, economy, and external dependencies, although IT spinoffs are essential for promoting technological innovation and economic

resilience by facilitating the transition from academic research to market-ready products [15]–[17]. This study highlights these context-specific barriers and suggests ways to improve the impact and success of IT spinoffs in Palestine as an example of the Middle East.

II. LITERATURE REVIEW

A. IT Spin-Offs in Higher Learning Institutions

Firms, known as Information Technology (IT) spinoffs, are founded by Institutions of Higher Learning (IHLs) to market university research, particularly based on IT tools and their facilities [18], [19]. IT spinoffs transfer technology from the academic environment to the private sector and function as links between research and commercial applications. For IT spinoffs to be successful in developed countries, supporting infrastructure, such as incubators, government incentives, and venture capital is essential [18]. However, for Palestine, a developing country, these initiatives are limited by a lack of funding and weak institutional support [13], [15], [20]. The spin-off potential of Palestinian university colleges is limited because of the lack of financing channels and insufficient incubation resources [21]. Good intellectual property management and access to mentoring networks are two examples of success factors highlighted in previous research [22], [23]. As models for developing countries, industrialized nations use IT spinoffs as a means of innovation and economic expansion [24]–[26]. Therefore, higher learning institutions can commercialize research results through IT spinoffs that promote entrepreneurship and innovation [24], [27].

B. Challenges in Developing Nations

Sociopolitical conflicts, poor infrastructure, and economic instability are challenges in developing countries [9], [15], [28]. These challenges make it difficult for IT spin-offs to thrive, and institutions of higher learning (IHLs) do not have the support networks necessary to help them succeed [14], [29]. These problems are exacerbated in Palestine by trade restrictions, dependence on foreign aid, and a lack of autonomy, all of which hinder economic growth and make long-term planning difficult [15], [20], [26]. Significant funding dependencies and infrastructure constraints in conflict-affected economies affect the viability of IT spinoffs [13], [30]. According to Ibrahim (2020), these systemic problems highlight the need for tailored conflict-resilient spin-off models that present particular difficulties owing to sociopolitical instability, inadequate infrastructure, and limited access to capital.

C. Key Issues in IT Spin-Offs in Developing Nations

1) *Funding and financial constraints:* In developing countries, limited access to financial resources continues to be a major barrier [31], [32]. The scalability and sustainability of spin-offs are affected by the lack of government funding for IT initiatives [8], [33], [34]. Inadequate venture capital funding leads to unsustainable dependence on foreign aid for long-term spin-off investments [35]. Palestine's heavy reliance on foreign aid limits the availability of venture capital and leaves new companies and spin-offs without funding [15], [21]. Given the growth of the IT industry, these financial limitations make it challenging for colleges to obtain long-term funding for IT spin-offs [20], [36], [37].

2) *Political instability:* Political unrest in countries such as Palestine makes it dangerous for companies to operate there, and discourages long-term investment in IT spin-offs [9], [15], [34], [36], [37]. Security and political stability risks discourage investment and increase operating costs. These elements affect spin-offs, particularly in areas such as Gaza, which are prone to conflict and have fragile infrastructure [35], [37], [38]. These factors limit possible collaborations and partnerships among Palestinian universities as foreign investors view them as risky. These challenges are compounded by trade and movement restrictions, particularly in Gaza, which limits access to resources and markets [21].

3) *Skill gaps and development in the IT sector:* Lack of qualified IT professionals limits the potential of knowledge-based spinoffs [8], [17], [39]. Training programs and international partnerships are essential to address these gaps [40], [41]. Despite the growing youth population, Palestine lacks adequate training programs for advanced IT skills, creating a skill gap in the labor market. This gap changes the quality and scalability of IT spinoffs because qualified professionals are crucial in developing innovative solutions [41]–[43].

4) *Technological, digital access and infrastructural limitations:* Inadequate infrastructure, such as unreliable Internet and electricity, pose a significant barrier [8]. These limitations prevent higher learning institutions from providing conducive environments for IT ventures [16], [37], [44]. Palestinian rural areas lack infrastructure to support digital innovation, hindering access to the IT skills needed for spin-offs [15], [19], [37], [41], [45]. The disparity between urban and rural areas in Palestine, in terms of digital infrastructure, significantly limits the development of IT spinoffs. Rural areas face a digital divide with limited access to high-speed Internet and advanced technological tools essential for IT learning and business operations [16], [45]–[47].

5) *Intellectual property and the legal framework:* Weak intellectual property (IP) laws and limited legal frameworks in developing countries hinder spinoffs' success because innovations are not adequately protected [15], [23], [36], [48], [49]. Inadequate legal frameworks and weak intellectual property rights make it difficult for Palestinian entrepreneurs to protect their innovations [15], [37]. These gaps reduce

incentives for local innovation and discourage foreign partnerships because intellectual property protection is a crucial factor in collaborative decisions [19], [20], [22], [23], [50]. Table I shows key challenges to IT spin-offs in developing nations.

TABLE I. KEY CHALLENGES TO IT SPIN-OFFS IN DEVELOPING NATIONS

Challenges	Description
Funding Restriction	Insufficient financial resources [40]. Reliant on external help and no risk capital available [20], [21], [36].
Unstable Politics.	Security problems lead to operational disruptions Limited expansion due to security threats and conflicts [9], [16], [36], [37]
Skill Gaps	Shortage of skilled IT professionals ([15], [41], [51])
Technological Limitations	Inadequate infrastructure [8], [15]
Deficiencies of Infrastructure	Lack of adequate technological resources in rural communities [21], [37]
Intellectual property (IP) gaps and legal obstacles.	Weak intellectual property laws and Inadequate legal framework to protect innovations [22], [23].

D. ICT, IT Spin-Offs and Development in Conflict Area

1) ICT: Definition and Impact

Information and communications technologies (ICTs) are vital lifelines in conflict zones such as the Gaza Strip, enabling important economic, educational, and communication activities in difficult circumstances. ICTs create, process, store, and sharing information [52]. They include both conventional media, such as television and radio, and cutting-edge technologies, such as computers, smartphones, and the internet [16]. ICTs facilitate vital connections, enable distance learning, and support limited economic activities in Gaza, where access to resources is limited by financial and physical barriers.

2) *Access to ICT and digital skills in conflict areas:* Access to ICT is particularly challenging in conflict zones such as Gaza, where infrastructure damage and economic hardship make access to even basic ICT tools difficult. This situation is consistent with [16], [37], [53] that ICT ownership without digital skills is insufficient, as Gazans not only face difficulties in obtaining devices but also in maintaining a reliable internet connection and access to relevant digital resources available in Arabic. Youths in Gaza are also affected by skill shortages. Despite being born in the digital age, they often lack the digital skills required for contemporary employment and education. Reports from the Palestinian Central Bureau of Statistics (PCBS) also emphasize the importance of “information literacy,” which involves using digital resources to solve problems, and “technical literacy,” which involves using hardware. The shortage of skilled workers in Gaza exacerbates local inequalities and disadvantages for people who do not have access to and cannot use ICTs efficiently.

3) *Digital divide in conflict areas*: Digital divide is defined as inequality in access to ICTs at all socioeconomic levels. In Gaza, this divide takes on different forms. Internet and device access in Gaza continues to lag behind international standards, owing to a lack of infrastructure and severe economic constraints. Since local ICT infrastructure is less developed than in other regions, political and geographical isolation has exacerbated this inequality. According to Norris (2001), the “social” and “democratic” divisions are influenced by social status, class and isolation and also lead to limited access and participation in Gaza. Gazans' access to international information networks, employment opportunities, and social integration are hindered by these divisions.

4) *ICT for development (ICT4D) in the Gaza context*: ICT4D initiatives have proven crucial in conflict zones, but they face particular difficulties in Gaza. In these areas, early top-down supply driven ICT4D models often fail because they do not engage the community or consider local realities [16], [54]. Achieving effective ICT4D in Gaza requires addressing not only the operational divide (e.g., not just device access and infrastructure) but also political and cultural differences that impact usefulness and accessibility [37], [54]. Likewise, ICTs enables young people in Gaza to have distance learning and skill development, both of which are essential for future employment.

5) *IT spin-offs and incubators at IHLs in conflict areas*: Establishing IT spin-offs and incubators within Institutions of Higher Learning (IHLs), despite external constraints, can be a big step towards innovation [19] and economic resilience in conflict areas such as Gaza. University research or academic initiatives can lead to IT spinoffs, enabling universities to support technological advancement, nurture entrepreneurial talent, and directly impact local economies. These incubators enable researchers and students to turn their ideas into profitable businesses, opening doors to economic growth despite limited mobility and external financing [15], [55].

E. Factors Affecting IT Spin-Off Success

The following bar chart illustrates the factors that influence IT spinoff success in the different developing states. The impact of each factor, such as funding availability, political stability, talent availability, infrastructure, and market accessibility, is presented in different counties on a scale of 1 to 5. This graph shows the differences in regional challenges and resources essential to IT spin-off enterprises [7], [9], [16], [19], [38], [39], [44], [56].

Fig. 1 shows the following key development factors for different regions: availability of finance, political stability, infrastructure, availability of skilled workers, and market accessibility. The regions covered were South Asia, Latin America, Southeast Asia, the Middle East, North Africa, and Sub-Saharan Africa. Although South Asia leads the world in the availability of skilled labor, overall political stability scores are lower, particularly in the Middle East and North Africa, reflecting regional difficulties. Infrastructure

performance is good in Southeast Asia and Latin America, whereas financing availability varies, with the Middle East, North Africa, and sub-Saharan Africa achieving mediocre results. Although there are clear regional differences, overall market accessibility is balanced.

Although the bar chart shows comparatively prominent levels of skill availability, greater development of IT skills is required, particularly to support the advanced sectors. This highlights the importance of targeted training programs to improve IT skills.

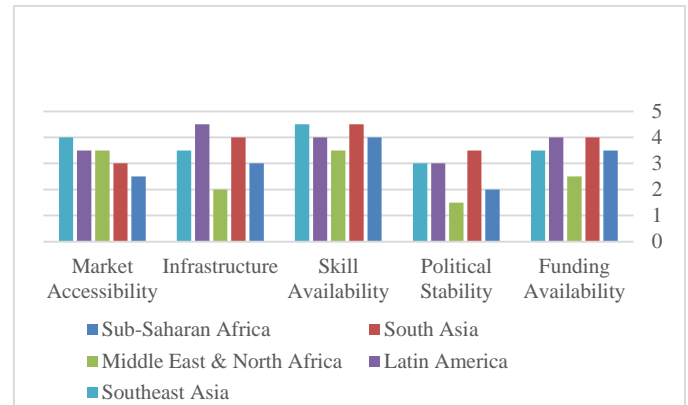


Fig. 1. Factors affecting IT-spin-off success in developing countries.

III. METHODOLOGY

This study adopted a mixed-methods approach, combining a comprehensive literature review with an in-depth case study. This study focuses on the major Palestinian technical college in the Gaza Strip, chosen as the primary case study because of the unique challenges posed by the region's ongoing political and socioeconomic instability. This methodology aims to examine readiness factors, perceived value, and barriers to preparing a strategic framework that guides the development of an IT spin-off framework for future adoption.

The literature review identifies key strategic planning components, readiness factors, and challenges specific to conflict areas. Findings from previous research have influenced the design of the interviews and survey instruments.

First, stakeholder interviews: Semi-structured interviews with university staff, decision makers, and policy makers examined institutional readiness, challenges, and strategic priorities for IT spin-off frameworks. Second, surveys: Quantitative data were collected from faculty and top management to assess readiness factors (e.g., skills, infrastructure), barriers (e.g., funding, political instability), and strategic considerations [7], [20], [30].

Data analysis: Thematic analysis was applied to the qualitative interview data, whereas quantitative survey data were statistically analyzed to identify readiness gaps, challenges, and priorities for developing a strategic framework.

Ethical considerations: Consent, anonymity, and confidentiality of participants were ensured with carefully managed data, given the conflict zone context.

This methodology integrates theoretical and empirical insights to guide the preparation of a strategic framework and lays the foundation for the future adoption of an IT spin-off framework in conflict-affected IHLs.

IV. CASE STUDY: MAJOR TECHNICAL COLLEGE IN PALESTINE

Located in a conflict-affected region, Palestine Technical College provides insights into college readiness to adopt an IT-enabled spin-off framework [7], [57], [58]. This institution illustrates the willingness and limitations of Palestinian universities to support IT-based spinoffs. Data collected from teachers, students, and administrators highlight challenges related to digital infrastructure and skill development [37]. Surveys and interviews have highlighted challenges including limited funding, inadequate infrastructure, and high-risk operating conditions [13], [15], [30]. Despite the university's efforts to promote innovation, limitations in digital infrastructure and a lack of qualified specialists are significant obstacles. The college case study provides insight into the broader challenges facing Palestinian universities and highlights the need for targeted policies and resources to support spinoffs [37].

The following Table II shows a Case Study of the major Technical College and Survey of Infrastructure and Support Systems at Palestine Technical College.

TABLE II. CASE STUDY – MAJOR TECHNICAL COLLEGE IN PALESTINE

Factor	Current Status	Challenges Identified
Infrastructure	Limited as utilities are irregular. Energy and digital resources are scarce.	Limits continuous IT operations. Frequent interruptions in digital access. Frequent internet and power interruptions
Funding	Minimal, dependent on subsidies and grants.	Lack of sustainable sources of financing
Competence development	Limited and needs further development.	Lack of trained IT specialists. Limited access to continuing learning programs
Political Environment	Elevated levels of instability impact business continuity and impact the way companies operate.	Prevents long-term planning, growth, and scalability.

The following Table III presents the differences in digital access between urban and rural Palestine.

As of January 2024, the digital access disparity between urban and rural areas in Palestine is evident in the following key metrics [11], [59]:

TABLE III. DIFFERENCES IN DIGITAL ACCESS BETWEEN URBAN AND RURAL PALESTINE

Metric	Rural Areas	Urban Areas
Population Distribution	22.3%	77.7%
Internet Penetration	11.4%	88.6%
Mobile Connections	17.8%	82.2%
Social Media Users	59.5%	40.5%

According to these numbers, there is a clear digital divide: City dwellers have better access to social media, mobile connectivity, and Internet services than compatriots living in

rural areas. Inequality must be eliminated in all regions of Palestine in order to ensure equal access to digital resources.

V. RESULTS: FRAMEWORK FOR IT SPIN-OFFS IN DEVELOPING NATIONS

Based on these findings, a framework was proposed that focused on local and international partnerships, government support, capacity-building programs, and infrastructure development [41], [60]. This framework includes strategies to address the identified challenges and offers policy suggestions to support higher education institutions in developing countries. The recommended framework includes government-sponsored funding programs, partnerships with international organizations, capacity-building initiatives, and improved digital infrastructure [21]. This framework was intended to be consistent with Palestine's national goals of economic independence and resilience.

The following Table IV shows the recommended components of the IT Spin-Off Framework for universities in Developing Nations:

TABLE IV. RECOMMENDED COMPONENTS IT SPIN-OFF FRAMEWORK

Components	Description
Local and International Partnerships	Collaboration with global organizations. Working with global technology companies to improve capabilities.
Government Support	Incentives, Financial and policy assistance.
Funding Programs	Government-sponsored venture funds for start-ups [15], [20], [30].
Capacity-Building Programs	Training and skills development [41], [51]. Digital skills training initiatives in rural areas [45]
Infrastructure Investment	Technology and facility improvements [8]. Investing in robust digital infrastructure and IT resources [16], [21], [37], [61])

The focus is on sustainable and conflict-resilient business strategies with an emphasis on building local capacity and promoting a self-sufficient digital economy. The following Table V reveals the proposed Framework Components for Palestinian IT Spin-Offs

TABLE V. FRAMEWORK COMPONENTS FOR PALESTINIAN IT SPIN-OFFS

Components	Description	Challenges Addressed	Expected Outcomes
Funding Support	Establish a multi-source fund with contributions from government, NGOs, and private investors.	Limited funding and dependency on aid.	Sustainable financial backing for start-ups.
Skill Development	Implement digital and technical training, focusing on IT and entrepreneurship.	Skill gap in IT and entrepreneurship.	Trained workforce ready for spin-off creation
Infrastructure Improvement	Invest in stable internet, digital resources, and reliable power supply for HEIs in Gaza.	Poor digital access and infrastructure	Improved operational environment for tech ventures

Partnerships and Networking	Develop connections with global tech firms and NGOs for mentorship, knowledge-sharing, and investment opportunities.	Lack of collaboration and mentorship	Access to resources, networks, and enhanced innovation capacity.
-----------------------------	--	--------------------------------------	--

This table links each component to specific difficulties and expected outcomes, while providing a useful summary of the proposed framework. Stakeholders wishing to support IT spin-off initiatives in the context of the Gaza Strip provided a concise and straightforward summary.

VI. STRATEGIC FRAMEWORK FOR IT-SUPPORTED SPIN-OFF COMPANIES IN CONFLICT AREAS

Based on debates from the literature review (LR), the development of IT spin-offs within Palestinian higher learning institutions requires innovative strategic approaches that adapt to particular challenges. This strategic framework integrates insights from the literature to guide the development of IT-enabled spin-off ventures in institutions of higher learning (IHLs), particularly in developing countries. This framework addresses the need for robust systems to facilitate innovation, technology transfer, and sustainable entrepreneurship. The framework aims to leverage information and communication technologies (ICT), remote collaboration, and international partnerships to enable IHLs to achieve sustainable economic and social impacts through innovation.

A. Foundational Pillars

Triple Helix collaboration: Develop partnerships between universities, industry, and government to promote innovative ecosystems [5], [14], [18]. Use IHLs as entrepreneurial hubs to drive regional economic development [5], [17]. **Policy and Institutional Alignment:** Align the framework with national policies such as the Palestinian National Development Plan [21]. Close policy gaps to promote entrepreneurship and innovation in IHLs [9], [19]. **Resource Optimization:** ICT is used to overcome resource limitations and enable remote operations and virtual collaboration [2], [5], [20]. Develop hybrid incubation models to connect local innovators with global markets and investors.

B. Key Components

Innovation and Research Development: Promoting interdisciplinary research addressing local and regional challenges [3], [5], [30]. Focus on resilience-focused technologies such as e-learning and agricultural innovation. **Leveraging Remote Collaboration and Digital Platforms:** Develop online incubators or hybrid incubation models that enable remote virtual mentoring, collaboration, and market engagement. Connect students and educators with global experts, investors, and partners to bypass local restrictions [1], [47]. **Localized curriculum for entrepreneurial and digital skills development:** Implementing specialized training programs in entrepreneurship, digital literacy, and IT management tailored to Gaza's constraints. Equip students with practical skills for both local and remote employment opportunities [27], [62]. **Technology Transfer Offices (TTOs):** TTOs should be strengthened to manage intellectual property,

licensing, and knowledge transfer between IHLs and industries [5], [30]. **Partnerships with international organizations for funding and expertise:** Partners with international organizations provide financial resources, mentorship, and access to advanced knowledge [5], [22], [47]. **ICT as a Key Enabler:** Use ICT platforms for analysis, scaling and international collaboration to overcome geographical and economic constraints [15], [16], [53], [61]. **Financial and Incubation Support:** Providing access to hybrid financing mechanisms, including grants, crowdfunding, and incubation programs tailored to the Gaza Strip context [22], [63].

C. Operational Framework

Opportunity Identification: Use ICT-based analytics to identify local and global market opportunities for spin-off companies [1], [2]. **Develop solutions targeting resilience-oriented technologies for conflict-affected regions** [6], [24]. **Framework Development:** Take a bottom-up approach involving local stakeholders to design spinoffs that address community needs [23]. **Implementation and Scaling:** ICT-enabled pilot spin-offs focus on local challenges and are scalable to similar global markets [6], [24].

D. Sustainability and Impact

Monitoring and Evaluation: Leverage ICT dashboards to track and evaluate spin-off performance in real-time [5], [15], [42], [53]. **Community engagement:** Engages local communities by involving students, researchers, and community leaders in the development of spin-offs [18], [64]. **Alignment with Sustainable Development Goals (SDGs):** Align spin-off initiatives with SDGs 4 (Quality Education), 8 (Decent Work and Economic Growth), and 9 (Industry, Innovation and Infrastructure) [37], [47]. Table VI shows strategic framework summary table.

TABLE VI. STRATEGIC FRAMEWORK SUMMARY TABLE

Key Area	Description
Foundational Pillars	Triple Helix Collaboration, Policy and Institutional Alignment, Resource Optimization
Key Components	Innovation and Research Development, Remote Collaboration, Localized Curriculum, Technology Transfer Offices, International Partnerships, ICT as Core Enabler, Financial and Incubation Support
Operational Framework	Opportunity Identification, Framework Development, Implementation and Scaling
Sustainability and Impact	Monitoring and Evaluation, Community Engagement, SDG Alignment

E. Expected Outcomes

Increased Spin-Off Creation: Increase in the number of IT-enabled spin-offs that address local and global challenges [3], [5], [23]. **Economic and Social Impact:** Strengthening local economies through job creation, e-learning, and agricultural technology solutions [20], [65], [66].

Improved IHLs Capacity: Universities are becoming entrepreneurial institutions that contribute to regional innovation [5], [12], [17], [19]. **Global Competitiveness:** Gaza-based spin-offs gain global recognition and scalability through the use of ICT and digital entrepreneurship strategies [25], [34].

The following Fig. 2 illustrates the strategic framework that includes themes and sub-themes, and provides practical opportunities that should be considered and explored if universities in the Gaza Strip could successfully launch and sustain IT spin-offs despite significant constraints.

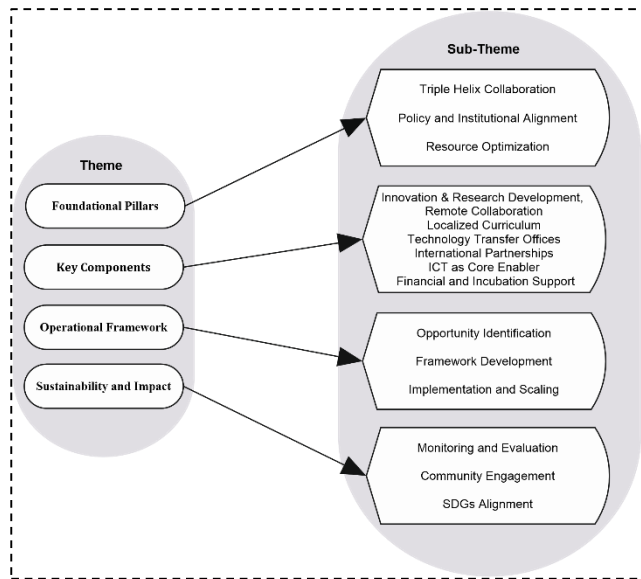


Fig. 2. Strategic framework.

By integrating ICT, remote collaboration, and local strategies, this framework enables IHLs in conflict zones, such as Gaza, to promote IT-enabled spin-offs with local relevance and global scalability. These strategies reflect a commitment to resilience and innovation, enabling universities to thrive despite adversity and achieve meaningful economic and social outcomes.

VII. CONCLUSION AND RECOMMENDATIONS

This study proposes a strategic framework for developing IT-enabled spin-off ventures in higher learning institutions in conflict-affected regions such as Palestine. It identifies key barriers such as funding constraints, political instability, skill gaps, infrastructure limitations, and weak intellectual property protection. Based on a case study of a major technical college in Palestine, this framework addressed specific challenges in the Palestinian context, including leveraging ICT, remote collaboration, and local strategies. This study emphasizes the need for healthy financing mechanisms, increased infrastructure investments, conflict-resistant strategies, skill development initiatives, and sustainable financing models that reduce dependence on foreign aid. It also highlights the importance of establishing partnerships and improving collaborations with international partners to promote IT-enabled entrepreneurship and innovation in challenging environments. Key recommendations include the following.

Although IT spinoffs hold significant potential for economic growth and translation of research into products or services, barriers to success still need to be overcome. Palestine's unique sociopolitical environment requires a tailored approach to promote IT spin-offs within institutions of higher learning (IHLs) that lead to the promotion of

entrepreneurship and innovation. While the current study provides valuable insights into the strategic framework, it is limited by the lack of empirical validation, robust statistical analysis, and longitudinal assessment of its long-term impact. Future research should address these shortcomings while incorporating mixed methods approaches, comparative analyses, and stakeholder engagement to enhance the framework's applicability and effectiveness across diverse contexts.

The findings of this study can serve as a basis for developing targeted strategies to promote sustainable IT ecosystems in universities in developing countries. The findings and recommendations provide insights that policymakers, higher learning administrators, and international organizations can use to create an enabling environment for IT spin-offs in Palestine, with the potential for broader applications in developing countries.

DECLARATION OF COMPETING INTEREST

The authors declare no conflicts of interest regarding the data or information reported in this study. The authors declare no conflicts of interest.

ACKNOWLEDGMENT

The authors would like to thank all the individuals and organizations who helped with this study. We would particularly like to thank supervisors at the National University of Malaysia and Palestinian universities for their insightful advice and support. The MIS supported this study through the Malaysian Ministry of Higher Education. The authors appreciate the resources and facilities provided by the Universiti Kebangsaan Malaysia (UKM).

The authors acknowledge the financial support from the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (under the TAP research grant). Support professionals such as Ts. Dr. Ibrahim Mohamed and Ts. Dr. Hasimi Sallehudin and Dr. Hussam F. Abushawish enriched the manuscript and constructively complemented the presentation. The Center for Software Technology Management (SOFTAM) and Palestine Technical College, Deir El-Balah (Palestine), for their mentorship and invaluable contribution. The commitment of the editor-in-chief, associate editor, and reviewers' comments for improvement is part of the effort to bring this paper into the best possible condition that it deserves. This work was supported by funding from the TAP and Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia under Grant TAP.

Word count: 4,517 words, excluding references. Ethical compliance: All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. Data Access Statement: All relevant data are included in manuscript and the supporting information files and research statistics supporting this publication are available on the Palestinian Central Bureau of Statistics website at <https://www.pcbs.gov.ps/default.aspx>. Author contributions: Mahmoud M. Musleh, and Ibrahim Mohamed contributed to

the design and implementation of the research, Mahmoud M. Musleh analyzed the results and wrote manuscript. Ibrahim Mohamed, Hasimi Sallehudin, and Hussam F Abushawish supervised the project.

REFERENCES

- [1] D. M. Steininger, "Linking information systems and entrepreneurship: A review and agenda for IT-associated and digital entrepreneurship research," *Inf. Syst. J.*, vol. 29, no. 2, pp. 363–407, 2019.
- [2] D. Vidmar, M. Marolt, and A. Pucihar, "Information technology for business sustainability: A literature review with automated content analysis," *Sustain.*, vol. 13, no. 3, pp. 1–24, 2021.
- [3] M. J. Bezanilla et al., "Developing the entrepreneurial university: Factors of influence," *Sustain.*, vol. 12, no. 3, 2020.
- [4] R. Pinheiro, E. Balbachevsky, P. Pillay, and A. Yonezawa, "Assessing the impact of COVID-19 on the institutional fabric of higher education. 2023."
- [5] J. B. Padilla Bejarano, J. W. Zartha Sossa, C. Ocampo-López, and M. Ramírez-Carmona, "University Technology Transfer from a Knowledge-Flow Approach—Systematic Literature Review," *Sustainability (Switzerland)*, vol. 15, no. 8, 2023.
- [6] J. N. Cubero, S. A. Gbadegeshin, and C. C. Segura, "Commercialization process of disruptive innovations in corporate ventures and spinoff companies: A comparison," *Adv. Sci. Technol. Eng. Syst.*, vol. 5, no. 2, pp. 621–634, 2020.
- [7] C. Crysdián, "The evaluation of higher education policy to drive university entrepreneurial activities in information technology learning," *Cogent Educ.*, vol. 9, no. 1, 2022.
- [8] H. J. Kim, "Determinants of technology-based spin-offs created by universities in Korea," *Asian J. Technol. Innov.*, vol. 28, no. 2, pp. 305–322, 2020.
- [9] C. ALDEMİR, "Educational Governance in Turkey: From the view of New Public Management and New Social Movement Theories," *Gaziantep Univ. J. Soc. Sci.*, vol. 17, no. 2, pp. 438–452, 2018.
- [10] M. Almodóvar-González, A. Fernández-Portillo, and J. C. Díaz-Casero, "Entrepreneurial activity and economic growth. A multi-country analysis," *Eur. Res. Manag. Bus. Econ.*, vol. 26, no. 1, pp. 9–17, Jan. 2020.
- [11] M. I. Aida Koni, Khalim Zainal, "AN OVERVIEW OF THE PALESTINIAN HIGHER EDUCATION," *Int. J. Asian Soc. Sci. Hasniza Yahya*, vol. 3, no. 9, pp. 1906–1912, 2013.
- [12] D. B. Audretsch and M. Belitski, "Three-ring entrepreneurial university: in search of a new business model," *Stud. High. Educ.*, vol. 46, no. 5, pp. 977–987, 2021.
- [13] P. N. Vicente, M. Lucas, V. Carlos, and P. Bem-Haja, "Higher education in a material world: Constraints to digital innovation in Portuguese universities and polytechnic institutes," *Educ. Inf. Technol.*, vol. 25, no. 6, pp. 5815–5833, 2020.
- [14] M. Ranga and H. Etzkowitz, "Triple Helix Systems: An Analytical Framework for Innovation Policy and Practice in the Knowledge Society," *Ind. High. Educ.*, vol. 27, no. 4, pp. 237–262, 2013.
- [15] Q. Alzaghaf and M. Mukhtar, "Factors affecting the success of incubators and the moderating role of information and communication technologies," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 2, pp. 538–545, 2017.
- [16] Q. Alzaghaf and M. Mukhtar, "Moderating effect of information and communication technology tools on the relationship between networking services and incubator success," *J. Eng. Appl. Sci.*, vol. 13, no. 14, pp. 5746–5755, 2018.
- [17] M. Guerrero and D. Urbano, "The development of an entrepreneurial university," *J. Technol. Transf.*, vol. 37, no. 1, pp. 43–74, 2012.
- [18] H. Etzkowitz, "Innovation in innovation: The Triple Helix of university-industry-government relations," *Soc. Sci. Inf.*, vol. 42, no. 3, pp. 293–337, 2003.
- [19] A. A. R. A. Abdullah, I. Mohamed, and N. S. M. Satar, "The Impact of E-Commerce Drivers on the Innovativeness in Organizational Practices," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 8, pp. 1348–1355, 2024.
- [20] A. H. Abukumail, "The Palestinian Information Technology Association of Companies: Moving Forward on ICT and Innovation," *MENA Knowl. Learn.*, no. 80, pp. 1–3, 2013.
- [21] Prime Minister's Office, "State of Palestine's National Development Plan: Resilience, Disengagement, and Cluster Development towards Independence (NDP 2021-2023)," pp. 1–83, 2020.
- [22] P. Pérez-Hernández, G. Calderón, and E. Noriega, "Generation of university spin off companies: Challenges from Mexico," *J. Technol. Manag. Innov.*, vol. 16, no. 1, pp. 14–22, 2021.
- [23] M. Sheriff and M. Muffatto, "University Spin-Offs: A New Framework Integrating Enablers, Stakeholders and Results," *Int. J. Innov. Technol. Manag.*, vol. 16, no. 2, 2019.
- [24] S. Battisti and A. Brem, "Digital entrepreneurs in technology-based spinoffs: an analysis of hybrid value creation in retail public-private partnerships to tackle showrooming," *J. Bus. Ind. Mark.*, vol. 36, no. 10, pp. 1780–1792, 2021.
- [25] P. C. Verhoef et al., "Digital transformation: A multidisciplinary reflection and research agenda," *J. Bus. Res.*, vol. 122, no. September 2019, pp. 889–901, 2021.
- [26] C. Fernandes, J. J. Ferreira, P. M. Veiga, S. Kraus, and M. Dabić, "Digital entrepreneurship platforms: Mapping the field and looking towards a holistic approach," *Technol. Soc.*, vol. 70, no. April, 2022.
- [27] M. J. Al Shobaki, S. S. Abu Naser, Y. M. A. Amuna, and S. A. El Talla, "The Level of Promotion of Entrepreneurship in Technical Colleges in Palestine," *Int. J. Eng. Inf. Syst.*, vol. 2, no. 1, pp. 168–189, 2018.
- [28] H. Berghaeuser and M. Hoelscher, "Reinventing the third mission of higher education in Germany: political frameworks and universities' reactions," *Tert. Educ. Manag.*, vol. 26, no. 1, pp. 57–76, 2020.
- [29] S. I. Ashmarina and G. M. Murzagalina, *Role of Universities in the Infrastructure to Support Small and Medium-Sized Business*, vol. 161 LNNS. 2021.
- [30] A. D. Daniel and L. Alves, "University-industry technology transfer: the commercialization of university's patents," *Knowl. Manag. Res. Pract.*, vol. 18, no. 3, pp. 276–296, 2020.
- [31] M. Neves and M. Franco, "Academic spin-off creation: barriers and how to overcome them," *R D Manag.*, vol. 48, no. 5, pp. 505–518, 2018.
- [32] V. Ratten, "Digital platforms and transformational entrepreneurship during the COVID-19 crisis," *Int. J. Inf. Manage.*, no. May, p. 102534, 2022.
- [33] D. W. Wainwright, B. J. Oates, H. M. Edwards, and S. Childs, "Evidence-based information systems: A new perspective and a road map for research-informed practice," *J. Assoc. Inf. Syst.*, vol. 19, no. 11, pp. 1035–1063, 2018.
- [34] G. Elia, A. Margherita, and G. Passiante, "Digital entrepreneurship ecosystem: How digital technologies and collective intelligence are reshaping the entrepreneurial process," *Technol. Forecast. Soc. Change*, vol. 150, no. 2, p. 120118, 2020.
- [35] S. Jenders, "Facility for New Market Development (FNMD) to Strengthen the Private Sector in the Occupied Palestinian Territories Final Evaluation Triple Line Consulting," no. May, pp. 1–148, 2012.
- [36] I. Ibrahim and S. Darwish, "University-Industry-Collaborations in Egypt: Academics' Perception of Motivators and Success Factors," 2022.
- [37] M. Ibrahim, "Implementing the 2030 Agenda for Sustainable Development in Palestine: An Innovation-Centric Economic Growth Perspective," pp. 4–7, 2020.
- [38] National Report, "The Higher Education System in Palestine," *RecoNow*, no. May, p. 129, 2016.
- [39] O. Abidi, V. Dzenopoljac, and A. Dzenopoljac, "Discussing the Role of Entrepreneurial Universities in COVID-19 Era in the Middle East," *Manag. Sustain. Bus. Manag. Solut. Emerg. Econ.*, pp. 1–12, 2021.
- [40] E. G. Carayannis, E. M. Rogers, K. Kurihara, and M. M. Allbritton, "High-Technology spin-offs from government R&D laboratories and research universities," *Technovation*, vol. 18, no. 1, pp. 1–11, 1998.

- [41] A. Carbonaro, J. A. M. B. Kuzelka, and F. Piccinini, "A new digital divide threatening resilience: exploring the need for educational, firm-based, and societal investments in ICT human capital," *J. E-Learning Knowl. Soc.*, vol. 18, no. 3, pp. 66–73, 2022.
- [42] R. Morrar, I. Abdeljawad, S. Jabr, A. Kisa, and M. Z. Younis, "The role of information and communications technology (ICT) in enhancing service sector productivity in Palestine: An international perspective," *J. Glob. Inf. Manag.*, vol. 27, no. 1, pp. 47–65, 2019.
- [43] A. N. Azmi, Y. Kamin, M. K. Noordin, and A. N. Ahmad, "Towards industrial revolution 4.0: Employers' expectations on fresh engineering graduates," *Int. J. Eng. Technol.*, vol. 7, no. 4, pp. 267–272, 2018.
- [44] F. Kitsios, M. Kamariotou, and E. Grigoroudis, "Digital Entrepreneurship Services Evolution: Analysis of Quadruple and Quintuple Helix Innovation Models for Open Data Ecosystems," *Sustain.*, vol. 13, no. 21, 2021.
- [45] J. Valentowitsch, F. Kianpour, T. Fritz, and W. Burr, "Doing Business in the Digital Age: Towards an Adjusted Resource-Based Model," *J. Competences, Strateg. Manag.*, vol. 12, pp. 1–22, 2024.
- [46] B. Y. R. Alharmoodi and M. M. Lakulu, "The Formulation and Validation of a Conceptual Framework for the Transition from E-government to M-government," *Eur. J. Interdiscip. Stud.*, vol. 8, no. 1, pp. 23–34, 2022.
- [47] P. Holzmann and P. Gregori, "The promise of digital technologies for sustainable entrepreneurship: A systematic literature review and research agenda," *Int. J. Inf. Manage.*, vol. 68, no. October 2022, p. 102593, 2023.
- [48] S. Tabib, "Assessing Entrepreneurship Practices at the Palestinian Higher Education Institutions," *Repos. Najah Univer*, pp. 1–169, 2021.
- [49] D. E. H. Tigelaar, D. H. J. M. Dolmans, I. H. A. P. Wolfhagen, and C. P. M. Van Der Vleuten, "The development and validation of a framework for teaching competencies in higher education," *High. Educ.*, vol. 48, no. 2, pp. 253–268, 2004.
- [50] A. C. N. Blumm and S. C. M. Barbalho, "Critical issues for an analytical framework in the relationship between academic spin-offs and their incubators," in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 2021, no. November, pp. 818–829.
- [51] Q. 'Aini ABDULLAH, N. HUMAIDI, and M. SHAHROM, "Industry revolution 4.0: the readiness of graduates of higher education institutions for fulfilling job demands," *Rev. Română Informatică și Autom.*, vol. 30, no. 2, pp. 15–26, 2020.
- [52] X. Sanchez and S. Bayona-Ore, "Strategic Alignment between Business and Information Technology in Companies," *Iber. Conf. Inf. Syst. Technol. Cist.*, vol. 2020-June, no. June, pp. 24–27, 2020.
- [53] A. Ibrahim, I. Mohamed, and N. S. M. Satar, "Factors Influencing Master Data Quality: A Systematic Review," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 2, pp. 181–192, 2021.
- [54] H. Z. Nuseibeh, A. R. Hevner, and R. W. Collins, "What can be controlled: actionable ICT4D in the case of Palestine," *Inf. Technol. Dev.*, vol. 25, no. 3, pp. 390–423, 2019.
- [55] D. S. Siegel and M. Wright, "Academic Entrepreneurship: Time for a Rethink?," *Br. J. Manag.*, vol. 26, no. 4, pp. 582–595, 2015.
- [56] T. Amjad, S. H. B. Abdul Rani, and S. B. Sa'atar, "Entrepreneurship development and pedagogical gaps in entrepreneurial marketing education," *Int. J. Manag. Educ.*, vol. 18, no. 2, p. 100379, 2020.
- [57] I. I. Tritasmoro, U. Ciptomulyono, W. Dhewanto, and T. A. Taufik, "Determinant factors of lean start-up-based incubation metrics on post-incubation start-up viability: case-based study," *J. Sci. Technol. Policy Manag.*, 2022.
- [58] D. Rutitis and T. Volkova, "Model for Development of Innovative ICT Products at High-Growth Potential Startups," *Eurasian Stud. Bus. Econ.*, vol. 19, pp. 229–241, 2021.
- [59] A. A. Mansour, "Investigating the Readiness of Ict Palestinian Organizations for Digital Transformation," 2022.
- [60] I. Siswanto and H. Raharjo, "Technology and Innovation Capitalization: A Comparative Study of Massachusetts Institute of Technology and University of Saskatchewan," *J. Phys. Conf. Ser.*, vol. 1273, no. 1, 2019.
- [61] H. Sallehudin, N. S. M. Satar, N. A. Abu Bakar, R. Baker, F. Yahya, and A. F. M. Fadzil, "Modelling the enterprise architecture implementation in the public sector using HOT-Fit framework," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 8, pp. 191–198, 2019.
- [62] D. Vidmar, M. Marolt, and A. Pucihar, "Information technology for business sustainability: A literature review with automated content analysis," *Sustain.*, vol. 13, no. 3, pp. 1–24, 2021.
- [63] M. J. Al Shobaki, S. S. Abu-Naser, Y. M. A. Amuna, and ..., "The Entrepreneurial Creativity Reality among Palestinian Universities Students," *Int. J. Acad. Manag. Sci. Res.*, vol. 2, no. 3, pp. 1–13, 2018.
- [64] M. McAdam, K. Miller, and R. McAdam, "Understanding Quadruple Helix relationships of university technology commercialisation: a micro-level approach," *Stud. High. Educ.*, vol. 43, no. 6, pp. 1058–1073, 2018.
- [65] M. Almodovar-González, M. C. Sánchez-Escobedo, and A. Fernández-Portillo, "Linking demographics, entrepreneurial activity, and economic growth," *Espacios*, vol. 40, no. 28, 2019.
- [66] M. Almodóvar-González, A. Fernández-Portillo, and J. C. Díaz-Casero, "Entrepreneurial activity and economic growth. A multi-country analysis," *Eur. Res. Manag. Bus. Econ.*, vol. 26, no. 1, pp. 9–17, 2020.

Multi-Factors Analysis Using Visualizations and SHAP: Comprehensive Case Analysis of Tennis Results Forecasting

Yuan Zhang

Physical Education Department, Northwest University, Xi'an, Shaanxi, 710069, China

Abstract—Explainable Artificial Intelligence (XAI) enhances interpretability in data-driven models, providing valuable insights into complex decision-making processes. By ensuring transparency, XAI bridges the gap between advanced Artificial Intelligence (AI) techniques and their practical applications, fostering trust and enabling data-informed strategies. In the realm of sports analytics, XAI proves particularly significant, as it unravels the multifaceted nature of factors influencing athletic performance. This work uses a rich data analysis flow that includes descriptive, predictive, and prescriptive analysis for the tennis match outcomes. Descriptive analysis uses XAI techniques such as SHAP (SHapley Additive exPlanations) with diverse factors such as physical, geographical, surface level and skill disparities. Top players are ranked; the trend of country-wise winning is presented for the last many decades. Correlation analysis presents inter-dependence of factors. Predictive analysis makes use of machine learning models, the highest overall accuracy of 80% according to the K-Nearest Neighbors classifier. Lastly, prescriptive analysis recommends specific details which can be helpful for players and coaches as well as for overall strategies planning and performance enhancement. The research underscores the significance of AI-driven insights in sports analytics, particularly for a fast-paced and strategic sport like tennis. By leveraging advanced data analytics methods, this study offers a nuanced understanding of the interplay between player attributes, match contexts, and historical trends, paving the way for enhanced performance and informed strategic planning in professional tennis.

Keywords—Artificial intelligence; data analytics; machine learning; match result prediction; XAI; SHAP

I. INTRODUCTION

In the age when technology plays a crucial role in the world, data has become a valuable commodity in any field; and sports analytics is no exception. The extension of big data in the professional sports realm has revolutionized the way performance, planning, and decision-making process, is approached. *Sports data analysis* is vital for advancing the understanding and performance of athletic activities, providing a foundation for evidence-based decision-making in sports. With the help of advanced techniques and latest analytical tools, it helps coaches, athletes, and teams to have deep insights by identifying patterns, optimizing game plan, and get better results. The analysis includes player and game statistics, game dynamics, physiological and even psychological data and thus by bringing sports science to modern computational sports

science. Moreover, the latest trends in data analysis include predictive modeling, offering insights into player fatigue, injury likelihood, and team performance under various conditions and help to predict the game outcome. Sport data analysis is being carried out in all types of sports worlds wide to gain optimal results [1]. Among sports, tennis, an aerobic and somewhat complex sport, presents an excellent chance to use data science for gaining a competitive edge and prognostic the outcomes of the matches.

Tennis is one of the most popular sports globally and is played by millions with a combination of energetics skills and physical strength, thinking ability and patience as elements such as athleticism, strategy, and power. With background foundations of 19th century, tennis has transformed into a highly competitive and technically demanding game, among both individual and team formats [1]. While other team sports tend to place their strength on the performer's abilities in the context of change and variability, such as the base depends on variant of surfaces, condition of weather, and opponents [2]. Every game turn into a battle, where a player needs to put focus on all capacities to win as fast as possible, to play using both strong and smart tactics [3]. The evolving shift of the tennis game, powered the need to enhance the technological innovations and progressive metamorphosis, continues to push the boundaries of human performance, raising tennis beyond the status of sport and turning it into both human physical and intellectual excellence based on their speed, endurance, and strength, coupled with technical precision and mental ability to boost [4]. Modern tennis involves dynamic interactions between players and surfaces, where factors such as court type, weather, and player tactics highly impact the outcomes of matches. This complexity makes tennis match a captivating sport, both as a form of entertainment and as a subject of in-depth analysis.

Match results prediction is accomplished by examining key performance indicators such as rally length, spin rates, serving efficiency, player movement, shot placement, and other factors that may influence match results. As sport evolves, data-driven approaches are becoming increasingly crucial for increasing in player performance, refining strategies of coaching, and improving in match outcomes. The integration of AI and data analytics in tennis research has opened new gateway for understanding player behavior, optimizing performance, and predicting outcomes [5]. By leveraging vast datasets that include player statistics, physical factors, match outcomes, and even skills analysis, AI models provide unpredictable in-depth

*Corresponding Author

analysis into various aspects of the game. AI-based predictive analytics use machine learning (ML) models to process data to predict outcomes based on player strengths, weaknesses, and historical performance against specific opponents. Such prediction is used by the coaches in defining training and modifying match strategies. Data analytics based on AI has been used in tennis for analysis of multi-perspective and to achieve enhanced precision [6]. In existing studies, the researchers have focused on the limb movements and force generation modes of athletes [7].

To provide a clear structure for the presentation of the research, the paper is divided into five sections. Section II provides a detailed analysis of our research contribution based on objectives. The existing literature review in Section III discusses prior research and insight research gaps. The specifics of dataset, data preprocessing, and the methods used for analysis are explained in Section IV of the study. Section V contains result and discussion, rely on descriptive, predictive and perspective analyses. The study findings are discussed in this section and related to the objectives of the research. Lastly, Section VI of the paper provides a recap of major conclusions and recommendations to extend the present study and advance the knowledge in the field.

II. OBJECTIVES

In this research study, we aim to carry out a comprehensive analysis of the tennis dataset based on real world data of more than three decades. It shares details about the exploration of various factors which may influence the match outcome. For exploratory data analysis, the factors of various perspectives are considered. The features of players are considered and then the features of losers and winners are separately considered. The distribution of aces is visualized based on diverse surface areas. The pair plot of winners and losers are explored. The correlation matrices are computed for diverse types of features. The predictive analysis consists of application of various data mining algorithms for classification which include K-Nearest Neighbor (KNN) and Ridge Classifier (RC) and Label-Propagation (LP). The results are evaluated based on accuracy, precision, recall, and F-measure. Lastly prescriptive analysis presents the recommendation of various strategies. The main contributions can be summarized as follows:

- Developed a data-driven framework that coupled with ML models on domain-specific factors to anticipate tennis match outcomes, providing actionable deep insights for players, coaches, and analysts.
- Analyzed the impact of seven key factors, including player ranking, performance ranking, player characteristics, demographic, physical health, surface level, and skill level analysis on tennis match predictions.
- Comprehensive data analytics are carried out using three diverse approaches of descriptive analysis, predictive analysis and prescriptive analysis.

- Exploration data analysis of real-world data varies out using state of the art Data visualization
- Using SHAP method for interpretation of top factors which is widely used in the latest eXplainable Artificial Intelligence perspective.
- Achievement of accuracy as high as 80% to predict the outcome of the match using lazy classifier of nearest neighbor, demonstrating its effectiveness for tennis match outcomes.

III. RELATED WORK

The use of data analytics for sports data analysis is an active research area due to its significance [8]. As one of the most powerful machine learning algorithms, it provides the highest precision and performance when computing. It is a favorite among researchers and extensively used in several fields. A vast number of works emphasize the capacity of ML in forecasting and assessing talented tennis players and performance, following series of steps from data collection to evaluation as shown in Fig. 1. Thus, Panjan [9] considered the determination of predicted results of young athlete's skills and physical measurements as one of the ML models getting high results especially in the female sportspersons' evaluation. It was a much better way to select coaches than just picking the one that has been in the industry for a long time. Siener [10] pointed out that more concepts are relevant for consideration, and excluded physical abilities and early performance measures as specific metrics cannot adequately capture prediction models. Related to this, ML has also been used in player categorization according to their performance. Filipic [11] conducted predictive analysis to classify professional tennis players into quality groups based on the ATP rates. It helps us as coaches and the players to analyze strengths and weaknesses of team and individuals. It also pointed out that there are other success factors besides performance strategies, including mental hardness, training approaches, and psychological strength [12]. Makino et al. [13] selected the ATP singles match to analyze point winners when influenced by the court and the players' style. To illustrate the ability to practice different and more creative forms of analyzing data, Almarashi et al. [14] took a different more creative approach by demonstrating the ability to predict trends of players' performance over a period. Chen and Groll [15] used decision tree algorithm, and the results showed that it yields high level of accuracy in predicting the match outcomes for both men and women's tennis as was also discovered by Ghosh et al [16] who used logistic regression. It has also been attempted to identify some sample employing unsupervised learning methods. Whiteside and Reid [17] applied k-means clustering to decide on the best locations for aces, where data points must be grouped based on their likeness. Li et al. [18] then trained convolutional neural networks (CNNs) on images to predict batting strength and angles, due to the success of the CNN for image recognition.

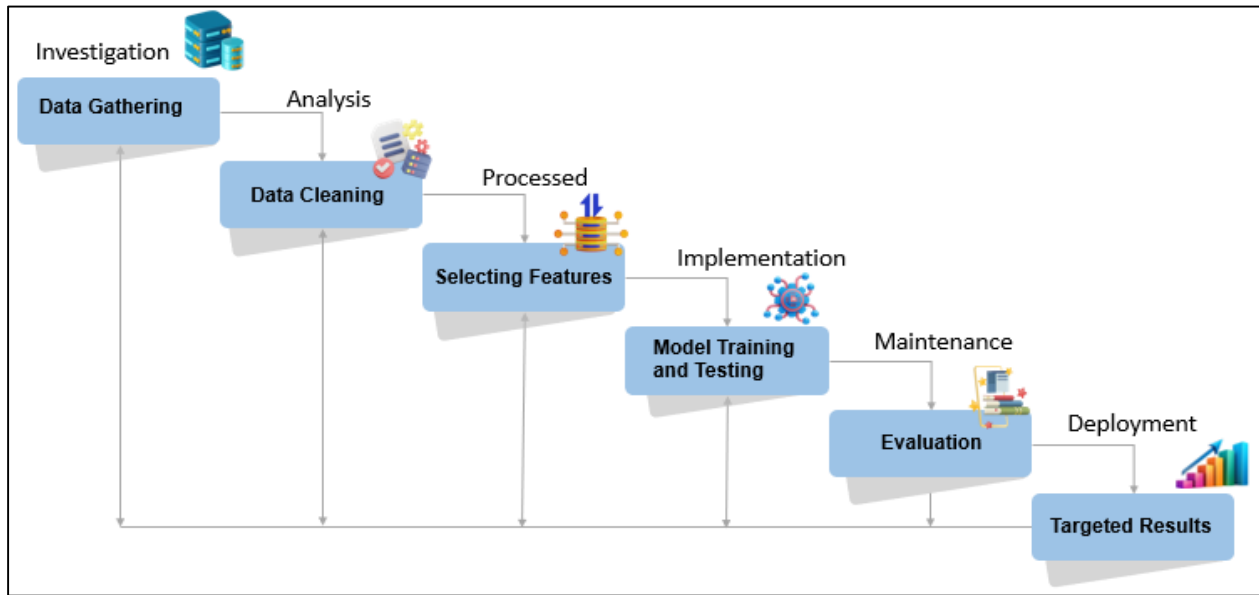


Fig. 1. An overview of the ML method, showing the iterative steps from raw data preprocessing to deploying the candidate model for applications.

In Zhou and Liu [19], different probabilities of different stances in the court were recently Bayes network predicted. Schulc et al. [20] used an LSTM network where the network learned from the video data to detect biomechanical signs of ACL injury risks. The LSTM network was able to accurately predict at-risk athletes with 75%-81% accuracy. Based on the data mining methodology, Jain et al. [21] explored sports performance, evaluated it according to benchmark models of key factors, technical aspects, and tactical difficulties confronting Chinese athletes. Together, these works establish the elaborate use of ML for the promotion of tennis proficiently in many areas such as talent recognition and changes analysis of performance and injury handicaps.

IV. MATERIALS AND METHODS

In the following part of this paper, we focus on the method of this research by considering the empirical data used for collecting, cleaning, and applying for the purpose of this research, which is a prediction of tennis matches. The data used are tennis match statistical and predictive analysis, available at open-source platforms, that includes attributes based on ranking of players, their characteristics, physical factors, skills factors, surface type, tournament conditions, and match duration. In this context, the data collected is preprocessed to clean it by using data imputation methods to handle missing data and applying techniques to erase or eliminate records with many missing entries to maintain internal consistency. These methods cover three approaches for comprehensive analysis. Descriptive Analysis highlights the feature analysis. Predictive Analysis explore the correlation between parameters. By employing various models including KNN, RC, and LP to create the overall model on the factors leading to match results. KNN classifier, as working illustrated in Fig. 2, assigns a class y to a data point x , its k -th nearest neighbors are determined using a distance metric $d(x, x_i) = \sqrt{\prod_{j=1}^n (x_j, x_{i,j})^2}$ based on predicted class neighborhood points $\hat{y} = mode\{y_i: x_i \in Neighbors(x)\}$.

Furthermore, RC algorithm based on linear model that minimizes a loss function with L2 regularization $X \in \mathbb{R}^{n \times p}$ (features) and $y \in \{-1, 1\}^n$ defining predictive labels using weight vectors w corresponding to regularization strength $\alpha > 0$ for prediction $\hat{y} = sign(Xw) \rightarrow \min_w \|Xw - y\|_2^2 + \alpha \|w\|_2^2$. Another graph based semi-supervised learning algorithm known as label propagation that propagates labels from labeled to unlabeled points iteratively depends on graph based-approach $G = (V, E)$ with weight matrix W and label distribution $F \in \mathbb{R}^{n \times c}$ based on classes $F^{(t+1)} = D^{-1}WF^{(t)}$. Prediction of labeled class computed using diagonal degree D with matrix of W . This process continues until convergence, and labels are assigned based on the maximum in F . Such models of analytics were trained and tested using historical information to provide predictions of the match outcomes with good levels of effectiveness. Furthermore, Perspective Analysis offers empirical evidence for a data-driven approach for making robust strategic planning, evaluation of performance, and decision making in professional tennis, and reveals how player characteristics and match conditions collectively determine performance.

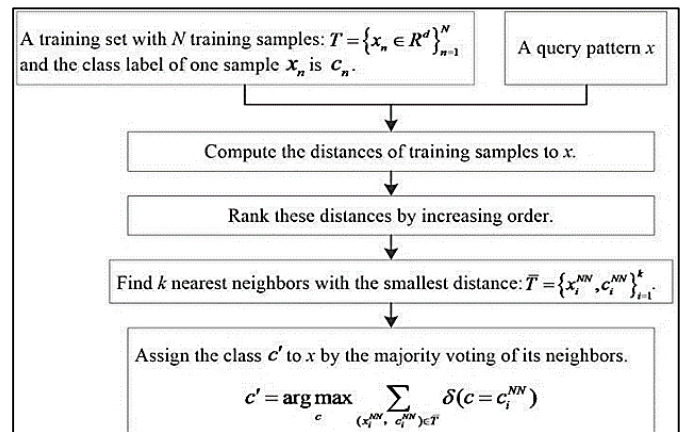


Fig. 2. The K-nearest neighbors (KNN) architecture.

V. RESULT ANALYSIS

A. Descriptive Analysis

A statistical analysis of the attributes of a tennis match shows a comprehensive analysis of how several factors correlate to make an impact on a player’s probability of winning and other factors as well, taxonomy shown in Fig. 3. The visualizations highlight the key attributes such as player rankings, physical fitness, and performance metrics, which collectively contribute to predicting match outcomes. All these

factors have a unique function of providing an understanding of the nature of competitive tennis. The related factors of physical fitness, as shown in Fig. 4, reveal the rank ratio of winner and loser as well, highlight the ability of ranking for higher predictions. Lower ranked players relatively closer to 1 are over-represented among winners demonstrating the player ranking – which is an average of the earlier performance progress, stability and competitiveness – is perhaps the single strongest determination of match outcomes.

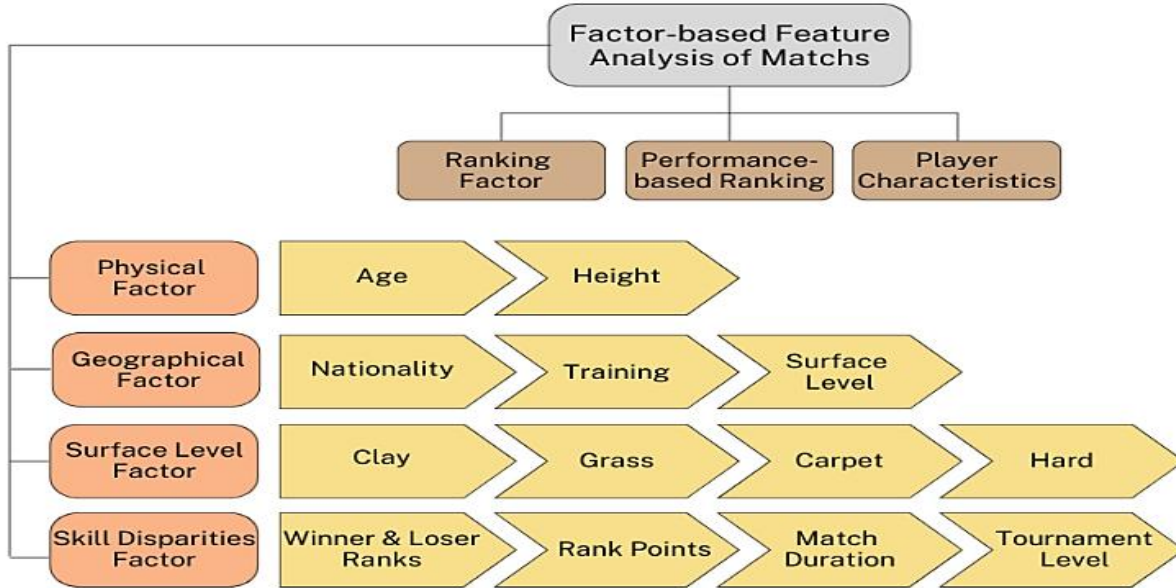


Fig. 3. Taxonomy of factor-based analysis of tennis match prediction.

This reinforces the idea that rankings are not merely statistical markers but reliable indicators of a player’s performance and fitness. Similarly, features like winner and loser rank points, which are measures of ranking points acquired over a certain period, additional support sustained like ranking and competitive success as factors affecting match outcomes. Player age also emerges as a significant factor, with winners mainly under the age of 25 according to the distributions of variables winner and loser age factors. This age group characterized the peak years of physical agility, strength, and psychological resilience. This argument is further emphasized by the fact that the decline in performance observed in older players is captured by the fact that the loser age distribution tapers off, which is evidence of physicality of tennis and the fact that with age, the performance of players reduces with increasing age regardless of league ranking. Height, as captured by winner and loser height shows a more complex interaction. The distribution of player heights according to general population values, and their average of 180-190 cm indicates that height can be useful – probably in serving and court coverage – but certainly is not as definitive as ranking or age. This simply means that, though factors such as height have additional marginal utility they outweigh in their skill, strategy and mental strength.

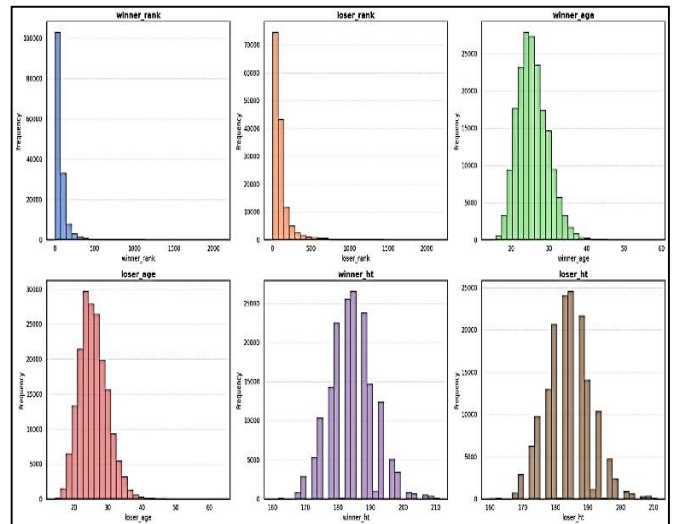


Fig. 4. Physical fitness factors affecting players performance.

The SHAP based summary plot, shown in Fig. 5, further supports these observations by showing how proposed features influence match outcomes. Factors like first, and second rank, and their associated ranking points dominate the feature

importance, emphasizing that prior performance is paramount in determining match success. Interestingly, variables such as first and second age, factors, and tournament-specific details include analysis about round, draw size, and tourney month with exhibit moderate importance meaning that even external factors affect performance, such as the tournament stage or environmental factors, can also influence progress. For instance, the level of the tournament or the number of sets played highlighted as best of attribute might benefit more experienced or physically conditioned players.

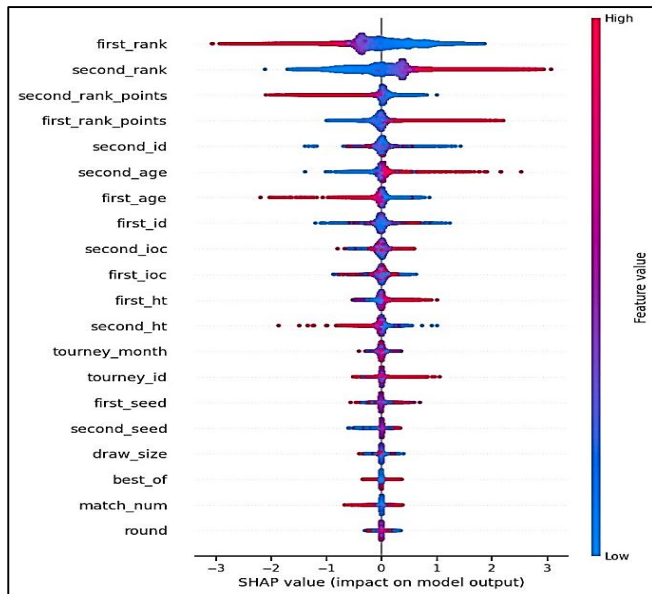


Fig. 5. SHAP analysis of features importance.

Fig. 6 provides a detailed breakdown of feature importance of predicting match outcomes. It underscores the significance of first rank pointed to winner's rank and second rank shows the progress analysis of loser's rank, guarantees that player rankings, which reflect skill, regularity, and past results, are the strongest indicators. Other features, such as second rank points highlight the ranking points of loser's players and second impactful factor first height attribute pointing to the chances of winners on the base of their height, reflecting that performance measures and physical fitness attributes are significant as secondary impact factors. The plot further underlines the contribution of other relatively significant variables with aspect

of nationality background as mentioned loser's nationality code, which characteristics of the player's performance in shaping geographic or cultural patterns. The visual strength of SHAP showing how much each feature contributes to match prediction while reinforcing the idea that, although rankings overwhelm, other features bring context.

The bar chart in Fig. 7 illustrates the Top 10 Players by career wins, including legends like Jimmy Connors, Roger Federer, and Rafael Nadal, also supports these findings. These players consistently rank among the best due to their ability to maintain high performance over longer periods, which is in tune with the ranking and ranking-points identified in the evaluation of the data. The success of these players also uncovers an important part of the equation, which is psychological factor, such as mental strength and match experience, which are inferred from performance measures such as rating. The boxplot as shown in Fig. 8 depicting the distribution of aces by surface (Clay, grass, carpet and hard) highlights into how playing conditions affect serve performance. According to the distribution, Grass courts exhibit the widest distribution and median number of aces, which shows the serve on this surface is preferable for powerful players. In contrast, clay courts are characterized by a lower median and distribution, suggesting that this slow playing surface reduces the impact of aces. This information emphasizes the fact that surface type must be considered important indicating match results particularly for those players who rely on serve base. The USA dominated tennis in the last parts of the twentieth century, as was mentioned, which could also be explained by the fact that the game during that time was especially suitable for players who use powerful strikes and played fast courts, as shown in Fig. 9. But this dominance was not sustained after the year 2000, since more emerging nations approached the game with new generation of players like Spaniards and the Serbian stars who demonstrate equal powers on clay and other surfaces. Spain happened to rise steadily at the same time as its emphasis on clay court preparations, while Serbia on a similar note rose with players like Novak Djokovic. This rise of the Swiss team in the period of Federer-Wawrinka partnership show how player generations can skew national statistics. The above-presented patterns indicate that player origin and era-specific patterns are functional contextual predictors that influence match outcomes due to the general competitive conditions and training processes.

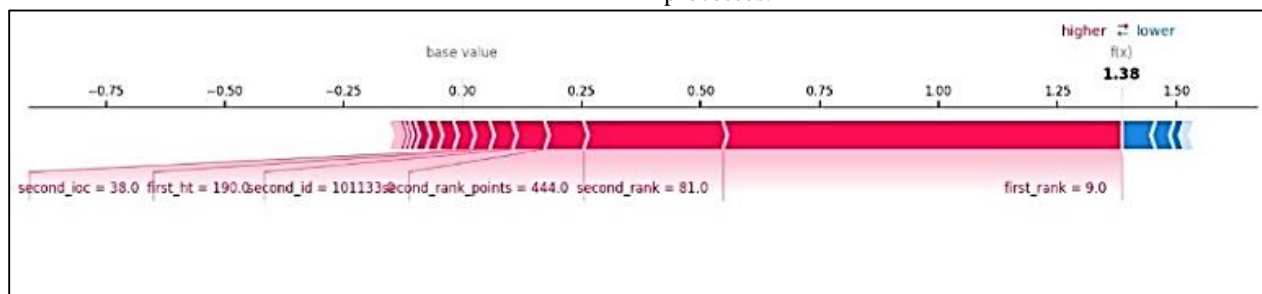


Fig. 6. Breakdown analysis of match outcomes.

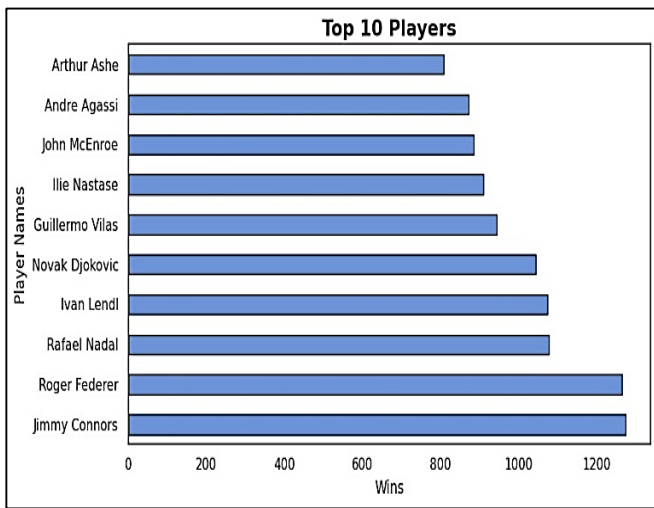


Fig. 7. Analysis of top 10 players performance.

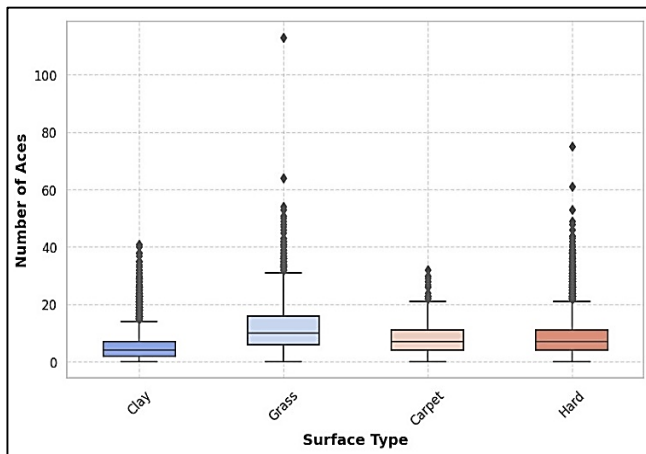


Fig. 8. Distribution of aces by surface.

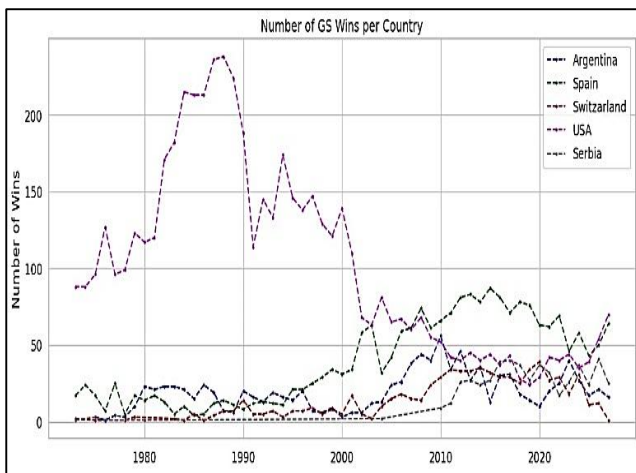


Fig. 9. Nation-wise performance of players statistics.

Analyzing variables using pair plot visualization as shown in Fig. 10, indicating factors such as winner and loser rank, match minutes that show the overall duration and tourney level indicating participating tournament score, highlighting the comprehensive analysis by examining the interplay between

rankings and match intensity. Matches characterized by players with higher ranks are generally shorter, pointing to their dominance and ability to close matches efficiently. Conversely, Players with low level ranking scores tend to engage in longer matches, indicating closely contested battles where differences in skill are less pronounced. The pair plot also separates Grand Slam matches (G) from Masters tournaments (M) where the duration is usually longer because of higher level of tension, stress, anxiety, among all still showing a best ranking with physical fitness, a significant factor to ranked as winner for players. This observation illustrates that tournament setting affects matches setting, by considering the external factors other than players' characteristics predicting the matches' model.

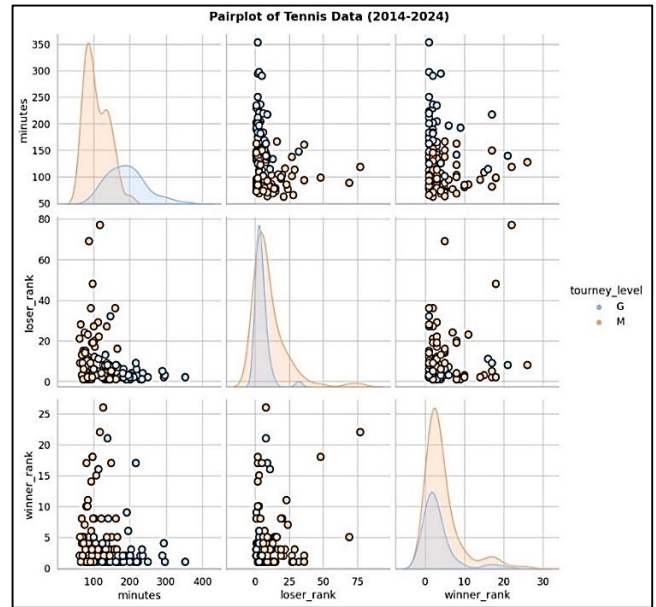


Fig. 10. Analysis of skill disparities factors.

The two correlation matrices as shown in Fig. 11 offer a comprehensive examination of the various relationships in relation to different variables in the database: tournament/match attributes and players' performance indicators. The features of this correlation include the date of the tournament; draw size; the match number; and performance indicators of the player who lost the match; aces served and committed, double faults, total serve attempts, first serves made, and break points faced. There is a very tight positive relationship between Attempts and First (0.93), suggesting that many serve points attempted deliver a high probability of successful first serve hitting. As with many of the other performance indicators, there is a strong relationship between break points saved and break points faced (0.92) – players that frequently find themselves on the wrong end of a break point usually show that they can sustain a lot of those situations. The strong positive relationship which is evident between the variable's games served and serve points attempted ($r=0.94$ mean, reveals the direct relationship between the number of service games played serves attempted. Low correlations between the date of the tournament and draw size and most of the performance indicators indicate that these characteristics have little influence on the result of ongoing inspired matches.

The Players Information and Performance Association Matrix looks at how certain variables in mutating with player characteristics which include the winner seed, winner height, winner rank, and match performance indicators which include aces served, double faults, break points faced, and games served by the winner, in Fig. 12. Cohesion between first serves in and serve points attempted by the winner is also evident with a coefficient value of 0.94 for the pair of variables. The finding that linkage of games served, and the break points faced by the winner ($r=0.94$) suggests that the consistency of the serving players is likely to go down with the game faced on their serve as he matches progress especially in terms of break points faced. The correlation -0.33 of winner rank and winner rank points indicate that players with low numerical value of rank will tend to gain more ranking points because they have performed better than other players over the season. They both showed some relationship with match duration to some key performance indicators, and this was evident in the breakdown of the break points saved and the first serves clinched to show how stamina and service comes in handy during long drawn-out matches.

Overall study highlights the evaluating probabilities of tennis match outcomes are a complex process and factors including player rankings, age, and prior performance emerging as the most influential features for the match outcomes. Height, tournament conditions, and match dynamics provide the secondary features that add more context and depth to the predictive model bringing more of the game into the analysis. The analysis highlights that while player rankings and previous performances predicting environmental and intrinsic aspects including nature of the ground, playing duration, and players' physical characteristics including height and serve effectiveness enhance the complexity of the game adding on to the analysis. Such results suggest that more global and data-driven strategy is required to accomplish successful modeling of match results. By leveraging data analytics methods, we can build more robust systems that reflect the interplay of skill, strategy, and resilience in tennis match. This approach does not only improve the efficiency of forecast, but can also define conceptual framework for action, improvement of performance, and decision-making in professional tennis.

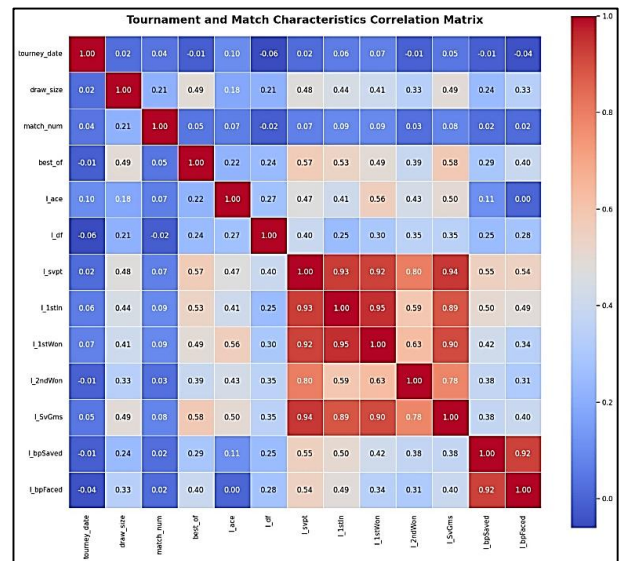


Fig. 12. The relationship between players' characteristics like age, height and ranking levels on the players' performance.

B. Predictive Analysis

The findings from the three classifiers include K-Neighbors Classifier, Ridge Classifier, and Label Propagation, in terms of accuracy, F1-score, and ROC based Area Under the Curves (AUC) values that show how the predictive models performed in the experiment with the goal to assess the strengths and limitations of using the selected algorithms for tennis match predictions. Detail analysis of results is shown in Table I.

The K-Neighbors Classifier gives the highest accuracy of 80%, with 63% F1-measure shows that the model provides a high probability of projecting the match outcome accurately most of the time, which is promising for applications where precise predictions are important. However, the values of the F1-score equal to 63% mean that the model makes moderate accurate predictions relying on players' performance with tournament levels, leads towards may some challenges in identifying less frequent match outcomes, potentially leading to some false positives or false negatives. Finally, the AUC of 78% also validates the model efficiency in determination of between the two classes of players as winner and loser but still more work is needed to be identified by the optimal decision threshold. The Ridge Classifier performs less accurate as compared to K-Neighbors Classifier achieves only 71% accuracy, utilizing both precision and recall, given its considerably higher F1-score, 69%. This implies that Ridge Classifier could be much more suitable for identifying both 'winners' and 'losers' particularly in formulation whose class distribution is skewed. Its lower AUC of 71% shows that the model does not perform as well regarding the ability to classify match outcomes throughout the probability distribution, with particular emphasis on the low success of the distinction between the positive and negative classes at various thresholds. This shows that, although the Ridge Classifier is quite balanced in terms of predicting outcomes.

Another classifier, Label Propagation tested with 77% accuracy is nearer to both models in the raw predicting power when it comes to predicting match outcomes. However, its F1-

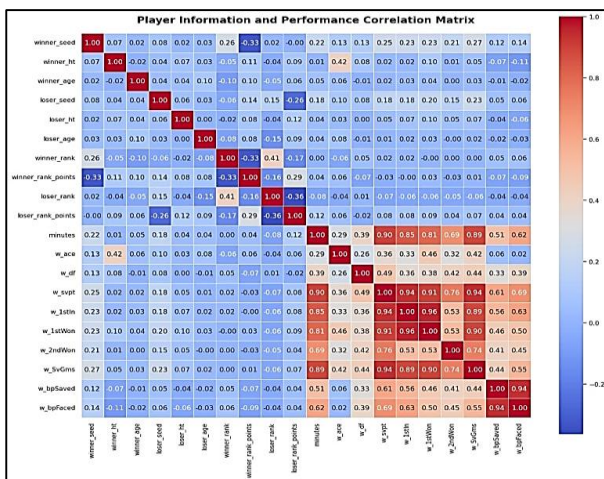


Fig. 11. Correlation analysis of player performance and tournament information.

score is 65% and it is lower than the two models we considered: K-Neighbors Classifier and Ridge Classifier which means that its precision/recall co-efficient is less accurate than these two. This has the implication that the model could be correctly classifying more instances, particularly in the minority class, resulting in either false classification as positive or as negatives. The AUC of 71% also shows us that it does not rank as high as the K-Neighbors Classifier in terms of the model’s ability to show the difference between the match outcomes based on various factors, but it is better than the Ridge Classifier. Overall, Label Propagation has a reasonable level of predictive accuracy as for match outcomes, but this model has a low ability to set a moderate ratio of precision and recall as well as it has weak discrimination in contrast to other models.

TABLE I. ANALYSIS OF CLASSIFIERS FOR PREDICTIVE MATCH OUTCOMES (%)

Model	Accuracy	Precision	Recall	F1-Score	AUC
KNN	80	75	70	63	78
RC	71	72	66	69	71
LP	77	71	60	65	71

Overall, K-Neighbors Classifier surpasses the other two in its accuracy AUC, meaning that K-Nearest Neighbors Classifier, indicating that it is better at making correct predictions and distinguishing between match outcomes. However, looking at the F1-score, it can be concluded that it could be allowed better ratio of precision and recall values. The Ridge Classifier proves to improve the balance of classification but offends accuracy and AUC value. Label Propagation, while offering good accuracy again is not good in f1-measures. These results highlight the trade-offs between model performance metrics and underscore the need to select a model based on the specific requirements of the task, such as whether the priority is maximizing prediction accuracy ability to distinguish between classes, as analysis shown in Fig. 13.

C. Perspective Analysis

Let us now focus on the third type of data analytics approach of prescriptive analysis which focuses on recommending specific strategies based on data analysis. The aim of this type of analysis is to get the desired outcomes based on analysis of historical data, and application of predictive models. Unlike descriptive analysis, which explains what has happened and which is main part of this manuscript as well, and predictive analysis, which forecasts what might happen, prescriptive analysis shares the answer to the main question of what is required to be done.

Prescriptive analysis uses data-driven insights to recommend specific training and strategies tailored to players’ needs and goals. For making robust strategic planning, evaluation of performance, and decision making in professional tennis, and reveals how player characteristics and match conditions collectively determine performance. By analyzing player attributes for prescriptive analytics provides actionable recommendations for optimizing performance. These insights into a player’s efficiency or their success on specific surfaces can guide them to match preparation

strategies. Fig. 14 shows Receiver Operating Characteristics (ROC) curve is shown for comparison and shows Area Under the Curve (AUC) too.

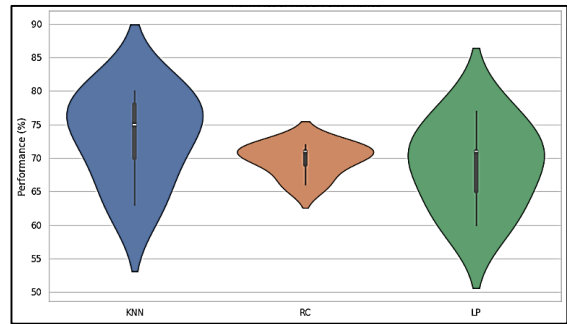


Fig. 13. This comparison reflects the accuracy measures of all applied models providing the performance evaluation of the ML approaches.

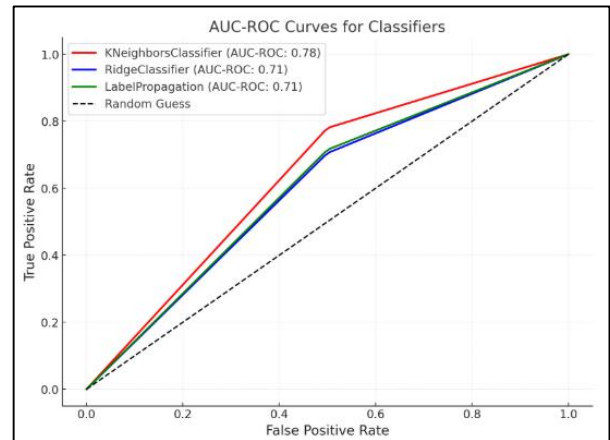


Fig. 14. The ROC-AUC curve of all applied models.

Additionally, understanding the strategies of opponent abilities and match dynamics enables players and coaches to follow strategies in real time, boost their competitive edge. Effective workload coordination in team settings to guard against injuries while maximizing on output from the players. With information concerning players, training frequency and types, courses can be constructed that would maximize their recovery period. Besides, this approach also improves the talents’ performance at the personal level and proper coordination between coaches, physiotherapists, and analysts. Further, based on the same perceptions tournament organizers and stakeholders can better schedule tournaments in a way that both protects fairness of competition and players’ health leading to enhanced experience. Such systems develop mutual constituencies of resources for supporting players and their sustainable performance in the sport.

In the previous works focused on tennis match result prediction by using the ML models, the numerous approaches and the features have been considered to improve accuracy, shown in Table II. The study in [22] (2022) used Logistic Regression (LR) with the features that aspect like surface type and being a winner or a loser besides the rank having an accuracy of 77%. Another approach made by [23] (2024) used Random Forest (RF) and concentrated on win/loss patterns. The suggested approach has a slightly lower accuracy of 70%.

Also, [24] (2024) put forward the Stochastic Forest Model with a player's win rate as the feature, with 74% accuracy. Work by [25] (2021) that integrated LR, DT, and RF models for changes in direction during matches gave a 75% result. Conversely, the proposed study (2024) presented the K-Nearest Neighbors (KNN) model that uses seven various features; the findings recorded a ninety percent accuracy; therefore, meaning better predictive capacity. This work sheds light on shifting paradigms of a predictive model of tennis match result based on machine learning involving feature evaluation and model selection as crucial success factors.

TABLE II. COMPARATIVE ANALYSIS WITH EXISTING STUDIES

Sr. No	Ref	Model	Features	Results (Acc: %)
1	[22] - 2022	LR	surface , Winner/Losser, Rank Rounds	77
2	[23] - 2024	RF	win/loss trends	70
	[24] - 2024	Stochastic Forest	player's win rate	74
3	[25] - 2021	LR, DT, RF	changes of direction	75
4	Proposed - 2025	KNN	Seven Features in Fig. 3	80

VI. CONCLUSION

Sports have always played a pivotal role in human culture, blending skill, strategy, and physical excellence. The coupling of artificial intelligence and data analytics into sports area highlights unparalleled opportunities to increase the power of decision-making, predict outcomes, and optimize performance rate. This research introduced three strategies of data analytics methods to investigate the factors influencing tennis match outcomes based on descriptive analysis, predictive analysis and perspective analysis, with a focus on feature-based analysis of attributes such as ranking attributes, physical attributes, and match conditions, emerges as significant predictors, providing valuable insights into the multifaceted nature of the sport. Among the models applied, the K-Neighbors Classifier achieved the highest accuracy of 80%, pointing out its potential as an effective tool for predictive analysis in tennis. This research highlights the potential of integrating advanced predictive models to help players, coaches, and analysts in strategic planning and performance optimization. Although the research study is helpful for understanding the factors for better tennis performance using XAI techniques however it is the limitation of the study that these findings are not generic and may not be applicable to other sports but only limited to tennis only. Considering the futuristic scope of the research work, let us share that the several future work can be considered for improvements can be made to increase the reliability of the predictions for more practical applications

- Higher level of data cleaning and applying diverse feature engineering techniques to refine and obtain increased quality data that would be used for developing predictive models

- Further tuning of advanced classifiers can be done by integrating hyper parameters of the classifiers to increase accuracy.
- Extending the work to conduct time-series analysis to make effective use of temporal characteristics of the data including player patterns over different tournaments.

This research not only offers understanding of the various factors of tennis match but also lays the groundwork for future explorations in sports analytical applications. This study thus clears the way for further enhancement to identify more robust methods and constructions through which data-driven approaches and models can be developed and deployed for sports and competition domains.

REFERENCES

- [1] Kaur, Amandeep, Ramandeep Kaur, and Gagandeep Jagdev. "Analyzing and exploring the impact of big data analytics in sports sector." SN Computer Science 2, no. 3 (2021): 184.
- [2] Liu, Sheng, Chenxi Wu, Shurong Xiao, Yaxi Liu, and Yingdong Song. "Optimizing young tennis players' development: Exploring the impact of emerging technologies on training effectiveness and technical skills acquisition." Plos one 19, no. 8 (2024): e0307882.
- [3] C. Janiesch, P. Zschech, K. Heinrich, Machine learning and deep learning, Electron. Mark. 31 (3) (2021) 685–695.
- [4] C. Shorten, T.M. Khoshgoftaar, B. Furht, Deep Learning applications for COVID-19, J. Big Data 8 (1) (2021) 1–54.
- [5] Y. Fang, B. Luo, T. Zhao, D. He, B. Jiang, Q. Liu, ST-SIGMA:spatio-temporal semantics and interaction graph aggregation for multi-agent perception and trajectory forecasting, CAAI. Trans. Intell. Technol. 7 (4) (2022) 744–757.
- [6] D.G. Ranganathan, A study to find facts behind preprocessing on deep learning algorithms, J. Innov. Image Process. 3 (1) (2021) 66–74.
- [7] J. Van der Laak, G. Litjens, F. Ciompi, Deep learning in histopathology: the path to the clinic, Nat. Med. 27 (5) (2021) 775–784.
- [8] Sampaio, Tatiana, João P. Oliveira, Daniel A. Marinho, Henrique P. Neiva, and Jorge E. Morais. "Applications of Machine Learning to Optimize Tennis Performance: A Systematic Review." Applied Sciences 14, no. 13 (2024): 5517.
- [9] Panjan, A.; Šarabon, N.; Filipčič, A. Prediction of the Successfulness of Tennis Players with Machine Learning Methods. Kinesiology 2010, 42, 98–106.
- [10] Siener, M.; Faber, I.; Hohmann, A. Prognostic Validity of Statistical Prediction Methods Used for Talent Identification in Youth Tennis Players Based on Motor Abilities. Appl. Sci. 2021, 11, 7051.
- [11] Filipčić, A.; Panjan, A.; Sarabon, N. Classification of Top Male Tennis Players. Int. J. Comput. Sci. Sport 2014, 13, 36–42.
- [12] Bozd'ech, M.; Zhán'el, J. Analyzing Game Statistics and Career Trajectories of Female Elite Junior Tennis Players: A Machine Learning Approach. PLoS ONE 2023, 18, e0295075.
- [13] Makino, M.; Odaka, T.; Kuroiwa, J.; Suwa, I.; Shirai, H. Feature Selection to Win the Point of ATP Tennis Players Using Rally Information. Int. J. Comput. Sci. Sport 2020, 19, 37–50.
- [14] Almarashi, A.M.; Daniyal, M.; Jamal, F. A Novel Comparative Study of NNAR Approach with Linear Stochastic Time Series Models in Predicting Tennis Player's Performance. Bmc Sports Sci. Med. Rehabil. 2024, 16, 28.
- [15] Dindorf, C.; Bartaguiz, E.; Gassmann, F.; Fröhlich, M. Conceptual Structure and Current Trends in Artificial Intelligence, Machine Learning, and Deep Learning Research in Sports: A Bibliometric Review. Int. J. Environ. Res. Public Health 2022, 20, 173
- [16] Ghosh, S.; Sadhu, S.; Biswas, S.; Sarkar, D.; Sarkar, P.P. A Comparison between Different Classifiers for Tennis Match Result Prediction. Malays. J. Comput. Sci. 2019, 32, 97–111.

- [17] Whiteside, D.; Reid, M. Spatial Characteristics of Professional Tennis Serves with Implications for Serving Aces: A Machine Learning Approach. *J. Sports Sci.* 2017, 35, 648–654.
- [18] Li, J.; Zhang, X.; Yang, G. The Biomechanical Analysis on the Tennis Batting Angle Selection Under Deep Learning. *IEEE Access* 2023, 11, 97758–97768.
- [19] Zhou, J.Q.; Liu, Y. Probability Prediction of Groundstroke Stances among Male Professional Tennis Players Using a TreeAugmented Bayesian Network. *Int. J. Perform. Anal. Sport* 2024, 1, 13.
- [20] Schulc, A.; Leite, C.B.G.; Csákvári, M.; Lattermann, L.; Zgoda, M.F.; Farina, E.M.; Lattermann, C.; Tóóser, Z.; Merkely, G. Identifying Anterior Cruciate Ligament Injuries through Automated Video Analysis of In-Game Motion Patterns. *Orthop. J. Sports Med.* 2024, 12, 23259671231221579.
- [21] Jain, Praphula Kumar, Waris Quamer, and Rajendra Pamula. "Sports result prediction using data mining techniques in comparison with base line model." *Opsearch* 58, no. 1 (2021): 54-70.
- [22] Solanki, Shivans, Vikas Jakir, Akshay Jatav, and Dishant Sharma. "Prediction of tennis match using machine learning." *International Journal of Progressive Research In Engineering Management And Science (IJPREAMS)* 2, no. 06 (2022).
- [23] Hu, Jinming, Xiaohua Yang, Zixuan Huang, and Jinqi Xie. "Machine Learning in Tennis Match Analysis: Predicting Score Point Victor and Momentum Shift." In *2024 5th International Conference on Machine Learning and Computer Application (ICMLCA)*, pp. 21-25. IEEE, 2024.
- [24] Lv, Yinghui. "Research on Tennis Match Strategies Based on Machine Learning and Markov Chain Modeling." *Highlights in Science, Engineering and Technology* 92 (2024): 459-466.
- [25] Giles, Brandon, Peter Peeling, Stephanie Kovalchik, and Machar Reid. "Differentiating movement styles in professional tennis: A machine learning and hierarchical clustering approach." *European Journal of Sport Science* 23, no. 1 (2023): 44-53.

Exploring Diverse Conventional and Deep Linguistic Features for Sentiment Analysis of Online Content

Yajun Tang*

Anhui Business and Technology College, Hefei City, Anhui Province, 231131, China

Abstract—Social media has changed the world by providing the facility to common person to share their views and generate their own content, known as Users Generated Content (UGC). Due to huge volume of UGC data being created at great velocity, so to analysis this big data, latest AI (Artificial Intelligence) and its sub-domain NLP (Natural Language Processing) are being used. Sentiment analysis of online content is an active research area due to its vast applications in business for review analysis, social and political issues. In this research study, we aim to carry out sentiment analysis of online content by exploring conventional features like Term Frequency – Inverse Document Frequency (TF-IDF), Count-Vectorization, and state of the art word embeddings based word2vec. Extensive exploratory data analysis has been carried out using the latest data visualization approaches. The main novelty lies in the application of unique and diverse machine learning algorithms on social media datasets and the results evaluation using standard performance evaluation measures reveal that the word2vec using Quadratic Discriminant analysis-based classifier show optimal results.

Keywords—Artificial intelligence; sentiment analysis; machine learning; word embeddings; natural language programming

I. INTRODUCTION

Opinion mining or sentiment analysis on the other hand is a highly important subfield of NLP and is used as the umbrella term for studying sentiments, opinions, and emotions in text. Due to the drastic increase in use of social networking sites and the internet, the analysis of public opinion has gained importance for various commerce, policies and academia. Text analytics include creating categories depending on whether the text is positive, negative or neutral which can be important in understanding uptake among consumers or any specific segment or the public. For sentiment classification, Extended SVM, Naïve Bayes and Logistic regression were used traditionally; yet they are highly dependent on feature engineering and could not capture the depth of human language effectively [1].

In the last few years, deep learning techniques have brought dramatic improvements in SA, because they apply neural network structures that learn multi-level representations of text data from scratch [2]. RNNs, LSTM and CNN have been found to provide better resolution in extracting sequential correlations and contextual information within textual data [3]. In addition, the rise of pre-trained language models such as BERT and GPT has enhanced the performance of the sentiment analysis systems since transfer learning reduces the rate of overfitting as well as making models better at generalizing between datasets with limited labeled data based on [1].

Sentiment analysis is not limited to the monitoring of social media such as Facebook, Instagram, and twitter [4] but can be practiced in areas such as product review analysis, the customers feedback evaluation and even in the healthcare field. Since sentiment analysis is rapidly becoming a standard method for assessing public opinion and improving customer interaction for various organizations, the issue of effective and accurate detection becomes critical. Future areas are still hard and require better solutions to decode sarcasm, different meanings in different contexts, and domain specific language which are also quite important to cover in deep learning techniques [2]. Thus, continuous study is required to overcome these challenges and equally to extend the effectiveness of sentiment analysis to help explain the multifaceted human emotions captured in the content that is generated online.

In this research study, our aim is to carry out sentiment analysis from online content of social media by exploring the role of various textual features such Count Vectorizer, and Term Frequency – Inverse Document Frequency (TF-IDF). The features are used as input to machine learning classifiers such as CalibratedClassifierCV (CACV), PassiveAggressiveClassifier (PAC), and Quadratic Discriminant Analysis (QDA). We also try to explore various deep features like word2vec which focuses on considering context for given words in local and global perspective respectively. Here local means within a sentence or few words before or after the word while global means within the whole document. The results evaluation is carried out using standard performance metrics of accuracy, precision, recall and f-measures. This paper contributes to the field of AI by carrying out machine learning analysis of human feelings and emotions derived from textual data, while contributing toward the general understanding of the relationship between textual encoding and the analysis of human behavior. The main contributions of this research study include:

- Application of an advanced feature engineering approach such as count vectorization, TF-IDF and Word2Vec embeddings proved helpful in improving model sentiment analysis performance.
- Also, other machine learning models like Calibrated Classifier CV, Passive Aggressive Classifier, and Quadratic Discriminant Analysis were used to classify the sentiment labels.
- The best results were achieved employing Word2Vec embeddings with CACV proving that the use of embedding enhances the performance of the system.

The paper is organized as follows: Section II presents an analysis of some of the prior work done regarding sentiment classification and feature extraction methods. Section III then describes the data processing methodology of this study, feature engineering techniques including TF-IDF and Word2Vec and the classification models used in this study. Section IV summarizes the findings of this work and discusses the effectiveness of the embedding techniques. Section V provides the overall conclusion of the paper.

II. BACKGROUND

Therefore, over the recent past, the use of sentiment analysis has grown to be more important as the world has advanced in issues such as social media and content creation. The field has graduated from decision tree type of solutions to highly enhanced machine learning and deep learning type methods, which can now learn the tone of the text to whether it is happy, sad, angry or otherwise. The latest publications point to a revolutionary effect of generative AI for enhancing the efficacy and flexibility of SA for identifying consumer sentiment [5]. The use of sophisticated NLP ensures that there is a means of processing huge volumes of data generated by customers over social media as well as ensuring that the firms can indeed derive tangible benefits from these big data sources [6]. A recent research study [7] focused on features-oriented sentiment analysis which is also known as aspect-oriented sentiment analysis. This type of analysis mainly does not focus on document. Due to the growing era of technology, different methods of real time sentiment tracking have been enhanced to enable organizations to measure the flow of public sentiment. Programs such as Brand24 and Sprout Social employment machine learning [8] [9] to identify the sentiment of text in different and even emojis, thereby giving a fuller understanding of customer feelings [10]. Moreover, aspect-based sentiment analysis (ABSA) has become another important approach which is to identify certain characteristics of the product or service and allows companies to assess feedback from customers on specific characteristics, such as, for example, battery life or usability, separately [11]. Considering deep learning models, attention-based model have been used in a recent study which mainly proposes multi-channel gated recurrent RNN algorithms for aspect-based sentiment classification purpose. The work proof that the proposal of multi-channels in the existing RNN model [12].

As these methodologies are progressing there are still some issues arising in sentiment analysis because of factors like sarcasm, cross cultural differences and—regarding social media—frequent changes of language [13]. These challenges have been pointed out in recent literature reviews and the community has called for more research to enhance the reliability of sentiment analysis approaches [14]. Also, the market for sentiment analysis tools is expected to expand rapidly owing to the rising need for timely analysis of the customers' sentiments and behavior [15]. Aspect-based sentiment analysis, the model with good contextual information, namely, Attention-based Bidirectional LSTM

(BiLSTM) networks, is more suitable when it comes to fine-grained tasks [16]. Similarly, other recent works by [17] suggested a convolutional neural network and BiLSTM with attention mechanisms to deliver higher accuracy than conventional methods of sentiment classification in product reviews. Transformer based models have dramatically influenced approaches used in sentiment analysis. Subsequently, [18] used gradient boosting algorithms for the sentiment analysis tasks and to their finding, it outperformed other previous models for identifying the complicated sentiment patterns in large contextual data. For the sentiment classification in particular domains, for example, financial or health care domains, and have shown that the domain-wise improvement of the classifier performance is possible in this case that combine sentiment analysis with other NLP tasks [19].

Altogether, rhetoric trends dynamic, and new developments in tools and methods are expected for better application efficiency and higher predictive results of sentiment detection in different environments. This suggests that, as organizations continue to use these insights for strategic decision making [20], expanded research will be required to respond to the limitations of current methodologies and to investigate new employment contexts in this rapidly expanding domain. In this paper, instead of investigating and comparing traditional machine learning methods to sentiment analysis of textual data as previous studies have done, the current study employs advanced supervised learning models that Calibrated Classifier CV, Passive Aggressive Classifier, and Quadratic Discriminant Analysis (QDA) networks. These strategies are intended to improve the reliability and stability of sentiment predicting which was mentioned to be a weakness in the prior researches.

III. PROPOSED RESEARCH METHODOLOGY

The following sections provide the details of the methodological approach, as illustrated in Fig. 1, used in this sentiment analysis study, by following steps of data preprocessing, feature extraction, model training and experiment.

A. Data Preprocessing

Data cleaning is very important to ensure that preprocessing on data is well done and well checked before applying any machine learning. Initially, for noise removal, following elimination of special characters, URLs, any numbers, all the stop words, including 'is', 'the', etc. To minimize model bias, data entries with redundancy or duality were spotted and disregarded. After this, preprocessing undertaken to the text included conversion of text to lower case to eliminate redundancy concerning the sensitivity of the upper and lower 'cases. For more refinement, lemmatization was applied to stem words, where it uses the smallest root for a word to avoid any complexity in text data [21]. Lastly, to improve text vectorization in the next steps, each sentence was broken down to individual words (tokens). This ensures that data is preprocessed and ready for model training.

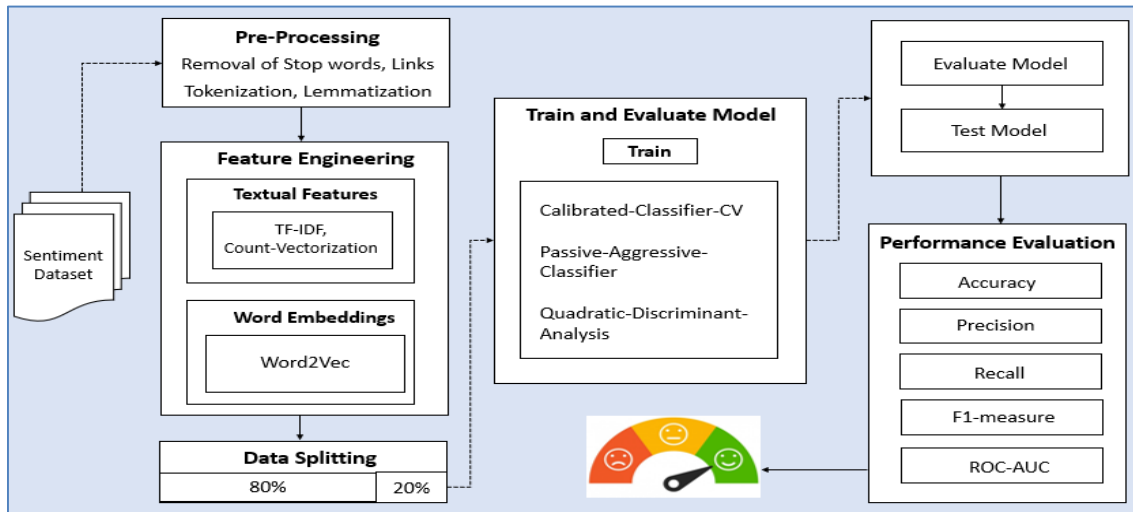


Fig. 1. Steps of proposed research methodology.

B. Feature Engineering

Feature engineering means getting the preprocessed textual data into forms directly understandable to the machine learning algorithms. In this study, three techniques were employed: Term frequency-Inverse document frequency (TF-IDF), Count Vectorization and Word2Vec. This paper provides the following overview of the mathematical basis and application of these methods.

1) *TFIDF*: TF-IDF refers to a technique of weighing words in a document against a corpus to determine the importance of the term in the document. It is the product of two components: Term Frequency (TF), and Inverse Document Frequency (IDF). To measure how much the term is exclusive or specific to a corresponding document. In this study, TF-IDF vectors were calculated, using Eq. (1) on the textual data and were sparse and of high dimensionality in representation of the documents [22]. Table I displays the description of symbols used in equations.

$$Document\ vector_d = [TF - IDF(t_1, d, D), \dots [TF - IDF(t_n, d, D)] \quad (1)$$

2) *Count vectorizer*: Frequency based vectoring or word-frequency vectorization derives a numerical value for each word based on the number of times the term appears in that document relative to a fixed list of terms. Every document is then converted to vector, whose elements are the vocabulary of the subjects and the values being the frequency of each of the terms used using Eq. (2). Although noncomplex, it does an excellent job of encoding the distribution of words in the dataset, which is represented as a sparse matrix for input into the machine learning algorithms.

$$Document\ vector_d = [x_1, x_2, \dots, x_n] \quad (2)$$

3) *Word2Vec representation*: Word2Vec gives dense words embedding in the continuum vector space to capture semantic relationship in between words, it learns word

embeddings from a large corpus. These embeddings capture context meaning to make the words that have similar contexts to have similar representations. For document-level representation, generally take the average of the word vector in applying machine learning models, so it is compact as well as semantically rich.

$$Document\ vector_d = \frac{1}{n} \sum_{i=1}^n Word2Vec(t_i) \quad (3)$$

C. Model Engineering

Algorithm selection, setting, and model training on the preprocessed dataset form the model engineering process based on three algorithms were employed: CalibratedClassifierCV, PassiveAggressiveClassifier, and Quadratic Discriminant Analysis (QDA). Equations defining each and principles which underline each are provided in the following:

CalibratedClassifierCV (CACV) is a meta-algorithm aimed towards increasing the accuracy of a base classifier when using probability estimates for decision making. It functions by using the raw outputs of the classifier, to which logistic regression model ensures to the raw outputs of the classifier, that ensures a monotonic relationship between probabilities and true outcomes, computed as in Eq. (4). This algorithm comes very handy especially when the base classifier gives unformatted probabilities or raw scores.

$$p(y = 1|x) = \frac{1}{1 + \exp(-a \cdot f(x) + b)} \quad (4)$$

PassiveAggressiveClassifier (PAC) is an online learning algorithm which is suitable for scaling and efficient classification paradigm. It adapts its model weights only when predictions are wrong or the decision margin is less than specified, which makes it reactive “aggressively or slightly” to mistakes. The model optimizes a hinge loss function that has been well applied in binary and multi-class classification and supports linear kernel-based learning, computed objective function as in Eq. (5). The model is especially useful for the cases of working with high dimensions and big data, like text classification tasks, at which it balances the speed of adaptation to new data and necessary computational resources.

$$L(w, x, y) = \max(0, 1 - y(w \cdot x)) \quad (5)$$

The model updates w iteratively as in Eq. (6):

$$w_{t+1} = w_t + \tau yx \quad (6)$$

Where $\tau = \frac{1-(w_t \cdot x)}{\|x\|^2}$ is the learning rate to ensure convergence while remaining sample of passive for correctly classified.

Quadratic Discriminant Analysis (QDA) is another generative classification algorithm, which implies that the model assumes features are normally distributed within the classes. It is an extension to Linear Discriminant Analysis (LDA) where covariances within each class may differ and therefore produces quadratic decision boundaries. The current implementation of QDA is based on the Bayes' theorem, where the likelihood of each class is the multivariate Gaussian probability density, as in Eq. (7). Unlike other machine learning algorithms which may not be well suited when dealing with non-linear feature-class space. In general, QDA is more complex than LDA, but it is more flexible; thus, it is preferable when the classes have different variance.

$$Q(d = k|g) = \frac{Q(g|d=k)Q(d=k)}{Q(g)} \quad (7)$$

This is subjected to Eq. (8):

$$Q(g|d = k) = \frac{1}{(2\lambda)^{e/2} |\Sigma_k|^{1/2}} \exp\left(-\frac{1}{2} (g - \mu_k)^T \Sigma_k^{-1} (g - \mu_k)\right) \quad (8)$$

The decision boundary for QDA is quadratic, computed as (9).

$$\delta_k(g) = -\frac{1}{2} \ln |\Sigma_k| - \frac{1}{2} (g - \mu_k)^T \Sigma_k^{-1} (g - \mu_k) + \ln Q(d = k) \quad (9)$$

Class predictions are calculated by maximizing the posterior probability using Eq. (10).

$$\hat{d} = \arg. \max_k Q(d = k|g) \quad (10)$$

D. Dataset

This study aims at the creation of a sentiment analysis system specializing in the analysis of emotional and opinionated posts in social media. Social media is quite popular and creates large textual data daily with useful knowledge of the public's perception of products, services, events, and social issues. This system is expected to utilize NLP tools to identify and sort sentiments of bilateral content, for example, positive sentiment or negative sentiment or even no sentiment. The dataset employed in this study is collected from open platform Kaggle, sourced from authentic social media data comprising of different styles of writing and different contexts such as brand tracking, in a crisis, for opinion mining and social trend analysis.

TABLE I. DESCRIPTION OF SYMBOLS USING IN EQUATIONS

Symbols	Description
t	Term in a document
d	Document
n	Total number of terms in a document
x	Frequency based on each word

$f(x)$	Raw output of the base classifier
a and b	Parameters optimized via logistics regression.
w, x	Weight and feature vector
y	True label (+1 or -1)
τ	Learning rate
μ_k	Mean vector of class k
\sum_k	Covariance matrix of class k
e	Dimensionality of feature space
TP, TN	True Positive and Negative
FP, FN	False Positive and Negative

E. Evaluation Measures

Measures of performance evaluation are metrics used in machine learning that provide a way of qualifying the several aspects of the model's predictions, as shown in Table II. Accuracy tends to give a broad view of the accurateness of the model since it quantifies the actual number of properly classified samples to the overall number of samples. Recall measures the ratio of true positives among all actual positive observations, or the ability to avoid false negative predictions. Recall (Sensitivity) shows how many actual positives were correctly identified, which focuses on minimizing the number of negative cases that are positive. F1-Score, this metric is the harmonic meaning between Precision and Recall, which is better when used when the distribution is uneven.

TABLE II. EQUATIONS OF PERFORMANCE MEASURES

Metrics	Equation
Accuracy	$\frac{TP+TN}{TF+FN+FP+TP}$
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
F1-score	$\frac{2(Precision \cdot Recall)}{Precision+Recall}$

AUC-ROC means Area Under the Receiver Operating Characteristic Curve, and it measures the model's conditional probability of correctly identifying a negative case using all the thresholds. Hence, these four metrics offer an uninterrupted way of evaluating the performance of the model such that the model's reliability and efficiency would be achieved.

IV. RESULTS

The results of the sentiment analysis experiments using three models, based on three different feature extraction techniques including TF-IDF, Count Vectorizer, and Word2Vec are summarized through confusion matrices and corresponding performance metrics. These results, as displayed in Table III help in directing focus to the model's strength and weakness aspect of correctly predicting sentiment labels namely negative, neutral and positive. The exploratory data analysis (EDA) visualizations summarize key textual patterns within the dataset:

A. Distribution of Text Length

From this histogram as shown in Fig. 2 (a), this graph describes the frequency of texts within the data set according to their length. The frequency distribution shows most texts are of lengths between 10 and 40 Words, although as text length increases, the number of texts decreases. We see that

distribution is right-skewed, which means that there are more texts of shorter length than texts of very long length. This bar chart in (b) shows 10 most frequently appearing words in the dataset and the frequency of each of these words. Among these, “I’m”, “day”, “like”, “know” are few of the most frequent words used in day-to-day conversation. These often-used words seem to indicate that the dataset samples a daily or personal interaction-oriented environment.

B. Word Frequency Distribution for Words with Frequency >10

The bar chart shown in (c), will give a detailed analysis of the words that appear more than 10 times in the dataset. Specifically, words that can be found in the list of 10 most frequent words like the “I’m,” “like,” and “know” are in the middle. Like ‘Interests’, ‘Excitement’, some extra word like ‘amazing’, ‘day’, ‘today’, ‘tomorrow’ seems to point toward sentiment –rich contexts or likely sentiment temporal relatedness in the data set.

C. Distribution of Labels

In Fig. 3 displaying bar chart (a) capturing the current distribution of labels in sentiment analysis. Overall, the data split over a broad range with the sentiment of neutral prevailing

over positive and negative sentiments, though with decreased number. The negative and positive sentiments are similar in the number of corresponding features, and they are between 125-175; however, the most prominently observed sentiments are the neutral ones with over 200 samples.

D. Word Cloud:

Fig. 3 preview (b) of the most often occurring words in the data set as a whole. The word cloud illustrates the frequency of words most often repeated in the dataset; it includes words such as love, going, day, I’m, know, among others. This shows the word frequency in sample texts, with possible positive words for the choice of ‘love’ and ‘day’ against possible negative words ‘don’t’ and ‘can’t’.

This analysis underlines the fact that the dataset has conversational and sentiment-related properties, has more short texts, and uses more often and more frequently the most common sentiment-related words. It is useful in understanding the structure and contents of the text, therefore assists in the preprocessing and feature extraction steps that may be followed in other downstream tasks such as sentiment analysis attempting to build methodologies for classification models based on textual characteristics.

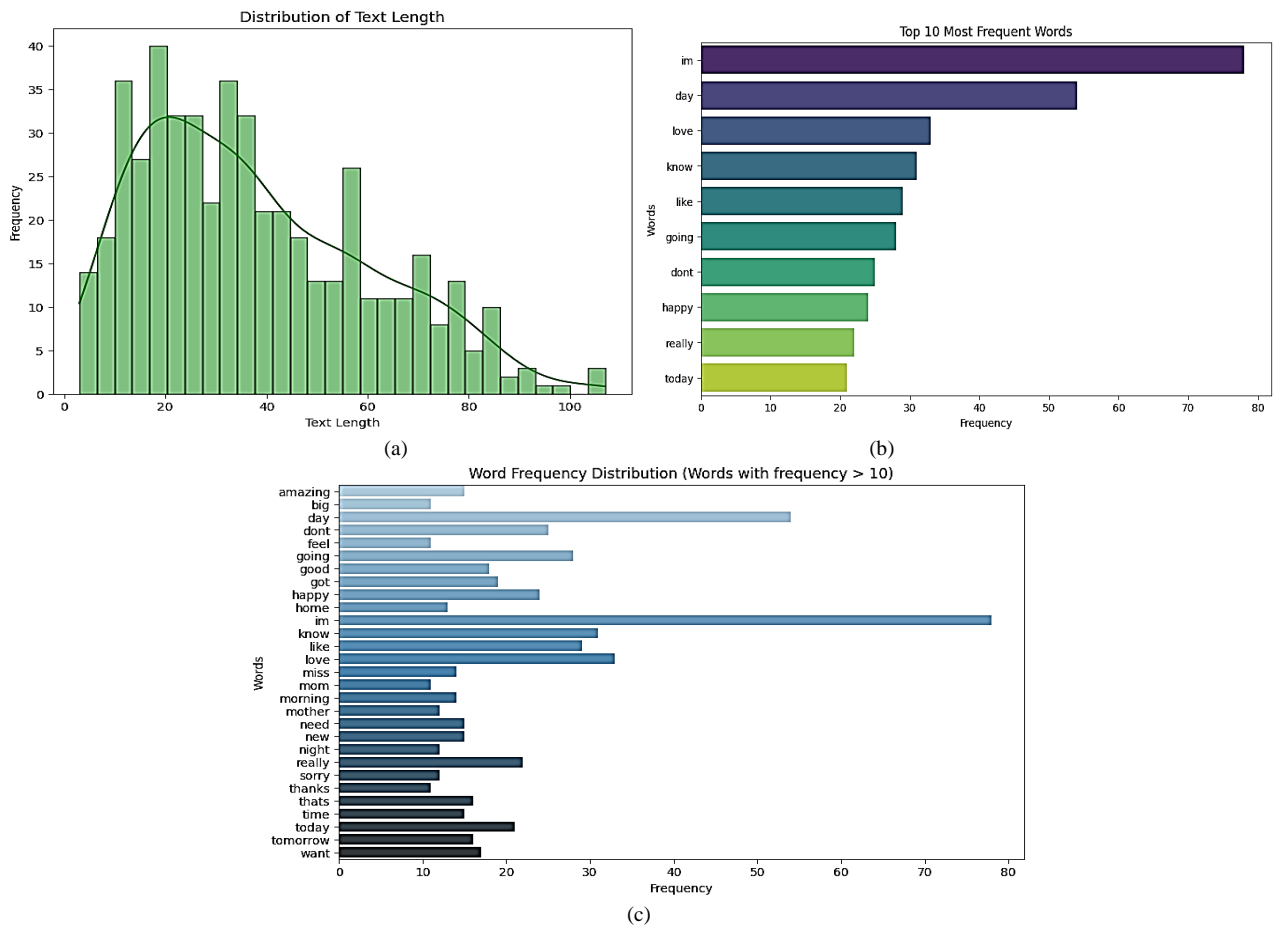


Fig. 2. Analysis of dataset text length along with frequency.

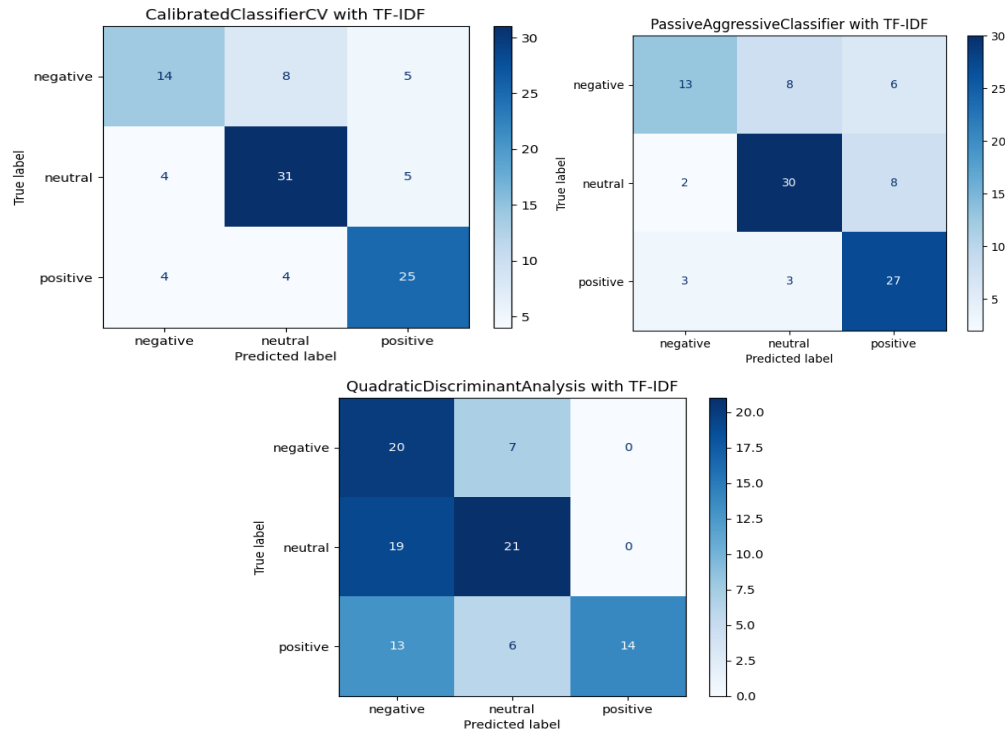


Fig. 4. Confusion Matrix of model performance using TF-IDF features.

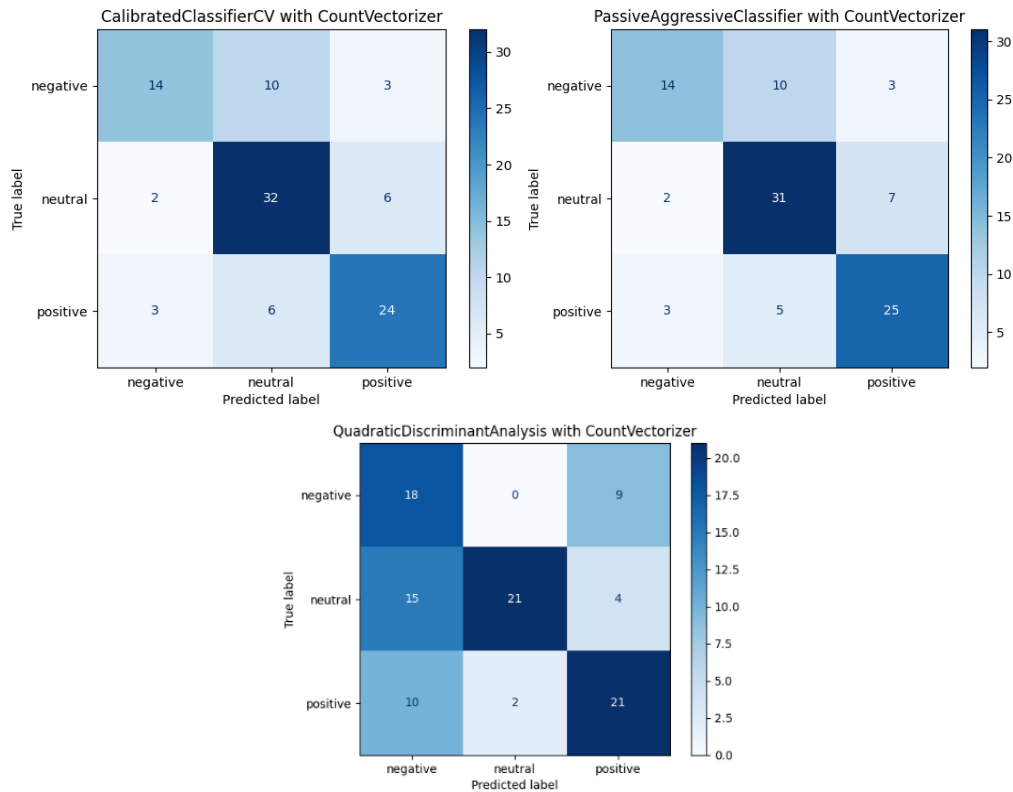


Fig. 5. Confusion matrix of model performance using count vectorizer features.

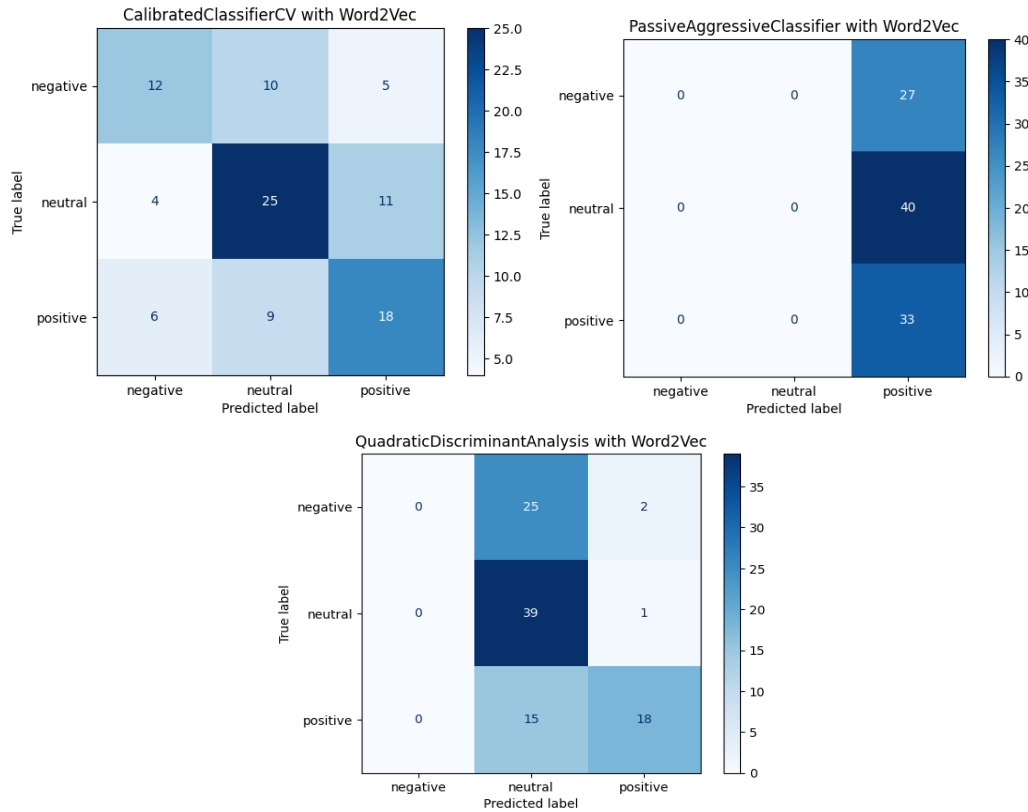
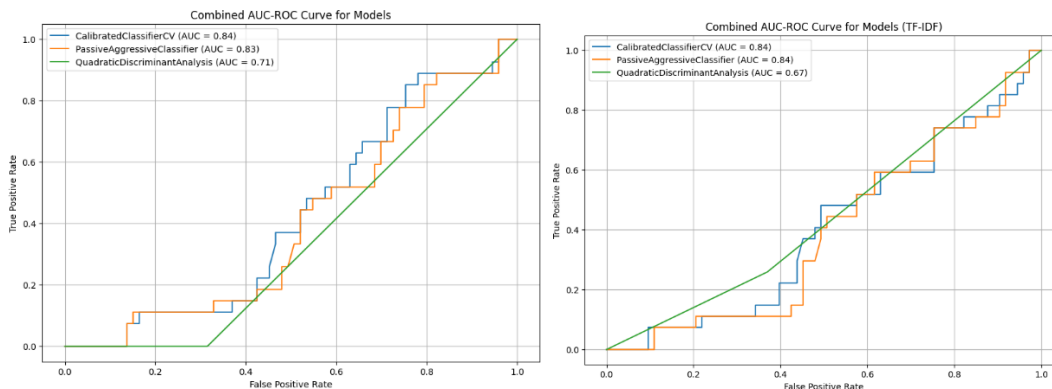


Fig. 6. Confusion matrix of model performance using Word2Vec features.

From the presented ROC curves, in Fig. 7, infer the overall classification performance of the models along with the examined feature extraction methods. For TF-IDF and Count Vectorizer, PAC and CACV have AUC of 0.76 – 0.81, which means the models are good in scenarios in which the separation between classes is clear; however, QDA has a problem with AUC values 0.66 – 0.72, suggesting a lower ability to distinguish between classes is lower. However, Word2Vec considerably enhances effectiveness; the best outcome is given by CACV (AUC = 0.73), followed by QDA (AUC = 0.66).

These curves show that even though few classifiers like PAC works well with traditional features such as TF-IDF, the embedding techniques like Word2Vec yielded a better class separation in most of the classifiers as supported by higher AUC scores throughout most of the curves. Moreover, the ROC diagrams show that some models are more efficient in terms of true positive and false positive rates which is evident when comparing Word2Vec representations showing that feature embeddings affect the classification performance.



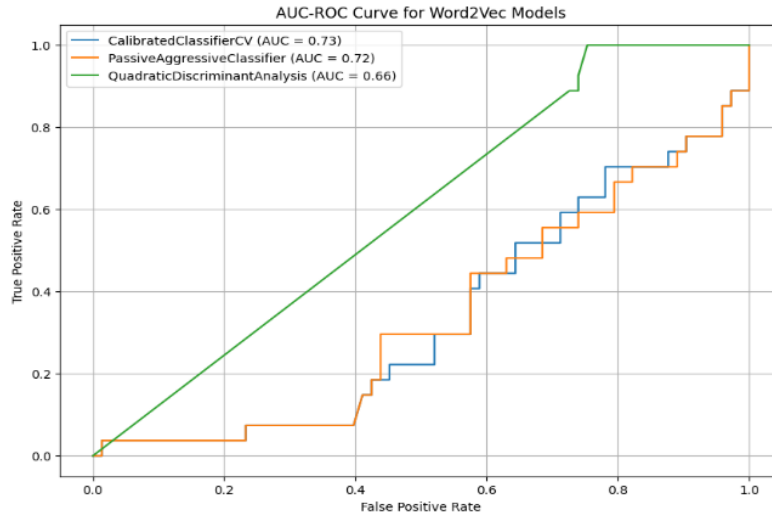


Fig. 7. Analysis of combined AUC-ROC model performance using features.

H. Discussion

To sum up, different feature extraction techniques have been effective in model performances of different degrees. Overall, Word2Vec was the best performer while QDA achieved the best accuracy of 80% and the best confusion matrix, further illustrating superior ability to deal with semantic resemblance for textual data, showing comparative analysis in term of accuracy measures illustrated as in Fig. 8. TF was more useful for PAC, and it scored the highest accuracy of 76% within its features, although it lower in

performance with Word2Vec. Count Vectorization was moderate in its performance, in this field QDA was the most effective out of all (74% accuracy), however it didn't come close to the Word2Vec results. These trends therefore show how features interact with model type and suggest that advanced method such as Word2Vec work best for sentiment analysis models because they are best suited for capturing context and relation between words. These results suggest that extraction methods and classifiers employ significant roles in sentiment analysis.

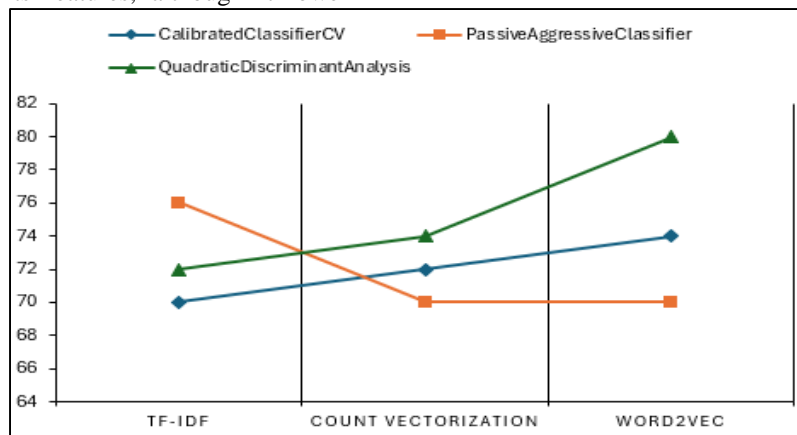


Fig. 8. Comparative analysis of accuracy measure along with models.

The contrast of the suggested model with other similar researches for sentiment prediction from textual information shows enhanced performances, as display in table IV. The current approaches, such as NB with 73%, RF with 74%, and SVC at 71%, have been computed in Twitter and IMBD datasets. On the other hand, the suggested Quadratic Discriminant Classifier in the Kaggle Social Media dataset just attained an 80% of accuracy. This considerable improvement demonstrates that the idea of the proposed model can learn various sentiment patterns, proving that the proposed model is capable of being a more feasible solution for the SA task as compared to the traditional machine learning models.

TABLE IV. COMPARISON WITH EXISTING STUDIES

Ref	Year	Models	Dataset	Results (%)
[18]	2020	RF	IMBD	74
[22]	2021	NB	Twitter	73
[23]	2023	SVC	Twitter	71
Proposed	2025	QDC	Social Media	80

V. CONCLUSION

Social media has become a digital world for users to share their opinions, views and interact through posts, messages, comments, and reviews, making it central mode for interpreting public sentiment. The role of AI in sentiment analysis is growing significantly, as it allows automated detection of emotions and opinions from UGC. In this study, we examined three feature extraction methods combined with various advanced AI models for sentiment classification. Among the models, QDA with Word2Vec embeddings achieved the highest accuracy of 80%, demonstrating its superior ability to capture semantic relationships and patterns in text. These findings show the effectiveness of integrating advanced feature representations with appropriate classifiers. Although the research study is helpful for predicting the sentiment analysis from online content, the limitation of the study that these findings are limited to textual data only. Future work highlights the significant exploration of deeper neural architectures and larger datasets to boost performance further. This study provides a gateway for developing more robust AI-driven tools for sentiment analysis, contributing to better understanding and leveraging user opinions in diverse applications.

REFERENCES

- [1] Zhang, L., Wang, S. and Liu, B., 2018. Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), p.e1253.
- [2] Iqbal, S., Khan, F., Khan, H.U., Iqbal, T. and Shah, J.H., 2022. Sentiment analysis of social media content in pashto language using deep learning algorithms. *Journal of Internet Technology*, 23(7), pp.1669-1677.
- [3] Suryawanshi, N.S., 2024. Sentiment analysis with machine learning and deep learning: A survey of techniques and applications. *International Journal of Science and Research Archive*, 12(2), pp.005-015.
- [4] Mahmood, A., Khan, H.U. and Ramzan, M., 2020. On modelling for bias-aware sentiment analysis and its impact in Twitter. *Journal of Web Engineering*, 19(1), pp.1-27.
- [5] Krugmann, J.O. and Hartmann, J., 2024. Sentiment Analysis in the Age of Generative AI. *Customer Needs and Solutions*, 11(1), p.3.
- [6] Convin.ai. (2024). Top Sentiment Analysis Tools to Watch in 2024 And Further. Retrieved from <https://convin.ai/blog/sentiment-analysis-tools-2024>.
- [7] Ahmad, W., Khan, H.U., Iqbal, T. and Iqbal, S., 2023. Attention-based multi-channel gated recurrent neural networks: a novel feature-centric approach for aspect-based sentiment classification. *IEEE Access*, 11, pp.54408-54427.
- [8] Kebede, D. and Tesfai, N., 2023. Ai-powered Text Analysis Tool for Sentiment Analysis.
- [9] Al-Otaibi, S.T. and Al-Rasheed, A.A., 2022. A review and comparative analysis of sentiment analysis techniques. *Informatica*, 46(6).
- [10] Ahmad, M., Aftab, S., Muhammad, S.S. and Ahmad, S., 2017. Machine learning techniques for sentiment analysis: A review. *Int. J. Multidiscip. Sci. Eng*, 8(3), p.27.
- [11] Eliot, L., 2020. Legal Sentiment Analysis and Opinion Mining (LSAOM): Assimilating Advances in Autonomous AI Legal Reasoning. *arXiv preprint arXiv:2010.02726*.
- [12] Ahmad, W., Khan, H.U., Iqbal, T., Khan, M.A., Tariq, U. and Cha, J.H., 2023. Hybrid multichannel-based deep models using deep features for feature-oriented sentiment analysis. *Sustainability*, 15(9), p.7213.
- [13] Ishfaq, U., Khan, H.U. and Iqbal, K., 2016. April. Modeling to find the top bloggers using sentiment features. In 2016 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube) (pp. 227-233). *IEEE*.
- [14] Mao, Y., Liu, Q. and Zhang, Y., 2024. Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University-Computer and Information Sciences*, p.102048.
- [15] Chakriswaran, P., Vincent, D.R., Srinivasan, K., Sharma, V., Chang, C.Y. and Reina, D.G., 2019. Emotion AI-driven sentiment analysis: A survey, future research directions, and open issues. *Applied Sciences*, 9(24), p.5462.
- [16] Wang, Y., Huang, M., Zhu, X. and Zhao, L., 2016, November. Attention-based LSTM for aspect-level sentiment classification. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 606-615).
- [17] Zhang, J., Liu, F.A., Xu, W. and Yu, H., 2019. Feature fusion text classification model combining CNN and BiGRU with multi-attention mechanism. *Future Internet*, 11(11), p.237.
- [18] Neelakandan, S. and Paulraj, D., 2020. A gradient boosted decision tree-based sentiment classification of twitter data. *International Journal of Wavelets, Multiresolution and Information Processing*, 18(04), p.2050027.
- [19] Ishfaq, U., Khan, H.U. and Iqbal, K., 2017. Identifying the influential bloggers: a modular approach based on sentiment analysis. *Journal of Web Engineering*, pp.505-523.
- [20] Naz, A., Khan, H.U., Alesawi, S., Abouola, O.I., Daud, A. and Ramzan, M., 2024. AI Knows You: Deep Learning Model for Prediction of Extroversion Personality Trait. *IEEE Access*.
- [21] Alsini, R., Naz, A., Khan, H.U., Bukhari, A., Daud, A. and Ramzan, M., 2024. Using deep learning and word embeddings for predicting human agreeableness behavior. *Scientific Reports*, 14(1), p.29875.
- [22] Qi, Y. and Shabrina, Z., 2023. Sentiment analysis using Twitter data: a comparative application of lexicon-and machine-learning-based approach. *Social Network Analysis and Mining*, 13(1), p.31.
- [23] Yadav, N., Kudale, O., Rao, A., Gupta, S. and Shitole, A., 2021. Twitter sentiment analysis using supervised machine learning. In *Intelligent data communication technologies and internet of things: Proceedings of ICICI 2020* (pp. 631-642). Springer Singapore.

An AI-Driven Approach for Advancing English Learning in Educational Information Systems Using Machine Learning

Xue Peng*, Yue Wang

Tourism College, Xinxiang Vocational and Technical College, Xinxiang 453000, Henan, China

Abstract—In current era of globalization, English language learning is important as it has become a global language and helps people to communicate from various regions and languages. For vocational students whose main aim is to get skills and get employed, learning English for communication is important. We here present a proposed framework for learning English language which can become a foundation for a complete Artificial Intelligence (AI) based system for help and guidance to the educators. This study explores the use of diverse Natural Language Processing (NLP) techniques to predict various grammatical aspects of English language content especially focused on tense prediction which lay the foundation of English content. Textual features of Bag of words (BoW) which considers each word as a separate token and Term Frequency –Inverse Document Frequency (TF-IDF) are explored. For both diverse features, the shallow machine learning models of Support Vector Machine (SVM) and Multinomial Naïve Bayes are applied. Moreover, the ensemble models based on Bagging and Calibrated are applied. The results reveal that BoW model input for SVM and Bagging technique using TF-IDF shows optimal results with high accuracy of 90% and 89% respectively. This empirical analysis confirms that such models can be integrated with web or android based systems which can be helpful for learners of English language.

Keywords—Artificial intelligence; information system; machine learning; English language learning; natural language processing

I. INTRODUCTION

The importance of the English language in today's globalized world is evident it serves as a universal medium of communication that connects people from diverse linguistic backgrounds. Fluency in English is essential for considering a broad range of information to be accessed in many ways, especially, in the institutions where many publications are in English [1]. Linguistic influence helps organize joint work and exchange results in terms of continuing the advancement of knowledge by researchers and scholars all over the world [2]. For non-native speakers, especially students in vocational colleges, the journey of learning English can be challenging, particularly when it comes to mastering grammar and sentence structure. Tenses, a core element of English grammar, pose a significant difficulty for learners, as they are essential for expressing actions in different time frames [3]. As education systems strive to provide more effective and efficient language learning, leveraging advanced technologies to address these challenges has become increasingly important. English Learning is the ability to master the capacity to write, read,

and comprehend the English language as well as knowledge of grammar, vocabulary, and pronunciation [4]. This learning may take place at school or university, in community classes, or through independent learning with the help of internet materials [5]. The use of English in global dimension has therefore become mandatory, and it entitles its users use in a multiplicity of sectors. In addition, cross-sectional study will ensure that the sample group is more heterogeneous and diverse regarding their grade level and learning which [6]. The importance of English learning goes beyond the need to communicate, English learning is key for success in academics and career. Knowledge in English breaks barriers and provide a drive to access information and materials since most information is found in English [7]. Many academic institutions particularly require English proficient skills to be used in reading of enhanced texts, discussions and the production of research. In addition, in today's business environments, English is often a requirement for job and promotions because of the effectiveness of the ability to facilitate cooperation with people of different teams in various countries [8]. In conclusion, learning English is not only about language mastery but is equally about improving the methods and the ways through which one can interact with the world.

The applications of AI in education, specifically in language learning, are vast and transformative. AI-powered tools can personalize learning experiences by adapting to individual learner needs, providing immediate feedback, and automating tedious tasks such as grading or content delivery [9]. In the domain of English language learning, AI can assist learners in improving their grammar, vocabulary, pronunciation, and understanding of complex linguistic concepts like tenses [10]. This personalization not only speeds up the learning process but also ensures that learners receive targeted support, thereby enhancing their chances of mastering English more efficiently [11]. Moreover, AI can bridge the gap between traditional classroom instruction and students who may not have access to high-quality educational resources, thus promoting equitable learning opportunities. The use of Education information systems in the teaching and learning of English language has emerged essential in the improvement of the teaching and learning process through use of Information and Communication Technologies (ICT) [12]. These systems provide interface to multiple learning material in form of e-books, online courses, and multimedia which enhance the reading, writing, listening and speaking ability of a learner [13] posit that the use of ICT in English Language Teaching (ELT)

increases learners' attentiveness as well as encourages learner interaction with the knowledge resources and among themselves. In addition, the teachers have also described that such systems help them to give the students individual feedback and sublime courses to fit the students' needs and wants about learning [14]. By integrating these systems with the use of Recurrent Neural Networks (RNNs) and item response theory, learner's input is then subjected to the detection of context and grammatical rules used in defining the correct tenses to be used from a pool of verbs [15]. In addition to error identification, it is effective in creating practice exercises based on the learner's achievement level fostering individual learning development programs. Environmental tense predictors of language exercises that adapt automatically or provide instant feedback for right and wrong contribute to the exciting and more efficient learning modes to help learners enhance their grammatical accuracy and fluency of English language usage [16].

In this research work, we work on teaching and improving English language skills of students studying in vocational colleges as their main concern is to learn skills so that they are readily be available to the market as skilled sources. So, we have worked on basic English language skills and the use of tense which is the pivotal concept in English Grammar in English language learning by using two main AI-based technique of machine learning and state-of-the-art ensemble models including Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), Bagging Classifier and CalibratedClassifierCV integrated with textual features of Bag-of-words (BoW), and Term Frequency-Inverse Document Frequency (TF-IDF), achieving highest accuracy of 90% with BoW feature when coupled with ensemble models. Further these results are evaluated using standard performance measures of accuracy, precision, recall and f1-score, showing a pathway for exploring more advanced learning abilities for vocational college students. Furthermore, contributions of this study are as follows:

- **Effective Feature Selection:** Used BoW and IDF features in tense prediction of tense on the English contents with the following results expressing suitability of BoW and TF-IDF for tense analysis as is relevant in classification of content.
- **Comparative Model Analysis:** Performed a comparison of BoW and RW performances between SVM & Multinomial Naïve Bayes; showed that SVM with BoW was 90% accurate and has outperformed other models in terms of grammar, especially tenses.
- **Implementation of Ensemble Techniques:** That incorporated ensemble methods including Bagging and Calibrated classifiers obtained 89% accuracy with Bagging and TF-IDF convenience features. As ensemble models enhanced the stability and the accuracy of their predictions.

The rest of the paper organization as follows: Section II presents the background knowledge of relevant literature in field of English language learning. Section III shows the comprehensive details of applied methodology including experimental setup. Section IV shares the analysis of results along with discussion. Section V provides the summary of paper in conclusion form with future directions.

II. LITERATURE REVIEW

The use of AI has become prominent in English Language Learning (ELL) processes as a strategy within educational information systems, improving the educational process and learners' interactions. In study [17], systematic review shows the positive and varied effects of AI technologies like ITS, NLP, and Speech Recognition on ELL. The review also notes that ITS seem to help learners the most around language proficiency or specific language skills, while speech recognition technology helps learners to improve pronunciation and speaking skills and, therefore, boosts their confidence. In addition, [18] embracing of Virtual Reality (VR) as well as Augmented Reality (AR) has changed the whole environment in which learners practice English. Another study [19] argues that these technologies foster contextually grounded environments which not only support the development of register specific language but also provide cultural context in which learners can apply the language. Appointment simulation enables learner to invariably assess their speaking performance in a supposedly realistic context, which is vital for language learning. Besides improving language skills [20], AI approaches help optimize and minimize the administrative work within the context of educational information systems. For instance, instruments like ChatGPT and education copilot help in developing courses and lesson plans, which gives such educators more time to deal with interactions and individual engaging with the student.

This automation not only decreases the possibility of grading bias but also allows instructors to give feedback promptly; instructional decisions reach a state in which it is based on the performance analytics data in real-time [21]. Still, there are obstacles in using AI in ELT because the incorporation of AI into teaching practice encounters certain difficulties. A study carried out by the Teaching English showed that schools around the world implement AI in their classrooms [22]. Similarly, concerns regarding the drawbacks of AI such as, in language use there could be bias and in terms of learning human interaction could be reduced when using the technologies [8]. Therefore, further research must reveal the effectiveness of using AI technology for learning English in the long term as well the existence of frameworks that can help educators work through its difficulties. In aggregate, the implementation of AI-inspired technologies into the scope of educational information systems is a breakthrough in learning English [11]. Introducing these systems in the classroom environments may also open the possibilities of improving the extent of teaching and learning capabilities, in the form of personalized learning experiences, coupled with the use of, for instance, immersive technologies and efficient administrative tasks. Nevertheless, the issues related to effective AI application are going to remain critical to achieve the potential benefits of utilizing AI in language instruction.

Modern progress in the areas of AI and ML have hugely impacted on improving the techniques of learning English. The implementation of some deep learning models like the CNNs and the RNNs in automating the essay grading task and appended feedbacks to the English language learners [23]. Another study, contributed to AI voice recognition in enhancing the precision on English pronunciation to help learners. Subsequent research has also examined the teaching of grammar

by using AI devices to improve the learning of English grammar since the difficulty level of the material is determined by the learner's performance [24]. Authors in study [25] developed an error prediction system they say helps in teaching learners how to write by pointing out any mistake they have made and the corrective action that needs to be taken thus enhancing the learners' writing skills to carry out an enhanced learning process for writing by offering the writers feedback on grammar, vocabulary and writing style. English listening comprehension has also been a focus of machine learning, perhaps as exemplified by study [26] who posited an assessment system, incorporating the use of AI to gauge the learners' listening skills based on their proficiency level for developing applications that offer differentiated vocabulary learning according to learners' performances. Furthermore, in study [27] employed RNN for real time error correction in English as foreign language, which provided feedback for learners as soon as they wrote incorrect sentences and paragraphs so that they could correct their grammatical and writing mistakes. Lastly, NLP and machine learning to teach English syntax and semantics to learners

establishing a strong foundation for learners to understand the complicated language rules [28]. These studies clearly show that English learning and teaching is the area where AI and ML are widely used as they can personalize the learning courses, develop the criterion reference assessment, and enhance the learners' skills within the different domain of the language.

III. METHODOLOGY

The following section provides a detailed procedure for specifying and categorizing understanding in the English grammar to improve teaching and learning experiences for the teachers and their learners. Concisely, the steps were carried out as Data Preprocessing, Feature Extraction methods, several Machine Learning Models, and Performance Measures have been employed as a basis for assessing the efficacy of the presented models. The, following structured approach, as shown in Fig. 1, guarantees a strong structure that responsibility helps to work out the difficulties of learning identification in English grammar.

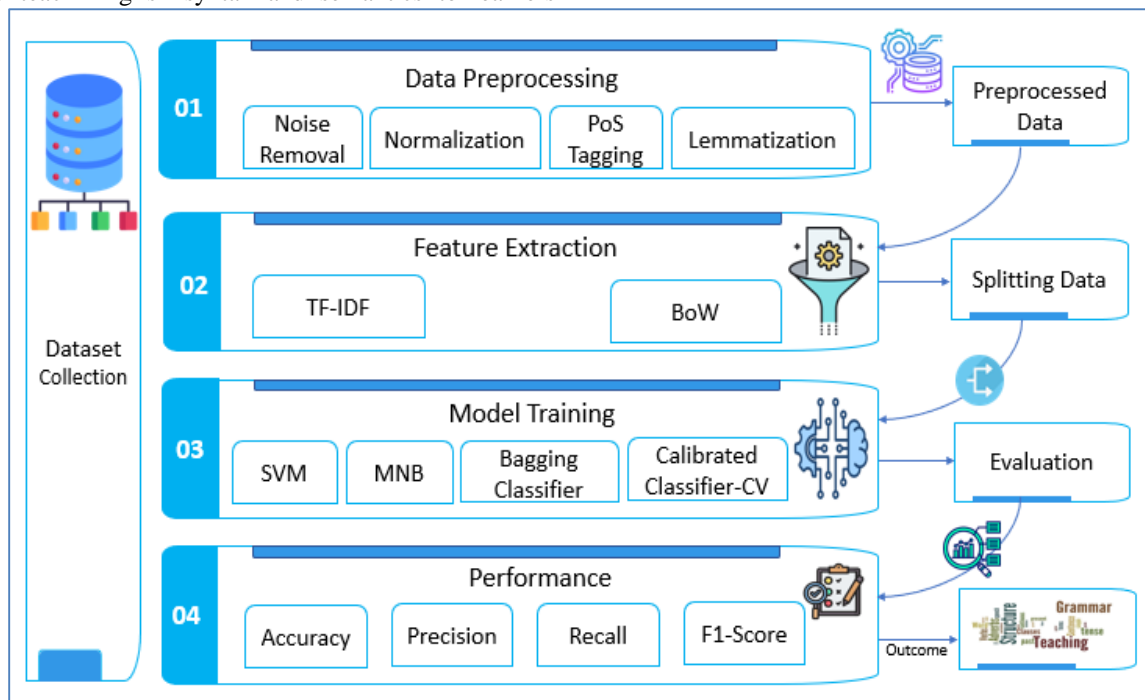


Fig. 1. The framework showing steps of the research study.

A. Data Collection and Preparation

This data set will be useful for English learners and teachers to learn and improve the usage of language in English. It includes example sentences and the tense that each of such sentences demonstrates. The data was then constructed carefully to ensure it produced sentences that clearly illustrated how the various English tenses can be used making it a rich resource for use in learning and teaching. Text preprocessing is an important step to be taken to transform textual data into NLP context. In this study, we identified several main processing procedures intended to improve the quality of the input data that is then provided to ML algorithms. There is the removal of stop-words and punctuation mark which is the first process in text mining in the process of filtering out noise words within the large datasets

so that models can learn from more important words. After this, lemmatization also applied that includes the removal of prefixes and suffixes of the words and bring them to their basic form, making it easier to normalize different forms. This is particularly important for tense identification as this means that different forms of any given verb will be treated in the same way. Also, transforming text into a standard form to remove more variation from the data. However, it also employed Part-of-Speech tagging which aims at assigning a role to each word it has identified as a noun, verb, adjective and the likes. This is important especially for recognition of tenses, because verbs are of key importance in tense definition.

B. Feature Extraction

Feature engineering on the other hand is the process of extracting more meaningful features from the preprocessed text that can then be used in the actual machine learning processes. In this research, employing two primary techniques: Understanding of Term Frequency-Inverse Document Frequency (TF-IDF), and Bag of Words (BoW). Table I shows the in-depth definition of symbols used in equations.

1) *Term frequency-inverse document frequency (TF-IDF)*: In its most simplified form, the Term Frequency (TF) can be, nevertheless, enhanced with normalization for document length differences. Normalized term frequency as is expressed by Eq. (1) can be defined as the normalized term frequency.

$$TF(t, d) = \frac{f_{t,d}}{\sum_{t \in d} f_{t,d}} \cdot (1 + \log\left(\frac{f_{t,d}}{\max_{\bar{t},d} f_{\bar{t},d}} + 1\right)) \quad (1)$$

Combining these advanced formulations, the TF-IDF score for a term t in document d relative to corpus D can be computed as in Eq. (2).

$$TFIDF(t, d, D) = TF(t, d) \cdot IDF(t, D) \quad (2)$$

Moreover, it is possible to explain the TF-IDF values based on the information gain of the terms with respect to the documents as computed using Eq. (3).

$$TFIDF(t, d, D) = \left(\frac{f_{t,d}}{\sum_{t \in d} f_{t,d}} \cdot \left(1 + \log\left(\frac{f_{t,d}}{\max_{\bar{t},d} f_{\bar{t},d}} + 1\right) \right) \right) \cdot \left(\log\left(\frac{N+1}{n_t+1}\right) \right) \quad (3)$$

The mutual probability distribution can be expressed as in Eq. (4).

$$M(t; d) = \wp(t|d) \cdot \wp(d) \cdot IDF(t) \quad (4)$$

This formulation captures what we have been aiming at, in this paper, that is, capturing as to how informative each term is regarding the associated documents to be able to have an even better understanding of them within the given corpus.

2) *Bag of Words (BoW)*: On the other hand, the Bag of Words utilized to reduce text data to a series of words while ignoring the position of words in the document and retains multiplicity. This makes word counting simple which is especially useful when trying to determine the commonly used words and particular, verb forms relating to varying tenses, by defining a weighted frequency representation that incorporates not only raw counts but also contextual importance through various normalization techniques, defined as in equation 5. In utilizing these feature engineering strategies to make a stronger representation of the text data would be formed, which will enable the identification of the appropriate tenses in learning of the English grammar.

$$V_d = [w_{1,d}, w_{2,d}, \dots, w_{n,d}] \quad (5)$$

Where $w_{i,d}$ is defined as in Eq. (6):

$$w_{i,d} = f_{i,d} \cdot \text{norm}(f_{i,d}) \cdot \text{context}(t_i, d) \quad (6)$$

The procedures presented in this research for establishing an approach for constructing an automated environment for using AI to support vocational educators and learners in enhancing their mastery of English grammar and tense identification.

C. Applied Models

In the realm of NLP in dealing with issues common with English learning identification and classification. This section investigates a few more complex forms of machine learning models and how they are implemented in different ways including models like Support Vector Machines (SVM), Multinomial Naive Bayes (MNB), Bagging Classifier, and Calibrated Classifier CV. Contrasting the principles and uses of these models, to clearer understanding of how these models can be used to improve educational outcomes in English learning for both teachers and learners will be gained.

1) *Support vector machine (SVM)*: SVM is a type of used learning method that is applied for classification problems. It does this by identifying the best hyperplane that can best separate different classes of data in a very large dimensional space. SVM works well in high-dimensional space and is not sensitive to the problem of overfitting, especially when the number of dimensions is large than the number of samples [29]. Thus, it uses a kernel size to map the data to a higher dimension so it can easily deal with non-linear relations using objective function, calculated as in Eq. (7).

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + F \sum_{i=1}^m \xi_i \quad (7)$$

Subject of constraint to Eq. (8).

$$y_i(w \cdot \phi(x_i) + b) \geq 1 - \xi_i, \quad \forall i = 1, 2, \dots, m \quad (8)$$

2) *Multinomial naive bayes (MNB)*: MNB is a probability-based classifier developed on the basic principles of Bayesian classifiers, and it is specifically useful for text classifications. It supposes that every feature is independent of other features on condition that the class is given. It is suitable for multi-class problems and most suitable with high-dimensional data such as text documents [30]. MNB performs the model by calculating the conditional probability using Eq. (9), defining each class against the words in the documents making this method simple especially for tasks like spam detection and sentiment analysis.

$$P(t|U) = \frac{P(t)P(U|t)}{P(U)} = P(t) \prod_{i=1}^n P(v_i|t) \quad (9)$$

For each class t , the probability of observing feature vector U is given by Eq. (10).

$$P(U|t) = \frac{\prod_{i=1}^n (f_{i,t} + 1)}{\sum_{k=1}^V (f_{k,t} + 1)} \quad (10)$$

The predicted class is determined by maximizing the posterior probability using Eq. (11).

$$\tilde{t} = \text{argmax}_t P(t) \prod_{i=1}^n P(v_i|t) \quad (11)$$

3) *Bagging classifier*: This process is known as bagging – Bootstrap Aggregating which is an ensemble technique that resolves the problem of variations of a learning machine. It operates differently by building multiple models, often decision

trees on different parts of the training data resulting from bootstrapping, that is random sampling with replacement based on weighted ensemble prediction, computed as in Eq. (12).

$$Y_{final} = g(\prod_{j=1}^J w_j h_j(x)) \quad (12)$$

The last decision is then performed by averaging or voting for these models, that Eq. (13).

$$Var(Y_{final}) = \frac{1}{J^2} \prod_{j=1}^J Var(h_j(x)) + (J - 1)Cov(h_j(x), h_k(x)) \quad (13)$$

4) *Calibrated classifier CV*: The Calibrated Classifier CV is an extension of a base classifier where cross-validation is used to enhance the probability estimation. This method refines the predicted probabilities obtained from classifiers to portray more accurate probabilities to improve decision makers in probabilistic systems, using calibrated function, defined in Eq. (14).

$$P_{calibrated}(y = 1|X) = \sigma(w^T X + b) \quad (14)$$

Where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the function used to computed objective function of calibrated model.

Cross validation is performed to make sure that calibration is done on unseen data, and thus provides higher accuracy on the prediction, computed using Eq. (15).

$$CV(L(w, b)) = \frac{1}{K} \sum_{k=1}^K L(w_k, b_k) \quad (15)$$

Where each fold provides a different set of parameters for calibration.

D. Performance Measure

In the context of performance evaluation of models for identification and classification for English grammar learning, several measures are used to evaluate whether models are effective in providing accurate predictions using metrics such as accuracy, precision and recall, F1 score.

Evaluation using accuracy is based on the concept of measuring the percentage probability that the model prediction for each data point is true with an overall performance, using Eq. (16).

$$Accuracy = \frac{True\ Positives + True\ Negatives}{Total\ Prediction} \quad (16)$$

Precision, also known as positive predictive value, assesses the accuracy of the positive predictions made by the model. It is defined as in Eq. (17), the ratio of true positive predictions to the total number of positive predictions.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (17)$$

Recall, or sensitivity, evaluates to what extent the model selects the number of correct cases when it is, and in percentage of positive prediction, how accurately the model identifies the true negative cases among the wrongly predicted positives. It is defined as in Eq. (18), the accuracy of positive predictions; these are the actual number of positive predictions divided by the total number of positive predictions made.

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \quad (18)$$

F1-score is the balance between precision and recall because the harmonic mean is the better measure than average in such cases, as defined in Eq. (19). They are especially valuable in the scenario of distinguishing between the data set and the data set that is wrongly classified as the opposite class or misclassified as belonging to the opposite class by a certain model.

$$F1 - Score = \frac{2(Precision * Recall)}{Precision + Recall} \quad (19)$$

TABLE I. DESCRIPTION OF SYMBOLS USED IN EQUATIONS

Symbols	Explanation
$f_{t,d}$	Frequency of term t in document d
n_t	Number of documents containing term t
$M(t; d)$	Mutual distribution
$\wp(t d)$	Conditional probability
$\wp(d)$	Prior probability
V_d	Vector representation for document d
$norm(f_{i,d})$	Normalization function to scale the frequency
$context(t_i, d)$	Contextual weighting factor shows semantic similarity measure
ξ_i	Misclassification slack variable
F	Regularization parameters
$f_{i,t}$	Frequency of feature v_i in class t
V	Vocabulary Size
$g(\cdot)$	Majority voting function
w_j	Base classifier

IV. RESULTS AND DISCUSSION

The descriptive analysis performed on the dataset gives an understanding of the linguistic and structural properties of the data. The Fig. 2 of distribution of tenses also represents well different tenses, and most importantly the present continuous and future tenses dominate while the tenses like future perfect continuous appear to be relatively less used. This implies a wide coverage on techniques for tension types that in turn help linguistic diversification. The analysis of the frequency of appearing of numbers of words in a sentence in Fig. 3 has indicated that most of the sentences contain between 40 and 60 words, which means that the examples used in the text should be rather appropriate for educational purposes as far as their length is concerned. The analysis of POS (Part-of-Speech) tags shows in Fig. 4 that the identified data contains a high number of nouns, verbs and determiners, and it is natural considering the focus on the sentence examples. Moreover, the word cloud of the preprocessed text graphically illustrates in Fig. 5 temporal and action-oriented words like ‘next,’ ‘year,’ ‘later’ and ‘gym’ which are evidence of time consciousness within this data set. Altogether, all these visualizations provide supporting evidence to augment the previous argument, regard to the suitability of the presented dataset for tense classification and educational purposes.

A. Results with TF-IDF

The findings from the analysis using TF-IDF accompanied by machine learning and ensemble models provide sufficient support to the proposed solution and confirm the potential of using advanced computational solutions to enhance English learning in educational systems. High accuracy of models such as SVM proved that the models can recognize and learn complex linguistic features with 88% accuracy as well as Bagging Classifier (89%) and CalibratedClassifierCV (87%) models.

These results as shown in Table II illustrating how such models can accurately identify English tenses, an essential

feature of language acquisition. This way, the obtained results demonstrate the possibility of applying these models in real-life settings, including the utilization of automated grammar evaluation, individual learning environments, and language learning assistance tools. The highly favorable efficiency of the methods such as bagging Classifier proves that such an algorithm functions stably and guarantees learners receive accurate feedback regardless of the input data. Just like, SVM demonstrates a very good generalization in its modeling, which makes it suitable for distinguishing small differences in the structure of work sentences as a way of mastering English learning.

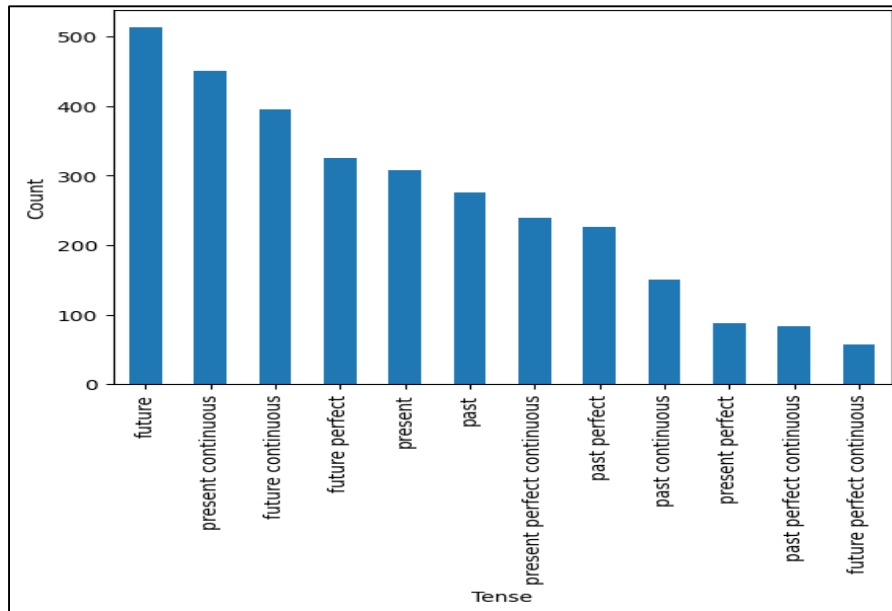


Fig. 2. Distribution of tense.

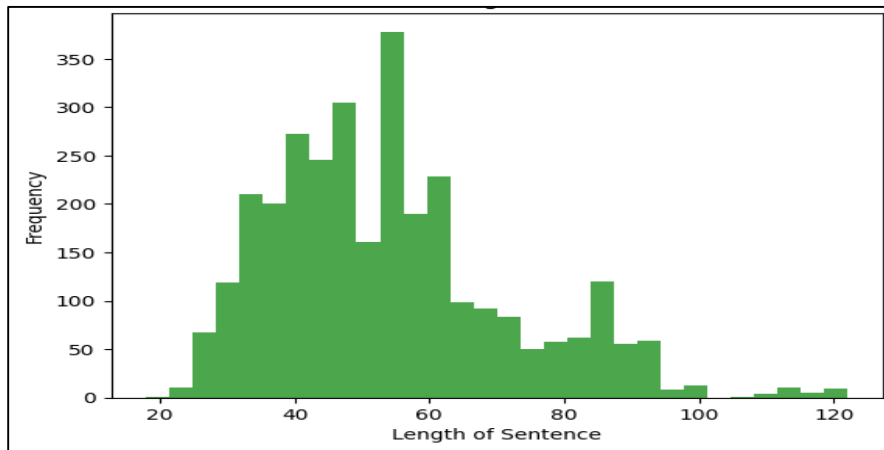


Fig. 3. Length distribution of label sentence.

through syntactic analysis. Including BoW-based models, educational platforms provide more accurate and linguistically grounded tools which contribute to progress of the learning process.

TABLE III. RESULTS OF APPLIED MODELS WITH BOW

Models	Accuracy	Precision	Recall	F1-Score
Shallow Machine Learning				
SVM	89	89	89	89
MNB	83	81	80	79
Ensemble Learning				
Bagging Classifier	90	89	90	89
CalibratedClassifierCV	84	84	84	84

The evaluation of the outcome using the TF-IDF and BoW on machine learning and related methods of ensemble brings out new perspectives of their performances, as shown in Fig. 6. In both feature extraction methods, the performances were excellent, and BoW was slightly better than TF-IDF in most of the models in terms of accuracy. Bagging Classifier achieved 90% using BoW while using TF-IDF it was 89%; for SVM BoW gives 89% while TF-IDF gives only 88%. This indicates that because of BoW's less complex representation, this model was able to capture the patterns in this dataset. Nevertheless, TF-IDF gave comparable results to the other algorithms and demonstrated good performance in evaluating term weight where the importance of terms is decisive. The two paradigm approaches emphasize on their qualities that would be useful in different areas of educational information systems for learning English.

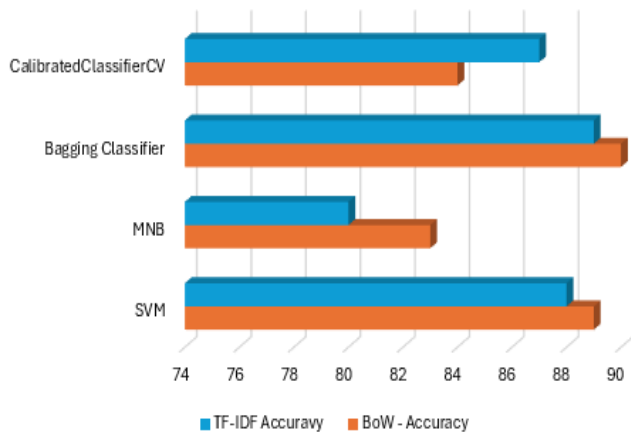


Fig. 6. Comparative analysis of applied features across accuracy measure.

V. CONCLUSION

In the modern educational system, the importance of learning cannot be overstated, as it is essential for students' academic and professional success. However, many learners, especially those in vocational colleges, face challenges in mastering key aspects of English, such as grammar, which forms the foundation of effective communication and learning abilities. Among the crucial components of English grammar, the use of tenses plays a pivotal role in ensuring clarity and

accuracy in vocational college system. The advancement of AI offers significant potential to address these issues by providing personalized and scalable solutions for learning. AI-powered tools can help students understand and apply grammatical concepts more effectively, thereby enhancing their overall learning experience. Our findings in this research demonstrate the efficacy of machine learning models, particularly Support Vector Machine (SVM) and Bagging Classifiers are highly efficient for tense usage classification with accuracies of 89% and 90%, respectively using BoW and TF-IDF features. The obtained results emphasize the possibilities of development and usage of AI-technologies for improving the English language acquisition, providing a robust framework for future educational tools. Moving forward, further research can explore more sophisticated AI techniques to incorporate more complex methods for interactive learning platforms. This study will therefore create a foundation for the development of more enhanced language education to the students at vocational colleges and other institutions.

REFERENCES

- Li, S. (2010). Vocabulary learning beliefs, strategies and language learning outcomes: A study of Chinese learners of English in higher vocational education (Doctoral dissertation, Auckland University of Technology).
- Meiyan, Z. (2023). English Teaching in Higher Vocational Colleges under the Background of the. *Advances in Vocational and Technical Education*, 5(11), 1-9.
- Tividad, M. J. (2024). The English Language Needs of a Technical Vocational Institution. *Mary Joy Tividad (2024). The English Language Needs of a Technical Vocational Institution. Psychology and Education: A Multidisciplinary Journal*, 16(3), 256-275.
- Lesia Viktorivna, K., Andrii Oleksandrovych, V., Iryna Oleksandrivna, K., & Nadia Oleksandrivna, K. (2022). Artificial Intelligence in Language Learning: What Are We Afraid Of. *Arab World English Journal*.
- Chen, J. (2020, December). Strategies for improving the effectiveness of English translation teaching in higher vocational colleges based on data mining. In *Journal of Physics: Conference Series (Vol. 1693, No. 1, p. 012021)*. IOP Publishing.
- Gayed, J. M., Carlon, M. K. J., Oriola, A. M., & Cross, J. S. (2022). Exploring an AI-based writing Assistant's impact on English language learners. *Computers and Education: Artificial Intelligence*, 3, 100055.s
- Guo, L., He, Y., & Wang, S. (2024). An evaluation of English-medium instruction in higher education: influencing factors and effects. *Journal of Multilingual and Multicultural Development*, 45(9), 3567-3584.
- Soodmand Afshar, H., & Doosti, M. (2016). An investigation into factors contributing to Iranian secondary school English teachers' job satisfaction and dissatisfaction. *Research Papers in Education*, 31(3), 274-298.
- Hou, Z. (2021, February). Research on adopting artificial intelligence technology to improve effectiveness of vocational college English learning. In *Journal of Physics: Conference Series (Vol. 1744, No. 4, p. 042122)*. IOP Publishing.
- Chiang, J. (2024). English Grammar Proficiency Predictability in the Prospect of Present Simple and Present Continuous Tenses: A Case from China. *Teaching English Language*, 18(1), 211-222.
- Ma, X. (2022). English Teaching in Artificial Intelligence-based Higher Vocational Education Using Machine Learning Techniques for Students' Feedback Analysis and Course Selection Recommendation. *JUCS: Journal of Universal Computer Science*, 28(9).
- Ginaya, G., Astuti, N. N. S., Mataram, I. G. A. B., & Nadra, N. M. (2020, July). English digital material development of information communication technology ICT in higher vocational education. In *Journal of Physics: Conference Series (Vol. 1569, No. 2, p. 022009)*. IOP Publishing.

- [13] Li, Y. (2020). Application of artificial intelligence in higher vocational english teaching in the information environment. In *Innovative Computing: IC 2020* (pp. 1169-1173). Springer Singapore.
- [14] Zinan, W., & Sai, G. T. B. (2017). STUDENTS' PERCEPTIONS OF THEIR ICT-BASED COLLEGE ENGLISH COURSE IN CHINA: A CASE STUDY. *Teaching English with Technology*, 17(3), 53-76.
- [15] Liao, D. (2022). Deep Learning: Investigation and Analysis on the Public English Course Learning of Vocational College Students. *Advances in Vocational and Technical Education*, 4(2), 53-59.
- [16] Jiang, H., & Wang, H. (2024, March). Designing and Implementing an Intelligent Machine Learning-Based Evaluation System for Assessing English Teaching Quality in Vocational Education. In *2024 International Conference on Interactive Intelligent Systems and Techniques (IIIST)* (pp. 36-40). IEEE.
- [17] Manire, E., Kilag, O. K., Cordova Jr, N., Tan, S. J., Poligrates, J., & Omaña, E. (2023). Artificial Intelligence and English Language Learning: A Systematic Review. *Excellencia: International Multi-disciplinary Journal of Education* (2994-9521), 1(5), 485-497.
- [18] Chen, Y. L., Hsu, C. C., Lin, C. Y., & Hsu, H. H. (2022). Robot-assisted language learning: Integrating artificial intelligence and virtual reality into English tour guide practice. *Education Sciences*, 12(7), 437.
- [19] Llor, M. A. M., Solorzano, D. M. A., Katherine, A., & Moreira, V. (2024). Integration of Artificial Intelligence in English Teaching. *Journal of Cleaner Production*, 289, 125834.
- [20] Amin, M. Y. M. (2023). AI and chat GPT in language teaching: Enhancing EFL classroom support and transforming assessment techniques. *International Journal of Higher Education Pedagogies*, 4(4), 1-15.
- [21] Zhu, J., Zhu, C., & Tsai, S. B. (2021). Construction and analysis of intelligent english teaching model assisted by personalized virtual corpus by big data analysis. *Mathematical Problems in Engineering*, 2021.
- [22] Shen, Y., Liu, Q., Zhang, K., & Zou, R. (2023, December). The Application of Artificial Intelligence Technology in Vocational College Training. In *International Conference on Educational Technology and Administration* (pp. 111-119). Cham: Springer Nature Switzerland.
- [23] Chen, Z., Zhang, J., Jiang, X., Hu, Z., Han, X., Xu, M., ... & Vivekananda, G. N. (2020). Education 4.0 using artificial intelligence for students performance analysis. *Inteligencia Artificial*, 23(66), 124-137.
- [24] Rane, N. L., Paramesha, M., Rane, J., & Kaya, O. (2024). Emerging trends and future research opportunities in artificial intelligence, machine learning, and deep learning. *Artificial Intelligence and Industry in Society*, 5, 2-96.
- [25] Che, Z., Amirthasarayanan, A., Al-Razgan, M., Awwad, E. M., Mohamed, M. Y. N., & Tyagi, V. B. (2024). A Novel Renewable Power Generation Prediction Through Enhanced Artificial Orcas Assisted Ensemble Dilated Deep Learning Network. *IEEE Access*.
- [26] Han, W. (2020). Implementing problem-based learning to enhance speaking skill in a vocational English class: an investigative study (Doctoral dissertation, Rangsit University).
- [27] Wang, X., & Zhong, W. (2022). Research and implementation of English grammar check and error correction based on Deep Learning. *Scientific Programming*, 2022(1), 4082082.
- [28] Blšták, M., & Rozinajová, V. (2022). Automatic question generation based on sentence structure analysis using machine learning approach. *Natural Language Engineering*, 28(4), 487-517.
- [29] Shah, S. M. S., Naqvi, H. A., Khan, J. I., Ramzan, M., & Khan, H. U. (2018). Shape based Pakistan sign language categorization using statistical features and support vector machines. *IEEE Access*, 6, 59242-59252.
- [30] Ishfaq, U., Shabbir, D., Khan, J., Khan, H. U., Naseer, S., Irshad, A., ... & Hamam, H. (2022). Empirical analysis of machine learning algorithms for multiclass prediction. *Wireless Communications and Mobile Computing*, 2022(1), 7451152.

Investigating Immersion and Presence in Virtual Reality for Architectural Visualization

Athira Azmi¹, Sharifah Mashita Syed Mohamad²

Department of Architecture-Faculty of Design and Architecture, Universiti Putra Malaysia, Serdang, Malaysia¹
Department of Computer Science-Faculty of Computer Science and Mathematics, Universiti Malaysia Terengganu,
Kuala Nerus, Malaysia²

Abstract—The architecture industry increasingly relies on virtual reality (VR) for architectural visualization, yet there is a critical issue of insufficient user involvement in the design process. This study investigates the sense of immersion and presence in the virtual environment among 60 Malaysian participants aged 20 to 40. The study utilized a 1000 sq. ft. apartment with three bedrooms and two bathrooms, was replicated in a 3D model based on real-world references. Our findings show that participants were moderately immersed in the virtual environment ($M = 4.86$), but the lack of sense of touch, lack of detail, and interactivity within the virtual environment affected their sense of immersion in VR for architectural visualization. This study has enhanced our understanding of human-computer interaction in VR, specifically for architectural visualization, and has emphasized the importance of improving these aspects to create more effective architectural visualization user experiences.

Keywords—Virtual environment; virtual reality; human-computer interaction; architectural visualization; sense of presence

I. INTRODUCTION

Computer simulation, such as virtual reality (VR) has become an intrinsic part in realizing the vision of Industrial Revolution 4.0 and has revolutionized the way human work, communicate, collaborate and interact with one another [1]. Within the architecture industry, VR has facilitated design, construction and management of the built environment, given its immersive and interactive visualization capabilities [2]. Professionals in the architecture, engineering and construction (AEC) industry relies heavily on visual modes of information transfer and interaction, such as sketches, two-dimensional drawings, computer imagery, visualization and simulation [3, 4].

VR has proven its value in the AEC industry, offering benefits from design reviews to construction simulations [2, 5]. Architects now have access to a variety of tools like Enscape, Lumion, Twinmotion, Unreal Engine, Chaos Vantage, and Chaos V-Ray for real-time visualization of their architectural design proposal. In the conceptual phase, VR is used to explore design ideas quickly, providing architects with a sense of proportion and scale. As projects progress, VR becomes crucial for design validation, allowing architects to experience detailed interiors and exteriors realistically [2]. This technology allows architects to make fast well-informed decisions, considering aesthetics, cost, and environmental impact. The design process, once time-consuming, now occurs in seconds.

Despite evidence showing the use of these digital visualization technology could bring significant improvement in

communication, exchange and interoperability of information, there exists a lack of end-user involvement and perspective in the design process [6]. According to Lee et al. [6], to effectively use digital visualization for architectural design collaboration, it is important to ensure the effectiveness of the visualization system from the end-user perspective.

Prabhakaran et al. [4] identify some hurdles within the architecture and construction sectors concerning immersive technologies like VR. These challenges encompass deficient communication among stakeholders, primarily attributable to the nascent state of VR infrastructure. Specifically, issues such as hardware requisites, user mobility constraints, ease of operation, and device ergonomics contribute to these communication inefficiencies [4]. Moreover, our earlier findings highlight the impact of VR hardware issues on users' sense of presence, a critical aspect underscored by Prabhakaran et al. [4] to enrich immersive experiences. The sense of presence, as highlighted by is pivotal for users to truly feel immersed in the virtual environment. While simulating spatial movement on a screen is feasible, Gomez-Tone et al. [8] argue that genuine immersion requires viewers to perceive themselves within the virtual space, fostering a profound sense of presence.

This study emphasizes the importance of end-user involvement in architectural design, utilizing VR technology to delve into user experiences. By integrating human emotions into digital visualization tools, the aim is to create architecture that caters to emotional needs. Facilitating improved collaboration between architects and users, the process enriches design through digital visualization. However, there remains a gap in understanding user responses to VR immersion in architectural contexts, especially in Malaysia. This research focuses on exploring Malaysian users' immersive experiences with VR when interacting with 3D building models. The goal is to pave the way for more empathetic architectural designs that prioritize user needs.

The paper is structured as follows: it begins with a background and literature review in Section I and II, providing context for the study. Following this, the experiment conducted to investigate the sense of presence among Malaysian users in VR is detailed in Section III. Subsequently, the results of the experiment are presented, accompanied by a thorough discussion is given in Section IV and Section V. Finally, the paper concludes by acknowledging the study's limitations and offering recommendations for future research in this domain in Section VI.

II. LITERATURE REVIEW

A. VR and Immersive Virtual Environment for Architectural Visualization

In the context of built environment, virtual reality (VR) can be defined as the experience of feeling present in a fictitious or envisioned environment through its representation [5]. An immersive VR system requires a three-dimensional (3D model), a head-mounted display (HMD), interaction devices or controllers and software to run the program. Immersive VR enables users to immerse themselves into the virtual environment [9].

As an immersive technology, VR enables human experience in the virtual environment through the sense of presence, which is the major factor in delivering lifelike experiences in the simulated environment [5]. Realism in the immersive virtual environment via VR is considered an important element for architectural design visualization, as the main objectives of VR applications in built environment field is to facilitate visualization and simulation of the architecture design [5].

Apart from tremendous benefit as learning tool for visualization in architecture education [10,11,12,13], most VR research in the construction industry proved that the technology benefits in the decision-making during the design process among design professionals [14,15]. However, there is a considerable gap in the effectiveness of VR for design collaboration with real-end users of buildings and clients from non-design background. Thus, it is important to investigate whether the use of VR for architectural visualization could achieve the level of realism as expected by these non-design users.

However, as stated earlier it is found that there is a lack of study that investigates how users respond to the design while experiencing it in VR, especially for the context of Malaysian user. Abdul Ghafar and Ibrahim [16] stated that there is lack of emphasis given to human factor when using these digital visualization tools. Azmi et al., [7] also argue that the use of VR for visualization of housing design for homebuyers are limited in terms of touch sensation and navigation. In addition, Delgado et al., [3], supported by Lyu et al., [17] argue that despite the advancement of VR technology in visual rendering of the immersive virtual environment, other sensory simulation such as auditory, tactile, thermal, olfactory and taste remain relatively underdeveloped. These arguments highlighted the importance of exploring end-user engagement through the visual representation using VR during the design stage to determine the effectiveness of this digital visualization technique during design collaboration.

B. Emotional Intelligence in Digital Architectural Visualization

Over the past decade, emotional intelligence has been the focus of research from different disciplines of studies to explore the advantage of applying the concept to benefit their respective fields. Salovey and Grewal [18] described emotional intelligence as the skill that brings together the fields of emotions and intelligence by viewing emotions as useful sources of information that help one to make sense and navigate the social environment. As human responds cognitively and

emotionally to the built environment, the use of VR to evaluate users' emotion during their immersive virtual environment experience serves as a promising framework for the future of design and research in the built environment [19].

However, user experience has been the main issue in VR for architectural visualization as VR is highly visual and does not really support other human sensations such as olfactory and haptic [7,19]. VR has the capacity to simulate the illusion of being in a place through the sense of presence, hence, it is crucial to meet the viewers' expectations, cognitive and emotional dimension of the built environment [19]. Research has shown that some design element in the built environment reflected higher sense of attraction of the brain to the surroundings, which impacts the psychological wellbeing of the inhabitants [20].

It is found that research in regard to emotional intelligence in architectural visualization within the virtual environment has not been widely studied. Thus, this paper argues the critical need for a study that investigates the synergistic relationship between VR and user experience. This study is trying to fill the considerable gap in studies that examine how the end-user, which is usually non-design professional perceive the digital simulation of an architecture design. This is pertinent to help architects to understand and improve the design of virtual spaces that has more meaning, physically and emotionally to the end-users. It is substantial to explore emotional intelligence during the design process that requires exchange of not only technical decisions but also the human experience, especially with the use of digital visualization technology such as VR.

C. Sense of Presence

Immersive virtual environment enables human experience in a given environment through the sense of presence. Presence is defined as the subjective experience of being in one place or environment, even when one is physically situated in another [21]. It has been proven by various recent research that human emotion in the virtual environment is similar to the emotion in the physical environment [7, 22, 23].

Caroux [24] indicates that while immersion would be typically related to sensory feedback that results in the sense of being surrounded by the virtual environment, presence would be more related to a cognitive psychological response that is the feeling of being in the virtual environment. Several factors affecting presence has been identified, including i) human factors - the level of experience and age [25], ii) the visual and sensory input [26], and iii) technological factors – stereopsis, field of view, and interactivity [5].

Following Paes et al., [5], identifying factors that affect presence in virtual environment is a vital step to improve the VR application in the built environment. This study aims to fully optimize the advancement of digital visualization technology in the AEC industry to ensure effective user experience in VR, especially for the context of Malaysian users. Hence, it is possible to create a new space for better discussions of design solutions between architects and end-users for a design that meet user needs. With the considerable gap in literature concerning users from Asian background in using VR technology for architectural visualization, following Azmi et al., [7], this study is focusing on the users' behavioral response in VR within the

context of Malaysian user in local architectural design. In the next section, this paper delineates the experimental research methodology employed to investigate the sense of presence among Malaysian users within VR environments for architectural visualization purposes.

III. RESEARCH METHODOLOGY

This study is an empirical and relational study as most human-computer interaction studies such as Paes et al., [5]. Employing a one-group posttreatment-only pre-experimental design, the study leverages survey questionnaires to assess user experience within virtual environments. This design entails exposing a single group of participants to an intervention, followed by measurement, as elucidated by Creswell [27]. Participants engage with a VR setup simulating the interior of a house, after which they provide feedback through questionnaires to gauge their immersion and presence within the virtual environment.

A. Participants

Participants were recruited for the experiment using purposive sampling method, via advertisements distributed in social media and word-of-mouth. In order to be included in the study, individuals had to meet the following criteria: (1) Malaysian nationality; (2) aged between 20 and 40 years old; (3) not physically or mentally impaired and (4) not under serious medications for health-related problems. In regard to the sample size, a power sample analysis using G*Power version 3.1.9.7 [28] was conducted with a level of statistical significance equal to 80% with a medium effect size of Cohen's $d = 0.5$ ($\alpha = .05$). This follows similar sample size estimation by other studies in VR experiments such as Pallavicini and Pepe [29]. Result of the power sample analysis using G*Power suggests an estimation of sample size for the research design and one sample case statistical test is 27 participants. In this study, a total of 60 samples were recruited, which is more than adequate to get a high statistical power.

Based on the post-hoc power analysis to compute achieved power based on 60 samples conducted in G*Power version 3.1.9.7, the study has 99% power to detect medium-sized effect $d = 0.5$ ($\alpha = .05$). The 60 participants recruited in this study include: 33 female (55.0%) and 27 male (45.0%); mean age of 30 years old (SD: 2.81). 95% of the participants are from Malay racial background. Only two participants have the experience of using VR for architectural visualization before the experiment.

B. 3D Model and VR Apparatus

An apartment designed in Selangor, Malaysia was selected as the experiment environment. This residential unit consists of 1000 square feet of living space with three bedrooms and two bathrooms. A 3D model of an interior of a house was developed in Sketchup Pro 2019 (version 19.3). Enscape software (version 2.6.1) was used for the real-time 3D visualization and as the VR plugin for Sketchup. Fig. 1 shows the 3D model of the kitchen of the apartment in Sketchup software. Fig. 2 shows the virtual environment as viewed in VR, rendered using Enscape software.



Fig. 1. 3D model of the kitchen in Sketchup software.



Fig. 2. VR view of the kitchen rendered in Enscape software.

For the VR apparatus, HTC Vive was used. HTC Vive consists of a head-mounted display (HMD), two base sensors for tracking position and orientation, and a set of controllers were used as the VR equipment. The computer used was a Dell G7 15 7590 laptop with Intel Core^{TM} i7-8750H processor and NVIDIA GeForce GTX 1050 (4GB GDDR5) graphics card. The computer has an 8GB RAM that operates with the Windows 10 operating system. The specifications on this computer satisfy the minimum system requirements for HTC Vive. Fig. 3 illustrates the VR setup which consists of the HTC Vive head mounted device (HMD), two base stations, two controllers.



Fig. 3. VR device setup.

The HTC Vive system allows for physical movement within a minimum play area of 2 meters x 2.5 meters, hence the experiment was set to comply within this play area. The play

area was setup using Viveport software. Fig. 4 shows the participants using the HTC Vive during the experiment.



Fig. 4. Participant using the HTC Vive during experiment.

C. Data Collection Instruments

The data collection instruments used in this study include:

- Demographics – the question includes their gender, age, marital status, racial background, level of education, profession and prior VR experience;
- Consent form - Before the experiment, every participants was asked to read carefully the consent form, as it is the right of every person to make informed decisions regarding their participation in a research study, after being informed of all aspects of their role in the study as required in the Belmont principle of respect for persons [30]. This research has also been approved by the Universiti Putra Malaysia Ethics Committee for Research Involving Human Subject (JKEUPM-2020-028) prior to the data collection;
- The Virtual Presence Questionnaire (VPQ) - After the participants had been exposed to the virtual environment, a Virtual Presence Questionnaire (VPQ) was given to each participant. The VPQ was adopted from the Presence Questionnaire, originally used by Witmer and Singer (1998). The Presence Questionnaire was used to evaluate the level of presence that each participant experienced during the VR experience to view the virtual environment. The VPQ developed in this study is adapted based on the instruments developed by prior research that includes Witmer and Singer [21], Westerdahl et al., [31] and Heydarian et al., [32].

The VPQ consists of nine Likert Scale-based questions (seven-point scale), two questions that requires a yes or no answer, and two open-ended questions. These questions were developed based on prior research to determine the level or immersion and presence of the participants in the virtual environment, including the level of realism of the virtual environment. The results from the VPQ would add to the understanding in regard to the participants' differences and abilities in a given virtual environment, and the characteristics of the virtual environment that may affect presence.

The two open-ended questions in the VPQ seek to obtain additional information based on the participants' experience in the virtual environment using the VR devices. In the open-ended questions, the first question invited the participants to comment on the kind of information that they think is lacking from the virtual environment; while the second question invited them to provide suggestions regarding the application of VR for architectural visualization in Malaysia. The two open-ended questions in the VPQ are:

- What kind of information do you think is lacking from the VR environment?
- What are your comments regarding the application of VR for architectural visualization?

According to Creswell [27], Neuert et al., [33], and Aithal and Aithal [34], open-ended questions in a set of questionnaires allow the respondent to express an opinion without being influenced by the researcher. The open-ended questions advantages include the possibility of discovering the responses that participants gave spontaneously, and thus avoiding the bias that may result from suggesting responses in close-ended questions such as Likert-scales [35].

IV. RESULT AND ANALYSIS

A descriptive statistical analyses to describe the central tendencies and variability in participants' response in this study were conducted using SPSS version 25 software package. Table I illustrates the mean and standard deviation for the participants' responses in the VPQ.

TABLE I. VPQ QUESTIONNAIRE ITEMS - MEAN AND SD

VPQ QUESTIONNAIRE ITEMS - MEAN AND SD		
VPQ Questionnaire Items	Source	Mean (SD)
"How physically fit do you feel today?" ^a	[21]	6.27 (0.73)
"How good are you at blocking out external distractions when you are involved in something?" ^a	[21]	5.75 (0.97)
"Are you easily disturbed or distracted when working on tasks?" ^a	[21]	3.87 (1.75)
"Did you get bored with the VR model during the viewing experience?" ^a	[31]	5.80 (1.61)
"Did the surfaces such as walls, floors, and furniture look real in the VR model when you view it?" ^a	[31]	4.95 (1.65)
"Could you orient yourself in the internal environment during the VR experience?" ^a	[31]	5.43 (1.53)
"How much did your experiences in the virtual environment seem consistent with your real-world experience?" ^a	[21]	4.95 (1.64)
"How realistic was your sense of movement around in the virtual environment?" ^a	[21]	4.60 (1.59)
"How difficult was it to understand the characteristics of the house in VR?" ^a	[32]	3.25 (1.72)
"Did you feel that the Virtual Reality (VR) model lacked information for you to understand the interior of the house?"	[31]	Yes = 55%
"Do you feel any discomfort or dizziness after the VR experience?"	[7]	No = 45%

^a. The question response format was a 7-point Likert scale.

Findings from Table I indicates that the participants were relatively fit during the experiment (M = 6.267, SD = 0.7333). The participants were also somewhat focused during the experiment based on their responses in Question 2 (M = 5.750, SD = 0.968) and Question 3 (M = 3.867, SD = 1.751). In the virtual environment, the result showed that most of the participants thought that the textures on walls, floors, and furniture look real in the VR model (M = 4.950, on a 7-point Likert scale where 1 = “not real at all” and 7 = “very real”). The same responses also applied to whether their experience seems consistent with real-world experience (M = 4.950, on a 7-point Likert scale where 1 = “not consistent at all” and 7 = “very consistent”).

The participants also felt that they were able to orient themselves in a virtual environment (M = 5.433, on a 7-point Likert scale where 1 = “not at all” and 7 = “totally”); and that their sense of movement in the virtual environment is not much realistic (M = 4.600, on a 7-point Likert scale where 1 = “not at all” and 7 = “totally”). Of the 60 participants, 33 participants (55%) felt that the VR lacked information for them to understand the interior of the house. Finally, a majority of participants (78.3%) did not feel any discomfort or dizziness after the VR experience. The results also reveal that participants were relatively focused during the experiment based on their ratings from Questions 1 to 4. The participants also indicated that the virtual environment was moderately realistic based on their ratings from Questions 5 to 10.

A. Analysis of Open-Ended Questions

Of the 60 participants, only 33 participants (55%) provided their answers to the open-ended questions. To identify participants’ evaluation of the virtual environment from the two

open-ended questions, this study deployed thematic analysis. Four main themes emerged from the participants’ answers in these two open-ended questions which are: “feel”, “detail”, “size” and “interactivity”. The thematic analysis is presented in Table II.

In the first open-ended question - what kind of information do you think is lacking from the VR model? 18 participants mentioned that they were unable to estimate the size and dimensions of the interior of the house based on their VR experience. These participants felt that it is essential for them to feel the size of the space in the architectural visualization. In addition, seven participants commented that the information about materiality and texture of materials in the house is lacking in the virtual environment. One participant commented that the VR model was unsatisfying since they were unable to touch the materials inside the house or feel the wind from the window or balconies. Two participants also commented on the lack of sense of sound and smell in the virtual model. On the other hand, five participants commented on the realism of the virtual environment that lacks detail and seems unrealistic.

In the second open-ended question, eight participants provided suggestions that the VR equipment or space provided should enable them to walk around freely in the virtual environment without being restricted. Five participants provided suggestions for a more realistic virtual environment with better resemblance of a real physical building. One participant suggested that the VR application could provide options in terms of furniture arrangement, while another participant suggested that the VR application to provide different color options in the architectural visualization.

TABLE II. THEMATIC ANALYSIS OF VPQ ANSWERS

Open-ended Question 1: What kind of information do you think is lacking from the VR environment?	
Participants’ Response	Theme
“I want to really feel the house to make sense of it in terms of touch and smell; I am not sure how that can be achieved virtually”.	Feel
“I can’t feel the element of wind to make sense of the open space concept of the house.	
“Physical environment is more satisfactory as there is physical touch. The feeling of walking inside a real house is not similar as in VR”.	
“I can’t imagine the quality of materials in the VR model”.	Details
“Lack of detailing in the VR model”.	
“I can’t estimate the size and dimension of space inside the virtual environment”	Size
“It’s difficult to understand the size of the room from the VR”	
“I think a lacking feature of the VR technology is that I can’t estimate the size of the house”	
Open-ended Question 2: What are your comments regarding the application of VR for architectural visualization?	
Participants’ Response	Theme
“Adding capability to vary household content so that it may suit different individuals (worker, students, family, big family), and having multiple household arrangement settings for the same house layout”.	Interactivity
“I would suggest a bigger space for people can walk freely while using the VR equipment”.	
“I would suggest including measurement in the VR model for potential homebuyers to realise the size of the house”.	
“It is suggested to include interactive information such as measurement (height and width) and colour options for walls or furniture”.	

V. DISCUSSION

The study acknowledges limitations in touch and interactivity, impacting the participants' sense of presence. Issues related to hardware constraints and restricted movements contribute to a less realistic experience. Based on participants' response, it can be discerned that the participants' experience in the virtual environment are moderately natural, which results to a moderate level of immersion and presence. The authors believes that a greater degree of immersion and presence is detracted due to the lack of sense of touch in the virtual environment.

Most participants highlighted that the inability to physically touch or feel the materials within the virtual environment was a significant limitation. The authors contend that this absence of tactile interaction compromised the overall sense of immersion and presence. Furthermore, additionally, participants pointed out the unrealism in navigation, as the VR device was tethered with wires, which restricted movement and affected their ability to move freely within the virtual space. We recognized the limitations of the VR device used in this study, which requires cables attached to the HMD that limits participants' movements; hence affecting the level of realism in the virtual environment.

Apart from that, most of the participants also opined that they were unable to sense the feeling of space or estimate the dimensions or size of the house that they view in the virtual environment, which is one of the critical factors in enhancing their experience and presence in the architectural visualization. We acknowledged these findings as the most important one, considering the assumption by past literature that the virtual environment could simulate the physical world conditions with sufficient accuracy and are efficient in representing spatial information [36].

Furthermore, according to Azmi et al., [7], due to the lack of interactivity in the virtual environment, the atmospheric qualities of spaces are diminished, hence the sense of presence was lacking. The virtual environment could not simulate the physical world accurately in terms of the quality of the surrounding and its relation to human senses. This corroborates findings from [37] that suggests integrating technology capable of aligning the visual perception of virtual objects with the tactile sensations of holding and touching real objects with bare hands can significantly enhance the quality of experiences and sense of presence in virtual environments.

This study also concurs with Higuera-Trujillo et al., [19] that user experience is a vital issue that needs to be addressed. The user experience includes enhancing the capacity of VR simulation to generate the 'place illusion' and the credibility of the 3D scenarios in the virtual environment to meet the users' expectations of the simulated environment [19]. Thus, the findings in this study allows for further understanding of the current user experience in VR for architectural visualization.

VI. CONCLUSION

This study aims to enhance the architecture industry's adaptation to Industry 4.0 by leveraging digital visualization, particularly VR, in architectural design. The study contributes to understanding user experiences in VR architectural

visualization, particularly in Malaysia. While VR has significantly improved design review and collaboration, there is still a gap in end-user involvement. By evaluating VR effectiveness from the user's perspective, this research addresses these challenges. Issues like hardware requirements and user mobility hinder seamless communication in VR architectural visualization. Additionally, the study emphasizes the importance of the user's sense of presence in VR environments. Findings from our empirical study shed light on participants' experiences, highlighting realism and challenges.

This study highlights the significant potential of VR technology to revolutionize both the real estate and architectural industries. In real estate, VR offers a more sustainable and versatile alternative to physical show units, allowing developers to enhance their marketing strategies, reach broader audiences, and provide homebuyers with a more comprehensive and immersive evaluation experience than traditional methods. In architecture, VR serves as a tool for understanding user emotions and behaviors within designed spaces, fostering empathetic and user-focused designs, with potentials to focus on groups such as the elderly or individuals with special needs.

Recommendations include enhancing VR interactivity for better user engagement. This research contributes to understanding users' needs in VR architectural visualization, particularly in Malaysia, aiming to improve design collaboration and user satisfaction. It anticipates VR technology's enhanced role in architectural visualization, aligning with users' preferences and needs. Future research directions include investigating innovative approaches to improve the realism and immersion of VR environments for architectural visualization and exploring advancements in VR hardware such as tactile gloves or body-kit to allow sense of touch in the virtual environment.

LIMITATION OF STUDY

Several limitations of this study should be acknowledged. First, due to budget constraints, the study used an older version of the VR equipment, which is HTC Vive which was succeeded by newer versions of the VR headset. The HTC Vive (released in 2016) was followed by the HTC Vive Pro in 2018, and later, the HTC Vive Cosmos and Vive Pro 2 were introduced in 2019 and 2021, respectively.

Additionally, the study focused solely on the responses of participants in Malaysia that fulfil the inclusion criteria, which limits the participant pool to other nationalities with various cultural differences, which might have different perception due to cultural differences. Lastly, the absence of haptic devices meant that tactile feedback was not incorporated into the virtual experience. Future research should consider integrating haptic technology to allow users to physically interact with textures in the virtual environment, and assess its impact on emotions and behaviors.

ACKNOWLEDGMENT

This study was part of a research funded by a grant from Universiti Putra Malaysia (Geran Putra IPM GP-IPM/2023/9772700).

REFERENCES

- [1] Gunal, Murat M. "Simulation and the fourth industrial revolution." *Simulation for Industry 4.0: Past, Present, and Future* (2019): 1-17. https://doi.org/10.1007/978-3-030-04137-3_1
- [2] Zhang, Yuxuan, Hexu Liu, Shih-Chung Kang, and Mohamed Al-Hussein. "Virtual reality applications for the built environment: Research trends and opportunities." *Automation in Construction*, Vol 118 (2020): 103311. <https://doi.org/10.1016/j.autcon.2020.103311>.
- [3] Delgado, Juan Manuel Davila, Lukumon Oyedele, Peter Demian, and Thomas Beach. "A research agenda for augmented and virtual reality in architecture, engineering and construction." *Advanced Engineering Informatics*, Vol 45 (2020): 101122. <https://doi.org/10.1016/j.aei.2020.101122>.
- [4] Prabhakaran, Abhinesh, Abdul-Majeed Mahamadu, and Lamie Mahdjoubi. "Understanding the challenges of immersive technology use in the architecture and construction industry: A systematic review." *Automation in Construction*, Vol 137 (2022): 104228. <https://doi.org/10.1016/j.autcon.2022.104228>.
- [5] Paes, Daniel, Javier Irizarry, and Diego Pujoni. "An evidence of cognitive benefits from immersive design review: Comparing three-dimensional perception and presence between immersive and non-immersive virtual environments." *Automation in Construction*, Vol 130 (2021): 103849. <https://doi.org/10.1016/j.autcon.2021.103849>.
- [6] Lee, Jin Gang, JoonOh Seo, Ali Abbas, and Minji Choi. "End-Users' augmented reality utilization for architectural design review." *Applied Sciences*, Vol 10, no. 15 (2020): 5363. <https://doi.org/10.3390/app10155363>.
- [7] Azmi, Athira, Rahinah Ibrahim, Maszura Abdul Ghafar, and Ali Rashidi. "Smarter real estate marketing using virtual reality to influence potential homebuyers' emotions and purchase intention." *Smart and Sustainable Built Environment*, Vol 11, no. 4 (2022): 870-890. <https://doi.org/10.1108/SASBE-03-2021-0056>.
- [8] Gómez-Tone, Hugo C., John Bustamante Escapa, Paola Bustamante Escapa, and Jorge Martin-Gutierrez. "The drawing and perception of architectural spaces through immersive virtual reality." *Sustainability*, vol 13, no. 11 (2021): 6223. <https://doi.org/10.3390/su13116223>.
- [9] Suh, Ayoung, and Jane Prophet. "The state of immersive technology research: A literature analysis." *Computers in Human Behavior*, vol 86 (2018): 77-90. <https://doi.org/10.1016/j.chb.2018.04.019>.
- [10] Elgewely, Maha Hosny, Wafaa Nadim, Ahmad ElKassed, Mohamed Yehiah, Mostafa Alaa Talaat, and Slim Abdennadher. "Immersive construction detailing education: building information modeling (BIM)-based virtual reality (VR)." *Open House International* 46, no. 3 (2021): 359-375. <https://doi.org/10.1108/OHI-02-2021-0032>
- [11] Ibrahim, Anwar, Amneh Ibrahim Al-Rababah, and Qanita Bani Baker. "Integrating virtual reality technology into architecture education: the case of architectural history courses." *Open House International* 46, no. 4 (2021): 498-509. <https://doi.org/10.1108/OHI-12-2020-0190>
- [12] Ummihusna, Annisa, and Mohd Zairul. "Exploring immersive learning technology as learning tools in experiential learning for architecture design education." *Open House International* 47, no. 4 (2022): 605-619. <https://doi.org/10.1108/OHI-01-2022-0020>
- [13] Pei, Wanyu, Tian Tian Sky Lo, and Xiangmin Guo. "Integrating Virtual Reality and interactive game for learning structures in architecture: the case of ancient Chinese dougong cognition." *Open House International* 48, no. 2 (2023): 237-257. <https://doi.org/10.1108/OHI-05-2022-0136>
- [14] Juan, Yi-Kai, Hao-Yun Chi, and Hsing-Hung Chen. "Virtual reality-based decision support model for interior design and decoration of an office building." *Engineering, Construction and Architectural Management* 28, no. 1 (2019): 229-245. <https://doi.org/10.1108/ECAM-03-2019-0138>
- [15] Ewart, Ian J., and Harry Johnson. "Virtual reality as a tool to investigate and predict occupant behaviour in the real world: the example of wayfinding." *ITcon* 26 (2021): 286-302. <https://doi.org/10.36680/j.itcon.2021.016>
- [16] Ghafar, Maszura Abdul, and Rahinah Ibrahim. "Effects of Human Culture Among AEC Professionals Towards Adaptation of Collaborative Technology in Industrialized Project Delivery." *International Journal of Digital Innovation in the Built Environment (IJDIBE)* 9, no. 1 (2020): 36-48. <https://doi.org/10.4018/IJDIBE.2020010103>
- [17] Lyu, Kun, Arianna Brambilla, Anastasia Globa, and Richard de Dear. "An immersive multisensory virtual reality approach to the study of human-built environment interactions." *Automation in construction* 150 (2023): 104836. <https://doi.org/10.1016/j.autcon.2023.104836>
- [18] Salovey, Peter, and Daisy Grewal. "The science of emotional intelligence." *Current directions in psychological science* 14, no. 6 (2005): 281-285. <https://doi.org/10.1111/j.0963-7214.2005.00381>.
- [19] Higuera-Trujillo, Juan Luis, Carmen Llinares, and Eduardo Macagno. "The cognitive-emotional design and study of architectural space: A scoping review of neuroarchitecture and its precursor approaches." *Sensors* 21, no. 6 (2021): 2193. <https://doi.org/10.3390/s21062193>.
- [20] Azzazy, Sameh, Amirhosein Ghaffarianhoseini, Ali GhaffarianHoseini, Nicola Naismith, and Zohreh Dobarjeh. "A critical review on the impact of built environment on users' measured brain activity." *Architectural Science Review* 64, no. 4 (2021): 319-335. <https://doi.org/10.1080/00038628.2020.1749980>
- [21] Witmer, Bob G., and Michael J. Singer. "Measuring presence in virtual environments: A presence questionnaire." *Presence* 7, no. 3 (1998): 225-240. <https://doi.org/10.1162/105474698565686>.
- [22] Lipp, Natalia, Natalia Dużmańska-Misiarczyk, Agnieszka Strojny, and Paweł Strojny. "Evoking emotions in virtual reality: schema activation via a freeze-frame stimulus." *Virtual Reality* 25, no. 2 (2021): 279-292. <https://doi.org/10.1007/s10055-020-00454-6>.
- [23] Yung, Ryan, Catheryn Khoo-Lattimore, and Leigh Ellen Potter. "Virtual reality and tourism marketing: Conceptualizing a framework on presence, emotion, and intention." *Current Issues in Tourism* 24, no. 11 (2021): 1505-1525. <https://doi.org/10.1080/13683500.2020.1820454>.
- [24] Caroux, Loïc. "Presence in video games: A systematic review and meta-analysis of the effects of game design choices." *Applied Ergonomics* 107 (2023): 103936. <https://doi.org/10.1016/j.apergo.2022.103936>.
- [25] Lavoie, Raymond, Kelley Main, Corey King, and Danielle King. "Virtual experience, real consequences: the potential negative emotional consequences of virtual reality gameplay." *Virtual Reality* 25, no. 1 (2021): 69-81. <https://doi.org/10.1007/s10055-020-00440-y>.
- [26] Kim, Yong Min, and Ilsun Rhiu. "A comparative study of navigation interfaces in virtual reality environments: A mixed-method approach." *Applied Ergonomics* 96 (2021): 103482. <https://doi.org/10.1016/j.apergo.2021.103482>
- [27] Creswell, John W., and J. David Creswell. *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications, 2017.
- [28] Faul, Franz, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. "Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses." *Behavior research methods* 41, no. 4 (2009): 1149-1160. <https://doi.org/10.3758/BRM.41.4.1149>.
- [29] Pallavicini, Federica, and Alessandro Pepe. "Virtual reality games and the role of body involvement in enhancing positive emotions and decreasing anxiety: within-subjects pilot study." *JMIR serious games* 8, no. 2 (2020): e15635. <https://doi.org/10.2196/15635>
- [30] Privitera, Gregory J. *Research methods for the behavioral sciences*. Sage Publications, 2022.
- [31] Westerdahl, Börje, Kaj Suneson, Claes Wernemyr, Mattias Roupé, Mikael Johansson, and Carl Martin Allwood. "Users' evaluation of a virtual reality architectural model compared with the experience of the completed building." *Automation in construction* 15, no. 2 (2006): 150-165. <https://doi.org/10.1016/j.autcon.2005.02.010>.
- [32] Heydarian, Arsalan, Joao P. Carneiro, David Gerber, Burcin Becerik-Gerber, Timothy Hayes, and Wendy Wood. "Immersive virtual environments versus physical built environments: A benchmarking study for building design and user-built environment explorations." *Automation in Construction* 54 (2015): 116-126. <https://doi.org/10.1016/j.autcon.2015.03.020>
- [33] Neuert, Cornelia E., Katharina Meitinger, and Dorothee Behr. "Open-ended versus closed probes: Assessing different formats of web probing." *Sociological Methods & Research* 52, no. 4 (2023): 1981-2015. <https://doi.org/10.1177/00491241211031271>
- [34] Aithal, Architha, and P. S. Aithal. "Development and validation of survey questionnaire & experimental data—a systematical review-based statistical approach." *International Journal of Management, Technology, and Social*

- Sciences (IJMTS) 5, no. 2 (2020): 233-251.
<http://dx.doi.org/10.2139/ssrn.3724105>
- [35] Reja, Urša, Katja Lozar Manfreda, Valentina Hlebec, and Vasja Vehovar. "Open-ended vs. close-ended questions in web questionnaires." *Developments in applied statistics* 19, no. 1 (2003): 159-177.
- [36] Paes, Daniel, Eduardo Arantes, and Javier Irizarry. "Immersive environment for improving the understanding of architectural 3D models: Comparing user spatial perception between immersive and traditional virtual reality systems." *Automation in Construction* 84 (2017): 292-303. <https://doi.org/10.1016/j.autcon.2017.09.016>.
- [37] Kim, Kihong, Ohyang Kwon, and Jeongmin Yu. "Evaluation of an HMD-based multisensory virtual museum experience for enhancing sense of presence." *IEEE Access*, vol 11, pp. 100295-100308, 2023, doi: 10.1109/ACCESS.2023.3311135

Data Mining MRO-BP Network-Based Evaluation Effectiveness of Music Teaching

Yifan Fan*

Henan Institute of Science and Technology, School of Music and Dance, Xinxiang 453003, Henan, China

Abstract—This study addresses the need for data analysis in evaluating the teaching outcomes of higher music education. It proposes a solution using data-driven algorithms to measure and analyze these outcomes. This study focuses on the issue of measuring and evaluating the outcomes of music education teaching. It analyzes the process of measuring and assessing these outcomes, designs a program for doing so, and introduces key technologies such as music education teaching process analysis, measurement of music teaching outcomes, construction of an assessment model for music teaching outcomes, and application of the assessment model. The study selects teaching content, practical skills, and social practice ability as the three aspects to evaluate. The results demonstrate that this method achieves higher assessment accuracy and requires less time, effectively addressing the challenge of measuring and evaluating the teaching outcomes of higher music education using big data. The findings demonstrate that the technique exhibits a high level of assessment accuracy and is less time-consuming. Additionally, it effectively addresses the challenge of measuring and evaluating the teaching accomplishments in higher music education from the viewpoint of big data.

Keywords—Mushroom propagation optimisation algorithm; BP neural network; higher music education teaching outcomes measurement; algorithm evaluation

I. INTRODUCTION

Currently, music education is progressing towards the establishment of a disciplinary structure, the application of scientific methods, the cultivation of higher cognitive skills, and the production of various academic accomplishments [1]. Evaluating and appraising music education instruction, as a crucial tool for advancing higher music education teaching, has immense importance in boosting teaching quality, driving instructional improvement, and fostering comprehensive student growth [2]. Scientific measurement and evaluation may aid instructors in comprehending students' learning development, pinpointing strengths and flaws in teaching, and therefore enhancing teaching techniques and approaches [3]. The emergence of intelligent disciplines has led to the adoption of data-driven models for measuring and assessing educational teaching approaches in higher music education. This trend has gained significant attention from professionals and researchers in the area. Hence, it is crucial to investigate the intelligent, scientific, and systematic approaches to measuring and evaluating teaching results in higher music education. This is essential for the robust advancement of the theory and practice of music education discipline.

Presently, the evaluation and measurement of teaching outcomes in higher music education mostly focuses on the study

of measuring indices, techniques, and assessment of teaching outcomes in music education. Yu and Zou [4] examined the current state of using artificial intelligence technology in the field of music and optimization strategies, and proposed a method for assessing music teaching based on artificial intelligence technology. Chen et al. [5] investigated the measurement and assessment methods of teaching outcomes in higher music education from a perspective of music education psychology. Yang [6] explored the current state of music education in higher vocational colleges and universities during the rise of aesthetic education, and suggested improvement and optimization measures in three teaching aspects; Liao and Huang [7] proposed a teaching evaluation method based on non-linear regression method for the teaching evaluation of vocal classroom of music education majors in higher education, and studied the problem of measuring and assessing the teaching results of music from the quantitative point of view; Peng [8] analysed the course process of music education by using the theory of OBE education, and put forward the music teaching evaluation method combined with simple machine learning algorithms; Jung [9] investigated the teaching evaluation method of national music culture transmission based on shallow network under multiculturalism, and analysed the evaluation model from various aspects such as cultural perspective and quantitative perspective; He and Liu [10] used the theory of multiple intelligences to construct the mapping relationship between the measurement value of the music teaching results and the teaching scores; Xu [11] researched the analysis and evaluation method of music teaching in combination with neural network based on the perspective of cultivating students' interests and hobbies; Wei et al. [12] studied the algorithm and assessment method of song arrangement generation in the field of music. Through the literature survey and network research analysis, the research on the measurement and assessment of higher music teaching results, although there have been a large number of academic results of measurement and a little music teaching evaluation system research, but there are still deficiencies, specifically in the following aspects [13]: 1) higher music teaching results measurement system is only limited to the measurement of the teacher's teaching process, ignoring the main body of teaching -- student feedback measurement; 2) higher music teaching results measurement system is only limited to the teacher's teaching process measurement, ignoring the main body of teaching --The feedback measurement of students; 3) The quantification of higher music teaching achievement measurement system is not objective enough, and the quantification of each index is comparable; 4) The higher music teaching achievement assessment method fails to portray the non-linear relationship between the measurement value and

the assessment value; 5) The higher music teaching achievement assessment method based on the neural network algorithm is prone to fall into the local optimum.

Neural networks are an algorithm used to create nonlinear mapping relationships. They are known for their simple structure and quick optimization convergence. Neural networks are commonly used for classification and prediction tasks, as well as in areas like network intrusion detection, charge prediction, machinery fault diagnosis, and condition assessment [14]. As the number of input parameters increases, the optimization convergence of neural networks may easily become stuck in a local optimum. However, the use of intelligent optimization techniques can enhance the speed and accuracy of convergence in neural networks [15].

This work addresses the issues related to measuring and assessing the achievements in higher music education teaching. To tackle these challenges, the study introduces a technique that combines the BP neural network [16] and intelligent optimization algorithm [17]. This approach is based on the MRO-BP model and aims to measure and analyze music education teaching achievements. This paper examines the issue of measuring and assessing teaching outcomes in higher music education. It analyzes research concepts and important quantitative technical aspects related to measuring teaching outcomes. It also addresses the measurement of teaching outcomes by analyzing the teaching process, identifying measurement indicators, and constructing a measurement system. Additionally, it proposes a methodology for assessing teaching outcomes in higher music education by combining neural networks and the MRO algorithm [18]. This methodology is based on the MRO-BP model. A novel teaching accomplishment assessment method based on MRO-BP is developed. The experimental section used statistical research data on teaching successes in higher music education. Through comparison analysis, it was confirmed that the MRO-BP model outperformed other models in terms of assessment effectiveness, as well as enhancing assessment time and efficiency.

The paper begins by introducing the significance of assessment in music education and the application of intelligent disciplines. The methodology section details the use of BP neural networks combined with the MRO algorithm to enhance the accuracy and efficiency of outcome assessments. Key techniques include process analysis, measurement of teaching achievements, and application of assessment models. The research explores the development of the MRO-BP network model, comparing it with other algorithms to demonstrate its effectiveness. The simulation and analysis section describes the data acquisition process, experimental environment, and parameter settings, followed by a performance comparison of different models. The conclusion highlights the model's superior results in accuracy and time efficiency, suggesting its applicability in higher music education while noting the need for further validation on other datasets.

II. MEASUREMENT AND ASSESSMENT

A. Analysis of Research Ideas

When teaching advanced music, instructors use many approaches, levels, and formats in the classroom [19]. This paper aims to address the issue of measuring and assessing the outcomes of higher music teaching. We focus on various aspects such as the course teaching process, teachers' level, students' knowledge demand, course practicability, and social value (Fig. 1). To achieve this, we conducted a detailed questionnaire survey to evaluate the cognitive abilities of music teachers. We then extracted indicators of the outcomes of the higher music teaching process and developed a measurement system for these outcomes. Additionally, we utilized the heterogeneous coupling optimization of the machine learning method to establish a mapping relationship between the measured values and assessment scores of higher music teaching. To fully analyze higher music teaching results, the study aims to establish a correlation between music teaching outcome measures and assessment scores. This particular research proposal is shown in Fig. 2.

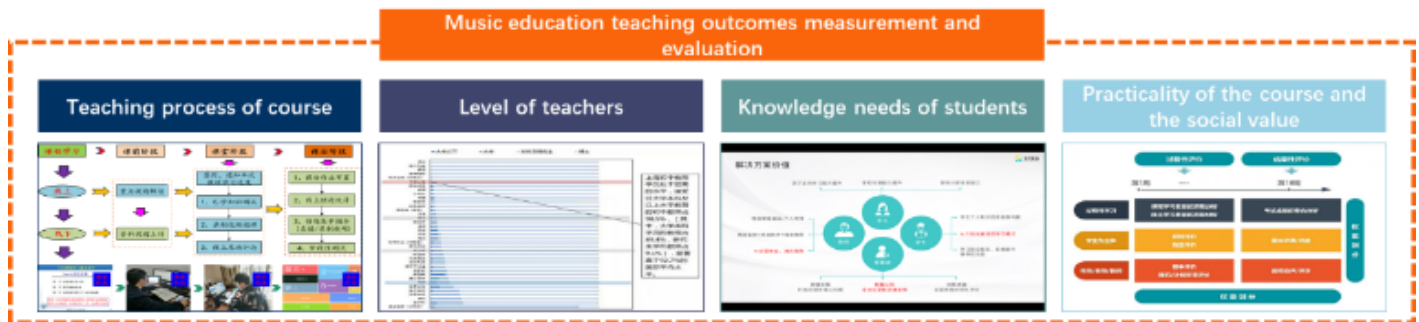


Fig. 1. Domains of measurement and assessment of teaching outcomes in higher music education.

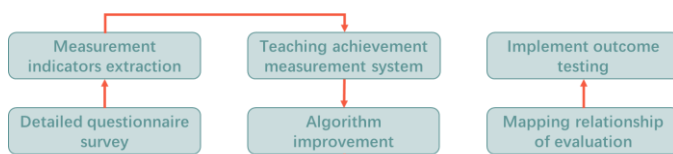


Fig. 2. Research ideas on measurement and assessment issues of teaching outcomes in higher music education.

B. Analysis of Key Technologies

This paper explores the measurement and assessment of teaching outcomes in higher music education, focusing on problem analysis, measurement of outcomes, assessment of outcomes, and application of a model. Specifically, it examines key technologies related to the analysis of the teaching process in music education, measurement of teaching outcomes in music, construction of an assessment model for teaching

outcomes, and application of the assessment model. These aspects are illustrated in Fig. 3.

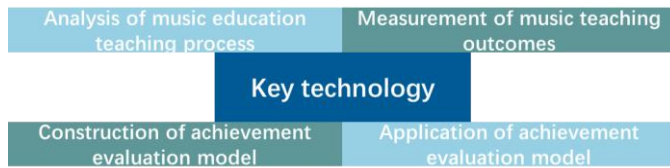


Fig. 3. Key techniques in the methodology for measuring and evaluating teaching outcomes in higher music education.

1) *Process analysis techniques*: In order to sort out the problem of higher education teaching outcome assessment, process analysis technique was proposed, mainly by analysing the process of higher education teaching outcome measurement and assessment, and carrying out process analysis from questionnaire survey, demand analysis, method design and other aspects [20], as shown in Fig. 4.

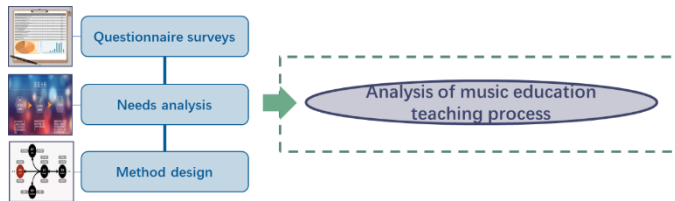


Fig. 4. Process of analysing the measurement and assessment of teaching outcomes in higher music education.

2) *Techniques for measuring teaching achievement*: Higher education music education teaching outcome measurement (as shown in Fig. 5) mainly assesses and measures the teaching outcomes of the pedagogues from the aspects of teaching content, practical skills and social practice ability. The input of the module is the results of the analysis of the process of measuring and assessing teaching outcomes in higher education, the assessment aspects, and the output is the teaching outcome measurement system.

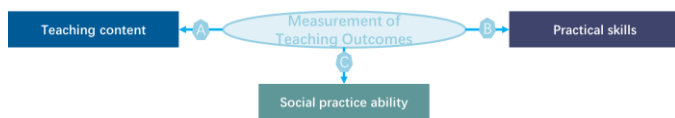


Fig. 5. Measurement of teaching outcomes in higher music education.

3) *Teaching achievement assessment techniques*: Higher Music Education Teaching Achievement Assessment (shown in Fig. 6) mainly combines a variety of intelligent algorithms to construct the mapping relationship between music education teaching measurements and assessment values. The input of this module is music education teaching measurement data, and the output is teaching outcome assessment scores.

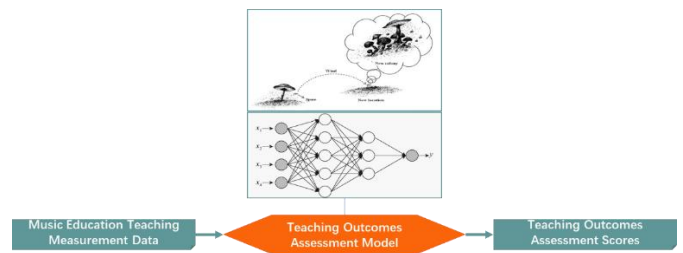


Fig. 6. Assessment of teaching outcomes in higher music education.

4) *Techniques for applying results-based assessment models*: Taking the music education teaching outcome data of higher education institutions as a case study, the trained outcome assessment model was applied to the data, and the teaching outcome measurements were collected, standardised and input into the outcome assessment model to obtain the music education teaching outcome assessment scores, as shown in Fig. 7.

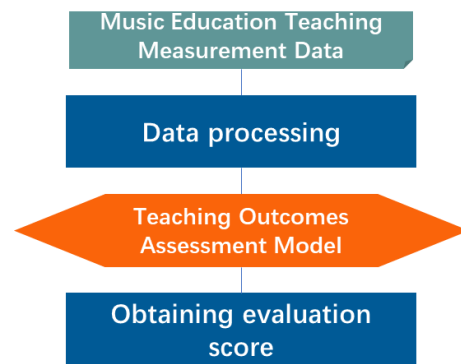


Fig. 7. Assessment of teaching outcomes in higher music education.

III. TEACHING OUTCOMES IN HIGHER MUSIC EDUCATION

According to the principles of demand-orientation, scientific, systematic and quantitative [21], this paper selects the measurement values of teaching results from three aspects, such as teaching content, practical skills and social practice ability [22], and the detailed extraction of the measurement values is shown in Fig. 8.

- Teaching content measures include measurements of regular grades, classroom practices, and final grades.
- Skills practice measurement includes academic salon activities focusing on professional skills demonstration and academic thinking exchange, and classroom practice activities focusing on teaching process design.
- Measurement of social practice ability includes internship assessment in off-campus teaching practice bases, and participation in various competitions for music education majors and teaching.

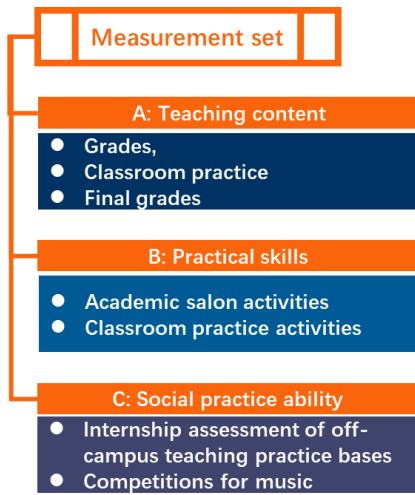


Fig. 8. Detailed extraction of the measurement values.

IV. RESEARCH ON THE ALGORITHM FOR EVALUATING MUSIC TEACHING ACHIEVEMENT BASED ON MRO-BP NETWORKS

A. BP Neural Network

BP neural network [23], or back-propagation neural network, is a multilayer feed-forward network that is widely used in the fields of function approximation, pattern recognition, classification, data compression, and time series prediction, as shown in Fig. 9.

The core of the BP network lies in its weight adjustment method, using the error back propagation algorithm, which adjusts the network weights to optimise the model performance by calculating the output error and back propagating it to each implicit layer, the structure of which is shown in Fig. 10 and Fig. 11.

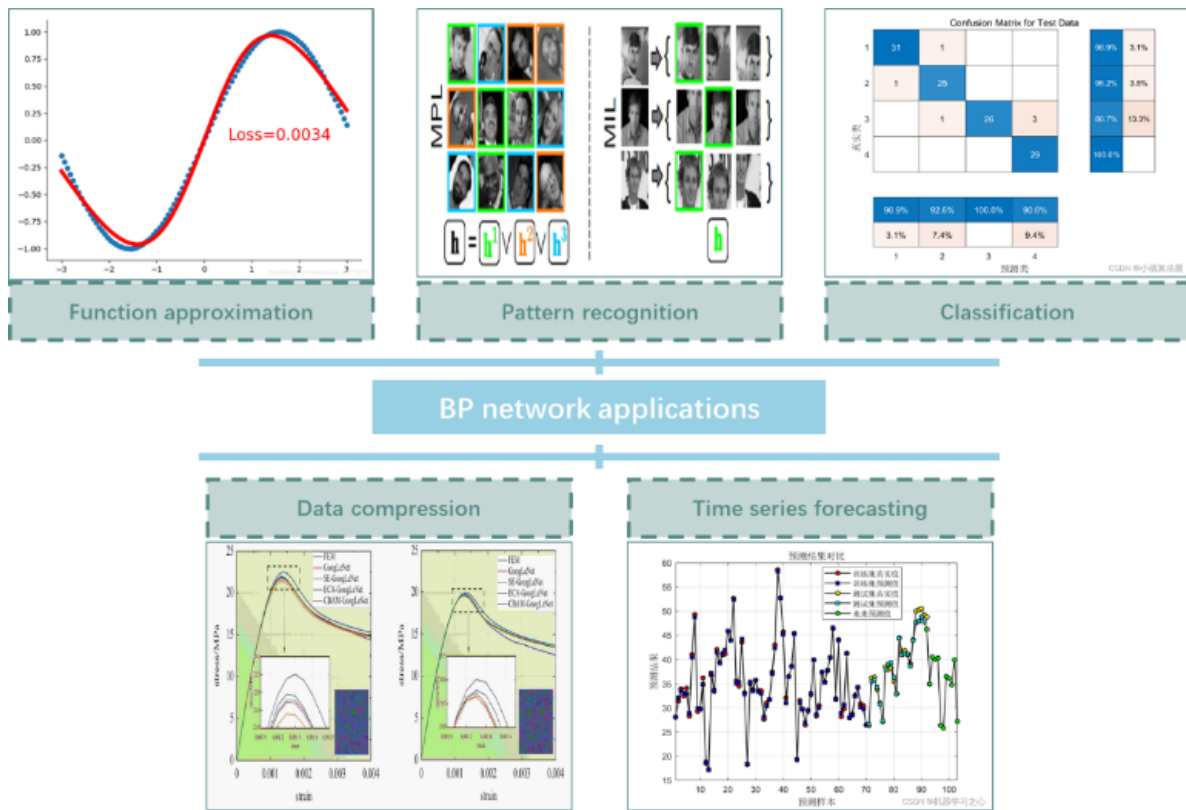


Fig. 9. BP neural network application.

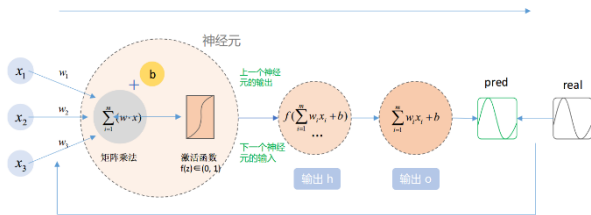


Fig. 10. Structure of BP neural network.

The basic structure of BP network includes input, hidden and output layers, with full connection between layers and no

connection between the same layers. The learning process of BP network includes two phases of forward propagation of signals and back propagation of errors. 1) In forward propagation, the input signals are transmitted through the network and generate predicted values in the output layer. 2) In back propagation, according to the error between predicted values and the actual target values, the error is reduced by calculating the error gradient to adjust the weights and thresholds of the network to reduce the error.

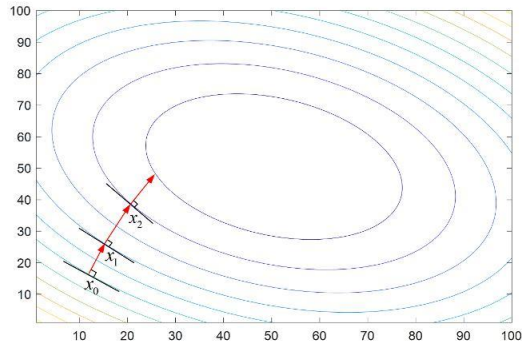


Fig. 11. Gradient descent method.

B. MRO-BP Network Model

1) *MRO algorithm*: Mushroom Reproduction Optimization (MRO) [24] is a population intelligence optimisation algorithm inspired by the mechanisms of mushroom growth and reproduction in nature. The algorithm mimics mushrooms exploring the reproduction region through spore propagation and finding a more optimal reproduction region by refining the search space (Fig. 12). The MRO algorithm decides whether to perform a local search or a global search by calculating the average fitness value of each colony as well as the average fitness value of all the colonies to improve the efficiency of the search and to avoid precocious convergence to the local optimal solution.



Fig. 12. Iterative process of mushroom propagation.

The flow pseudo-code of the MRO algorithm is shown in Table I:

TABLE I. PSEUDO-CODE OF THE MRO ALGORITHM

Algorithm 1: Mushroom Reproduction Optimisation Algorithm
M parent mushroom populations were randomly generated; Calculate the initialised mushroom population fitness value, with the average fitness value, and update the optimal mushroom individuals;
While the iteration condition is satisfied
For i=1:M
If Ave+Tave/c<Tave
An artificial wind is used to disperse the population, select the optimal solution, and update the optimal solution;
End if
Randomly mutate the population, compute Ave, select the optimal solution, and update the optimal solution;
End for
Calculate Tave
End while

The specific implementation steps of the MRO algorithm (Fig. 13) are as follows:

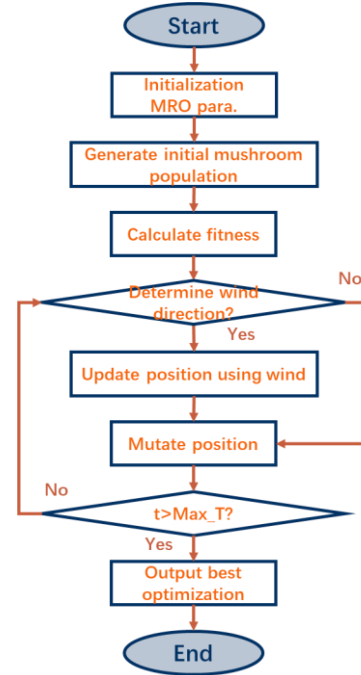


Fig. 13. Flowchart of MRO algorithm.

- 1) Initialise the parameters of the MRO algorithm with the population. Initialise the maximum number of iterations T_{max} , the population size $npop$, and the population individual search range. Initialise the population position:

$$X_{ij} = rand() \times (ub - lb) + lb \quad (1)$$

where ub and lb are the upper and lower boundaries of the mushroom search space, respectively.

- 2) Artificial wind determination of conditions. Artificial wind was operated on mushroom individuals that met the conditions, otherwise mushroom individuals were randomly searched within the breeding area.

$$Avg(i) + \frac{T_{Avg}}{c} > T_{Avg} \quad (2)$$

Where, $Avg(i)$ is the average fitness value of the i^{th} mushroom individual, T_{Avg} denotes the global average fitness value, and c is the coefficient of determination.

- 3) Artificial wind mechanism. The simulated artificial wind operation on individual mushrooms makes the algorithm with global optimisation seeking capability.

$$Mov_j^{wind} = (X_i^* - X_k^*) \times \left(\frac{Avg(i)}{T_{avg}} \right)^m \times Rand(-\delta, \delta) \times rs + Rand(-r, r) \quad (3)$$

Where, Mov_j^{wind} denotes the distance moved by the j^{th} individual, X_i^* is the mushroom individual with the best fitness value in the i^{th} colony, and X_k^* denotes the path with the best fitness value in the k^{th} colony. $Avg(i)$ is the average fitness value value of the i^{th} colony, $Tavg$ is the average fitness value of all colonies. m is the customisation coefficient. δ is the direction coefficient. r is the step length control coefficient.

- 4) Breeding area search. Individual mushrooms search for better adapted locations within nearby breeding areas.

$$X_{ij} = X_i^{parent} + \overline{Rand(-r, r)} \quad (4)$$

Where X_{ij} is the location of the mushroom individual, i is the number of this individual in the population, and j is the mushroom dimension. r is the random search radius. X_i^{parent} denotes the parent mushroom.

- 5) Finding the optimal fitness value individual and updating the mushroom location.

$$[bestmushroom] = \min(f(X_i)) \quad (5)$$

- 6) Determine whether the maximum number of iterations is reached, if the maximum number of iterations is reached, end the loop and output the result, otherwise carry out the next iteration.

The MRO method has superior capabilities in both global search and local refinement, making it well-suited for addressing intricate optimization issues. It may preserve population variety while searching, enabling it to escape local optimum solutions and discover the global optimal solution or a superior solution, as seen in Fig. 14. The MRO method has been used to address challenges in several domains, such as engineering design optimization and data mining [25].

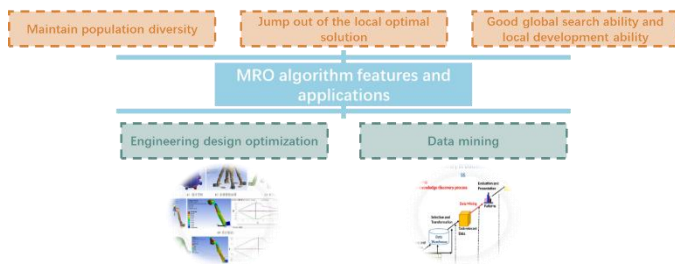


Fig. 14. MRO algorithm characteristics and applications.

2) *MRO-BP network*: In this paper, the BP network structure parameter weights and biases are used as decision variables, and the mean square error between the evaluated value and the true value is used as the fitness value of the MRO-BP model, the specific structure is shown in Fig. 15, and its pseudo-code is shown in Table II.

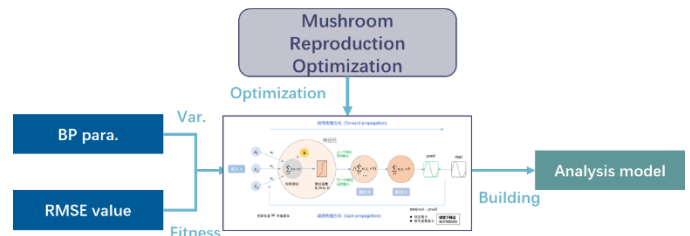


Fig. 15. MRO-BP network structure.

TABLE II. PSEUDO-CODE OF MRO-BP NETWORK ALGORITHM

Algorithm 2: MRO-BP network pseudo-code

The MRO algorithm parameters are set, the MRO optimisation decision variables are identified, and the BP weights and biases are encoded in real numbers;
 The RMSE is calculated as the fitness value to update the optimal mushroom individual, i.e., the current optimal BP network parameters;
 Whether the While iteration condition is satisfied
 Calculation of simulated artificial wind operations on individual mushrooms based on artificial wind determination conditions;
 Mutational manipulation of individual mushrooms using breeding area search;
 Update the network parameter individual information as well as the optimal network parameter individual;
 End while
 Output optimal network parameters;
 Constructing the MRO-BP network.

C. Improved Network Modelling Applications

From Fig. 16, the MRO algorithm is employed in this paper to enhance the accuracy of the BP network model in the assessment of teaching achievement in higher music education. Additionally, the measurement system of teaching achievement in higher music education is investigated. The precise sequence of actions in the procedure is as follows:

- Through the examination of the challenges associated with measuring and evaluating the outcomes of higher music education, we develop a framework for analyzing the teaching outcomes of higher music education based on three key dimensions: teaching content, practical skills, and social practice ability.
- Acquire the measurement data for evaluating the outcomes of higher music education. Utilize the distinctive indicators of the measurement system to input the data into the MRO-BP model. Train and optimize the model to get an assessment model for evaluating the outcomes of higher music education.
- Choose validation data to get improved measures of music education teaching outcomes, input them into the MRO-BP network-based model, and generate better scores for assessing music education teaching outcomes.

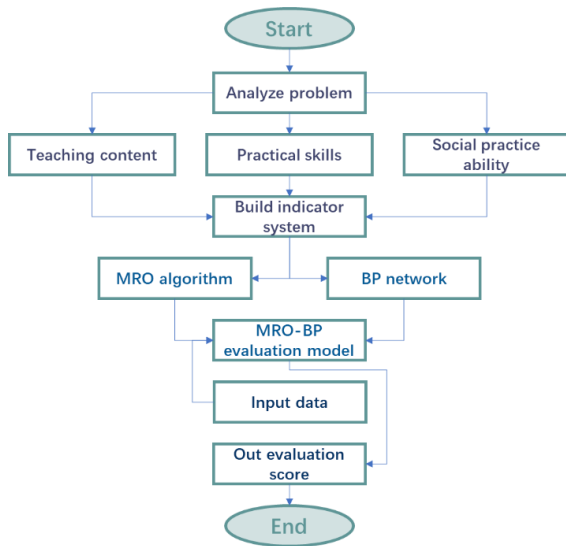


Fig. 16. MRO-BP network model application.

V. SIMULATION AND ANALYSIS

A. Experimental Data Acquisition

The cognitive assessment of music learners was conducted by a comprehensive questionnaire, which took into consideration several factors such as the teaching method, the proficiency of instructors, the knowledge requirements of students, and the practicality and social significance of the course. The survey findings were used to extract the higher music education teaching outcome measurement data. This data was then split into three sets: the MRO-BP model training set, validation set, and test set. The particular division ratio, number, and purpose of each set are provided in Table III.

B. Experimental Environmental Setup

The specific settings of hardware environment, software environment and other experimental environments used for algorithm verification in this paper are shown in Table IV.

TABLE III. EXPERIMENTAL DATA SETTINGS

Serial number	Data set	Proportions	Quantities	Goal
1	test set	15%	552	Testing the evaluation performance of the MRO-BP model
2	validation set	15%	553	Calculating the fitness value of the MRO algorithm for optimising BP networks
3	training set	70 per cent	2580	Training the optimal BP network model, i.e. MRO-BP model

TABLE IV. EXPERIMENTAL ENVIRONMENTAL SETTINGS

Name of the environment	Parameterisation
software	AMD Ryzen 9 5900HX with Radeon Graphics 3.30 GHz
operating system	Windows 10
programming software	Python 3.8
visualisation software	Matlab2021a

C. Contrast Algorithm Parameter Settings

Methods for measuring and assessing teaching outcomes in higher music education employing comparison algorithms such as BP, TLBO-BP, GWO-BP, MPA-BP, AVOA-BP, and MRO-BP. The BP model utilizes the gradient descent method to calculate error feedback. The number of nodes in the hidden

layer is determined based on the analysis in section 5.4. The TLBO [24], GWO [25], MPA [26], AVOA [27], and MRO algorithms have a maximum iteration limit of 1000, and the number of populations is determined based on the experiments in Section V (D). The specific parameter settings for the comparison algorithms can be found in Table V.

TABLE V. COMPARISON OF ALGORITHM PARAMETER SETTINGS

Serial number	Arithmetic	Parameterisation
1	BP	The activation function is a radial basis function
2	TLBO-BP	TF=round[(1+rand)]
3	GWO-BP	a decreases linearly from 2 to 0
4	MPA-BP	P = 0.5, R is a uniformly distributed random vector, FADs = 0.2, U = 0 or 1
5	AVOA-BP	L1 = 0.8, L2 = 0.2, w = 2.5, P1 = 0.6, P2 = 0.4, P3 = 0.3
6	MRO-BP	Fmin=0.07, Fmax=0.75, τ=4.125, a0=6.25, a1=100, a2=0.0005

D. Analysis of Results

1) *Parameter setting analysis*: To ensure that the parameters are in accordance with the BP and optimization algorithms, this paper evaluates the accuracy and time consumption of the assessment models of teaching outcomes in

higher music education using varying numbers of hidden layer nodes and populations. The results are illustrated in Fig. 17, Fig. 18, Fig. 19 and Fig. 20. According to the Fig. 17. From the data, it is evident that as the number of hidden layer nodes in the BP algorithm increases, the accuracy of the measurement and

assessment of music education teaching outcomes in each model initially decreases and then stabilizes. Additionally, the time required for the assessment model also increases. As the population size increases, the precision of the algorithm used to measure and assess the outcomes of higher music education teaching improves. However, after reaching a population size of 75, the accuracy remains stable and does not decrease. Additionally, the time required for each model to process the data increases as the population size increases. The paper's detailed study reveals that the number of hidden layer nodes in the BP algorithm is 90, and each optimization technique has a population size of 75.

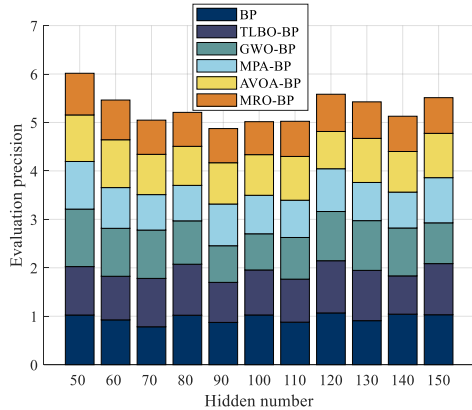


Fig. 17. Accuracy analysis of the evaluation model based on different number of hidden layer nodes.

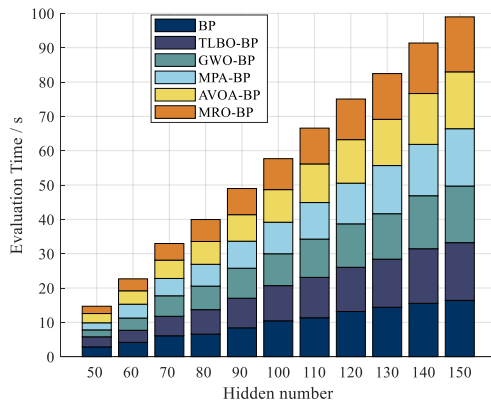


Fig. 18. Time-consuming analysis of the evaluation model based on different number of hidden layer nodes.

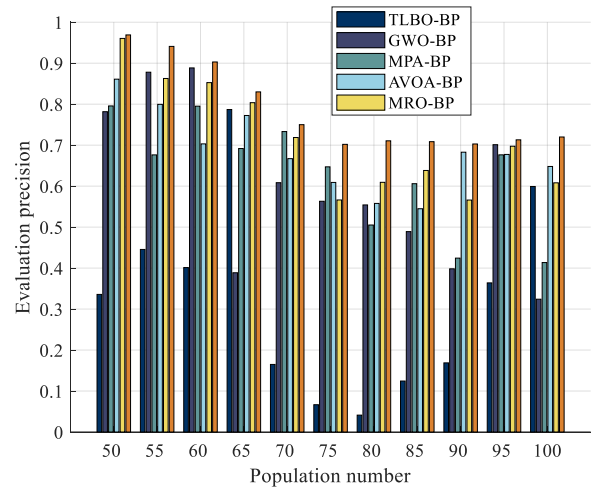


Fig. 19. Accuracy analysis of the assessment model based on different population sizes.

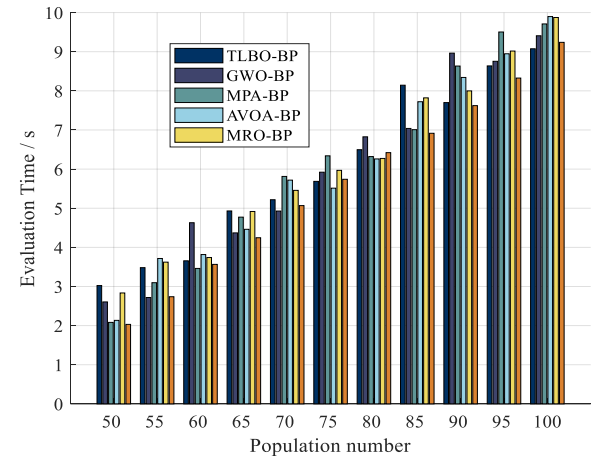


Fig. 20. Time-consuming analysis of assessment models based on different population sizes.

2) *Evaluation performance analysis:* To analyze and compare the effectiveness of the measurement and evaluation methods for teaching outcomes in higher music education proposed in this paper, we utilized several techniques: BP, TLBO-BP, GWO-BP, MPA-BP, AVOA-BP, and MRO-BP. The comparative analysis results are presented in Fig. 21 and Table VI.

According to the figure. Based on the data, it is evident that the MRO-BP network has the highest convergence accuracy for measuring and assessing teaching outcomes in higher music education. It is followed by AVOA-BP, MPA-BP, GWO-BP, and TLBO-BP. Additionally, the speed of convergence for each optimization network in measuring and assessing teaching outcomes in higher music education is approximately equal.

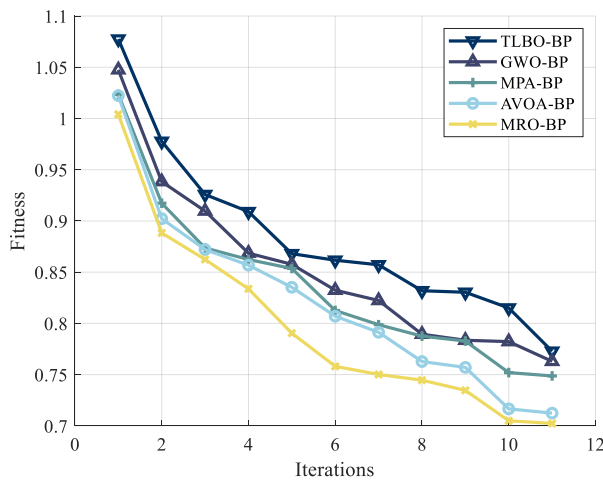


Fig. 21. Optimisation convergence curves of different optimisation algorithms.

Table VI shows that the MRO-BP network-based measurement and assessment of teaching outcomes in higher music education has the smallest RMSE value, followed by GWO-BP, MPA-BP, AVOA-BP, TLBO-BP, and BP. In terms of MAPE, the MRO-BP network-based measurement and assessment of teaching outcomes in higher music education has the smallest MAPE value of 0.7314, followed by AVOA-BP, MPA-BP, GWO-BP, TLBO-BP, and BP. The MRO-BP model also has the smallest MAE value of 0.49, followed by AVOA-BP, GWO-BP, MPA-BP, BP, and TLBO-BP. In terms of time-consuming, the MRO-BP model takes 7.744, followed by AVOA-BP, MPA-BP, GWO-BP, TLBO-BP, and BP. In summary, the MRO-BP model is the most accurate and requires the least amount of time.

TABLE VI. RESULTS OF PERFORMANCE COMPARISON OF DIFFERENT ASSESSMENT MODELS

arithmetic	RMSE	MAPE	MAE	Time/s
BP	0.892	0.9342	0.78	8.822
TLBO-BP	0.829	0.8561	0.91	8.729
GWO-BP	0.728	0.8373	0.63	8.397
MPA-BP	0.731	0.7783	0.71	7.758
AVOA-BP	0.737	0.7761	0.61	7.647
MRO-BP	0.701	0.7314	0.49	7.744

VI. CONCLUSION

As music education teaching data continues to grow, the analysis of music behavior data and the assessment of music teaching outcomes have emerged as key areas of focus in the field of data-driven music research. This study addresses the issue of measuring and evaluating the teaching outcomes of data-driven algorithmic applications. It provides a technique for measuring and evaluating the teaching outcomes of higher music education using a combination of the MRO algorithm and BP network, known as the MRO-BP network model. This paper examines the issue of measuring and evaluating the teaching outcomes of music education. It focuses on designing the main technology and identifying the measurement indicators for teaching outcomes from three perspectives: teaching content,

practical skills, and social practice ability. The paper constructs a measurement system for music teaching outcomes by combining the MRO algorithm and BP network. Additionally, it proposes an algorithm for assessing the teaching outcomes of music based on the MRO-BP network. The method proposed in this paper demonstrates superior results in terms of RMSE, MAPE, MAE, and time consumption when analyzing teaching outcome data in higher education. It effectively addresses the challenge of measuring and assessing teaching outcomes. The MRO-BP network model outperforms other models and is applicable to analyzing teaching outcomes in higher education, specifically in music education. However, its generalization and stability should be further validated using other datasets. Subsequently, it is necessary to use the MRO-BP network model for addressing further difficulties and enhancing the overall performance of the MRO-BP network.

REFERENCES

- [1] Wang Y, Teng Y. Research on the New Mode Teaching of "Three Dimensional Five Movements" Music Appreciation Course. *Creative Education*, 2024, 15(4):9.
- [2] Shu Zhang. Application status and optimisation strategy of artificial intelligence in music education. *Journal of Wuhu Institute of Vocational Technology*, 2024, 26(02):80-82+88.
- [3] Du B. The Introduction of Popular Music in Music Education Teaching in Vocational Colleges. *Journal of Education and Culture Studies*, 2022.
- [4] Yu H, Zou Z. The music education and teaching innovation using blockchain technology supported by artificial intelligence. *International Journal of Grid and Utility Computing*, 2023.
- [5] Chen M, Mohammadi M, Izadpanah S. Language learning through music on the academic achievement, creative thinking, and self-esteem of the English as a foreign language (EFL) learners. *Acta Psychologica*, 2024, 247.
- [6] Yang Y. Diversified Teaching Method in Basic Piano Course Instruction for Higher Education Teachers. *Contemporary Education Research (Baitu)*, 2023, 7(3):1-7.
- [7] Liao M, Huang F, Hawamdeh S. Music Education Teaching Quality Evaluation System Based on Convolutional Neural Network. *Journal of Information & Knowledge Management*, 2024.
- [8] Peng F. Music copywriting and the problems of music education: overcoming prohibitions and the use of music in teaching. *Research*, 2022, 25:1-12.
- [9] Jung E J. A Study on the Teaching and Learning Method of Goryeogayo in Music Education. *Korean Association For Learner-Centred Curriculum And Instruction*, 2022.
- [10] He L, Liu H. A music main melody extraction algorithm based on multi-feature fusion and compressed excitation model. *Computer Application and Software*, 2023, 40(5):160-166.
- [11] Xu D. Study on the Application of Experiential Teaching in Secondary Vocational Music Education. *Foreign Language Edition: Educational Science*, 2022(2):194-197.
- [12] Wei M, Jiang W, Hu X. Residual storey drift estimation of the MDOF system with the weak storey under seismic excitations using the BP network. *Structures*, 2023.
- [13] Li Z. Application of the BP Neural Network Model of Gray Relational Analysis in Economic Management. *Journal of Mathematics*, 2022.
- [14] Xu B, Yuan X. A Novel Method of BP Neural Network Based Green Building Design-The Case of Hotel Buildings in Hot Summer and Cold Winter Region of China. *Sustainability*, 2022, 14.
- [15] Li S C, Wu J F. Indoor positioning based on intelligent optimisation algorithm and its optimised BP neural network. *Science Technology and Engineering*, 2024, 24(20):8568-8576.
- [16] Zhang Y H, Zhang H Z, Ma L, Zhu N. Recycling chain shop site selection problem and mushroom propagation algorithm solution. *Journal of Shanghai University of Technology*, 2023, 45(04):405-414.

- [17] Ren H, Xie F. The Application of Multiple Music Cultures in College Music Teaching in the Background of Internet. *Applied Mathematics and Nonlinear Sciences*, 2024, 9(1).
- [18] Qiao F. Vocal Music Education: the Reference and Application of Choral Aesthetic Education to Music Education in Colleges and Universities. *Curriculum and Teaching Methodology*, 2023.
- [19] Yin X L. Educational Innovation of Piano Teaching Course in Universities. *Education and Information Technologies*, 2023:1-16.
- [20] Yoo H. Teaching Traditional and Transformed Versions of Culturally Diverse Musics With Integrity. *Journal of General Music Education*, 2023.
- [21] Yang Y. Research on Blended Teaching Evaluation Model of College English Based on Entropy Method and BP Neural Network. *Open Access Library Journal*, 2024, 11(6):9.
- [22] Liu F, Zhang H Z, Zhou X. Hybrid discrete mushroom propagation algorithm for open site selection path problem with fuzzy demand. *Computer Application Research*, 2021, 38(03):738-744+750.
- [23] Bidar M, Mouhoub M, Sadaoui S, Kanan H R. A Novel Nature-Inspired Technique Based on Mushroom Reproduction for Constraint Solving and Optimization. *International Journal of Computational Intelligence and Applications*, 2020, 19(2):2050010.
- [24] Wang P C, Feng H J, Li L R. An optimisation algorithm for teaching and learning based on adaptive competitive learning. *Computer Applications*, 2023, 43(12):3868-3874.
- [25] Feng C, Tang L Y. Turbine blade feature extraction based on GWO-VMD. *Journal of Harbin University of Commerce (Natural Science Edition)*, 2024, 40(04):387-396.
- [26] Liu Y H, Song Y B, Zhu D P. Fault diagnosis method for rolling bearings based on ELDA dimensionality reduction and MPA-SVM. *Noise and Vibration Control*, 2024, 44 (03): 117-124 (in Chinese)
- [27] Chen Q Y, Shao J, Wang C Q, Chen L, Tai Xingyu Improved African Vulture Optimization Algorithm Based on Dual Dynamic Adjustment. *Foreign Electronic Measurement Technology*, 2024, 43 (01): 20-29 (in Chinese)

Employing Data-Driven NOA-LSSVM Algorithm for Indoor Spatial Environment Design

A Case Study of Physical Bookstores

Di Wang, Hui Ma*, Tingting Lv

College of Art and Design, Jilin Jianzhu University, Changchun 130118, Jilin, China

Abstract—This study aims to enhance the precision and efficiency of indoor spatial design for college physical bookstores in the context of the new media environment. To achieve this, a novel intelligent analysis model was developed by integrating the Navigator Optimization Algorithm (NOA) with the Least Squares Support Vector Machine (LSSVM). The research analyzes the relationship between the new media environment and bookstore design, identifies key design principles, and establishes performance metrics. The proposed NOA-LSSVM model optimizes design parameters by utilizing a hybrid convergence-divergence search mechanism, achieving improved accuracy and computational efficiency. A case study of Jilin Jianzhu University's bookstore was conducted to evaluate the model's performance. The NOA-LSSVM model was compared with three other optimization algorithms: the Flower Pollination Algorithm (FPA), Whale Optimization Algorithm (WOA), and Sine Cosine Algorithm (SCA). Results showed that the NOA-LSSVM model achieved superior accuracy, with a Mean Absolute Percentage Error (MAPE) of 2.9, significantly lower than FPA (4.6), WOA (3.8), and SCA (4.2). Additionally, the model exhibited faster convergence and enhanced design efficiency, optimizing the bookstore's functional zones and spatial layout to balance dynamic and quiet areas effectively. In conclusion, the NOA-LSSVM model demonstrates a robust capability to optimize indoor spatial design in the new media environment, outperforming traditional methods in accuracy and practicality. This study provides valuable insights for integrating intelligent algorithms into spatial design processes, with the potential for broader applications in other commercial or educational spaces. Future research should focus on extending the model's generalizability and incorporating advanced media technologies for enhanced user experiences.

Keywords—New media environments; data-driven algorithms; indoor spatial environment design; mariner optimization method

I. INTRODUCTION

In order to enhance students' learning materials and services, the growth of college physical bookshops is being actively supported to strengthen their impact in the area of education [1]. The proliferation and use of new media, facilitated by the advancement of Internet technology and the exponential growth of Internet users, has presented unparalleled prospects and difficulties across all sectors of society. Consequently, physical bookshops are also confronted with the need for modernization and adaptation. To adapt to the growing market need, physical bookshops are integrating new media technologies to expedite their transition and bolster their competitiveness [2]. Conducting research on the interior space

design of college physical bookstores in the new media environment may stimulate innovation in bookstore space design and enhance the quality of services. Additionally, it can serve as a valuable resource for the growth of campus bookstore markets and industry transformation [3].

The study on the interior space design of college physical bookshops in the new media environment is presently in its first phase. It primarily focuses on two aspects: the use of new media technology in the design of physical shops, and the analysis of how new media technology is applied [4]. Wang and Wang [5] examine the successful collaboration between college physical bookstores and publishers in order to address the high cost of textbooks. Sari [6] introduces new design concepts to modify the layout of college physical bookstores. Bae [7] investigates the current state of development of college physical bookstores and proposes effective strategies to renovate them. Soureshjani et al. [8] suggests several reform measures to transform bookstores into vibrant spaces that promote a new culture of wisdom and reading. Al-Ansari et al. [9] thoroughly discusses the appeal of college campus bookstores and presents a new development strategy that is suitable for the digital era. Nyboer [10] analyzes the challenges from various perspectives such as cultural experience, construction mode, and business model, and provides a range of practical solutions. The integration of intelligent algorithms and the interior space design of college physical bookshops in the new media environment has become an essential approach for the future development of intelligent and digital space design. Intelligent algorithms may be categorized as supervised learning, unsupervised learning, semi-supervised learning, and other ways. The utilization of data-driven intelligent algorithms in analyzing the application of new media technology in indoor space design has become increasingly prevalent. This method plays a crucial role in researching the design of college physical bookstores in the new media environment. It enhances the efficiency of indoor space design and expedites the design process [11]. Despite the ongoing improvement in educational conditions and the increasing importance of campus physical bookstores, there are still several issues in the process of integrating intelligent technology into the interior space design of college bookstores. These include: 1) a lack of research on the application of new media technology in the interior space design of college bookstores; 2) ineffective utilization of data generated during the design process of college bookstores; and 3) the immaturity of intelligent technology used in college bookstore interior space design.

*Corresponding Author

This paper strives to address the intellectual requirements of college students in physical bookstores by integrating new media technology and artificial intelligence. It proposes an application analysis model for the interior space design of college physical bookstores in the new media environment, using an improved data-driven algorithm. The study investigates the correlation between the new media environment and the layout of indoor spaces in bookstores. It takes advantage of various research methods such as literature research, survey research, interdisciplinary research, and inductive comparison. The study analyzes the design process of indoor spaces in college physical bookstores and develops an intelligent application analysis scheme for such spaces. This scheme combines the navigator optimization algorithm with the LSSVM model. Additionally, the study proposes a new media model based on the NOA-LSSVM model. A novel media-based NOA-LSSVM model is offered as an analytical tool for the interior space design of college physical bookshops, using the Voyager optimization algorithm and the LSSVM model. Using the physical bookshop of Jilin Jianzhu University as a case study, the effectiveness of the proposed NOA-LSSVM model is examined and evaluated by comparing it with other application analysis methodologies.

This paper is structured as follows: Section II explores the relationship between the new media environment and the physical layout of college bookstores, providing a detailed analysis of the design process and identifying key principles and metrics for indoor spatial design. Section III introduces the NOA-LSSVM model, detailing the principles of the Navigator Optimization Algorithm (NOA) and Least Squares Support Vector Machine (LSSVM), and describes how these are integrated into a data-driven intelligent analysis framework. Section IV presents a case study of Jilin Jianzhu University's bookstore, demonstrating the model's application, design outcomes, and a comparative evaluation of its performance against other optimization algorithms. Finally, Section V concludes with key findings, highlighting the NOA-LSSVM model's superior accuracy and efficiency in optimizing design processes while outlining potential areas for further research and practical implementation.

II. DESIGN INTERIOR SPACE OF COLLEGE PHYSICAL BOOKSTORES

A. The Connection between the Contemporary Media Landscape and the Physical Layout of Bookstores

The emergence of new media complements the growth of physical bookshops. Its implementation impacts individuals' perception of the reading experience, enhances the interior spatial setting of bookstores, and maximizes the creation of a favorable reading ambiance [12]. The indoor space design of college bookstores has undergone significant changes in response to the new media environment. These changes primarily include: 1) improving the overall spatial experience; 2) diversifying the nature of the space; and 3) enhancing the spatial environment, as depicted in Fig. 1.

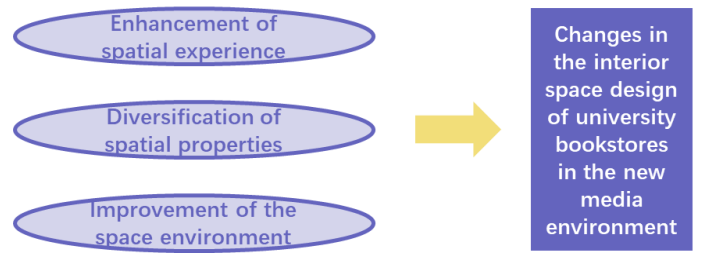


Fig. 1. Changes in the new media environment on the interior space design of college bookstores.

B. Analysis of the Interior Space Design Process of College Physical Bookstores

1) *The key aspects of designing the interior space of college physical bookshops:* The need for college physical bookshops among college students has seen significant changes with the progress of time. The interior space design of these bookstores has evolved to focus on utility, industry diversification, experience elements, and artistic aspects [13], as shown in Fig. 2.



Fig. 2. Trends in the design of physical bookstores in universities.

2) *Analysis of the design process:* The interior space design of college physical bookstores follows specific development direction and design principles, namely the humane principle, functional principle, experiential principle, and epochal principle (as depicted in Fig. 4). This design process includes various steps such as space planning and layout, illumination design, color matching and style, display and exhibition design, furniture and decorations selection, air-conditioning and ventilation system, sound and noise control, online and offline combination, out-of-store time and budget control, detail design, and others [14] (as highlighted in Fig. 3).

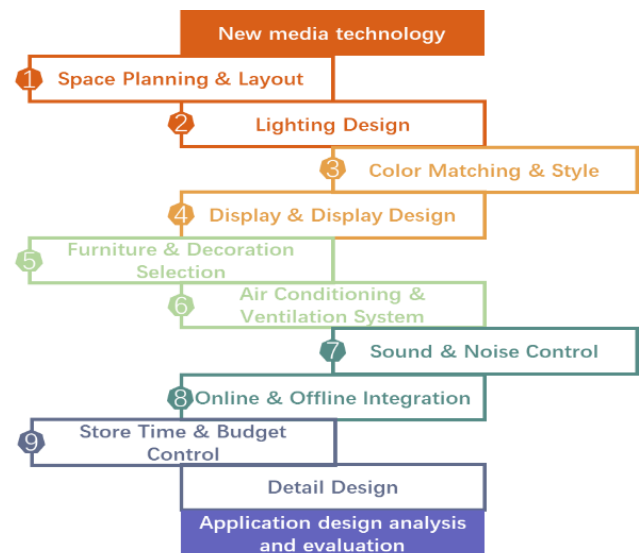


Fig. 3. Design process of physical bookstores in universities.

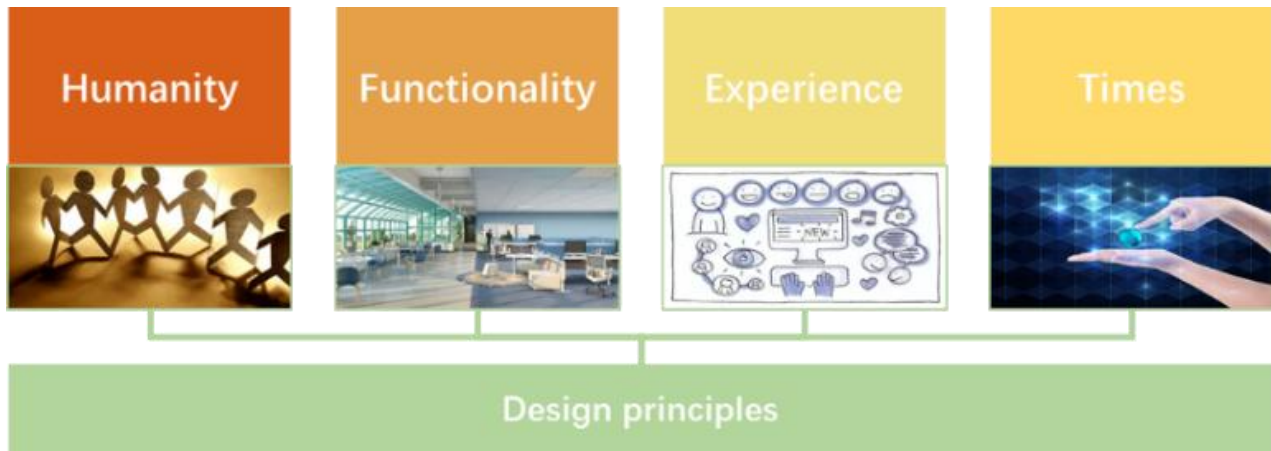


Fig. 4. Design principles.

3) *Develop a process to obtain metrics for analyzing application design:* This paper tests the application analysis indexes for indoor space design of college physical bookstores in the new media environment. The analysis focuses on three aspects: design elements S, functional space K, and design research Y, as illustrated in Fig. 5.

Spatial design application analysis metrics		
Design element S	Functional space K	Design study Y
<ul style="list-style-type: none"> <input type="checkbox"/> Space size S1 <input type="checkbox"/> Space furnishings S2 <input type="checkbox"/> Material selection S3 <input type="checkbox"/> Color selection S4 <input type="checkbox"/> Lighting design S5 <input type="checkbox"/> Plant configuration S6 	<ul style="list-style-type: none"> <input type="checkbox"/> Multi-purpose space K1 <input type="checkbox"/> Virtual space K2 	<ul style="list-style-type: none"> <input type="checkbox"/> Personalized service Y1 <input type="checkbox"/> Interactive experience Y2 <input type="checkbox"/> Online and offline integration Y3

Fig. 5. Indicator analysis of the application of interior space design of college physical bookstores in the new media environment.

- Design element S comprises of space size S1, space furnishings S2, material selection S3, colour selection S4, lighting design S5 (Fig. 6), and plant setup S6 (Fig. 7);



Fig. 6. Lighting design.



Fig. 7. Plant configuration.

- The functional space K comprises of two components: multifunctional space K1 and virtual space K2;
- Design study Y comprises of personalized service Y1, interactive experience Y2, and the integration of offline and online platforms Y3.

C. Program for Designing the Interior Space of Physical Bookstores in Higher Education Institutions

Focusing on the problem of application analysis of interior space design of college physical bookstores in media environment, this paper proposes a method of application analysis of interior space design of college physical bookstores based on data-driven algorithm, and the specific design scheme is shown in Fig. 8. The scheme analyzes the process of interior space design of college physical bookstores in media environment, extracts relevant application analysis indexes, collects application analysis data, standardizes and annotates the dataset, combines Voyager optimization algorithm [15] and LSSVM model [16], constructs the application analysis model of interior space design of college physical bookstores in media environment, and carries out the performance validation and analysis of the model by using examples.

According to the design scheme, the research on the application analysis method for the interior space design of college physical bookstores in the media environment includes key technologies such as the extraction and construction of application analysis indexes, data collection and pre-processing, design model construction and optimisation, and case validation and analysis, as shown in Fig. 9.

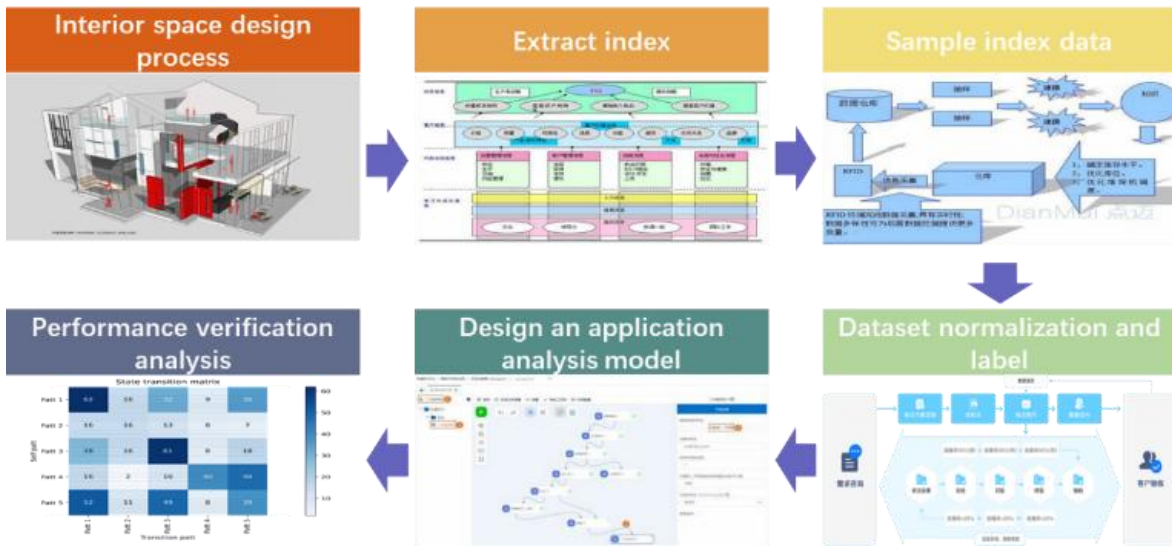


Fig. 8. Applied analysis of interior space design of physical bookstores in higher education design scheme.

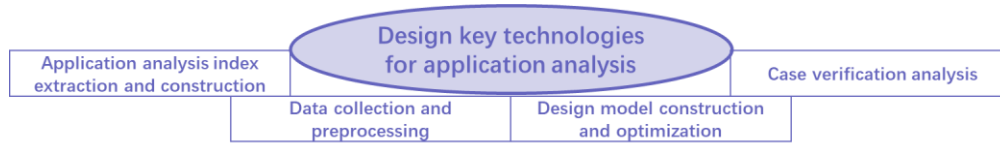


Fig. 9. Key techniques for applying analysis to interior space design of physical bookstores in universities.

III. INTELLIGENT ANALYSIS ALGORITHM FOR INTERIOR SPACE DESIGN

A. NOA-LSSVM Model

1) *Voyager optimisation algorithm*: Navigator optimization algorithm (NOA) [15] is a new type of meta-heuristic intelligent algorithm inspired by the exploratory behaviour of navigators. The NOA algorithm alternates between "searching" and "exploiting". The NOA algorithm alternates between "searching" and "exploiting", when the search period is even, the navigators search for the solution by divergence; when the search period is odd, the navigators search for the solution by convergence, and find the optimal solution by alternating iterations.

The navigator position is a candidate solution and the expression is:

$$P = \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1d} \\ P_{21} & P_{22} & \cdots & P_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ P_{n1} & P_{n2} & \cdots & P_{nd} \end{bmatrix} \quad (1)$$

where n is the number of navigators and d is the number of decision variables.

The navigator adaptation values are expressed as follows:

$$F = [F_1, F_2, \dots, F_n]^T \quad (2)$$

a) *Convergence mode*: The position update formula for the convergent search performed by each navigator during the convergence cycle is as follows:

$$P_{ij}^{k+1} = P_{gj}^k + D_{ij} \cdot e^{bt} \cdot \cos(2\pi t) \quad (3)$$

$$D_{ij} = |P_{gj} - P_{ij}| \quad (4)$$

Where, k is the number of iterations; D_{ij} is the distance between the position of mariner i and the current optimal position in the j^{th} dimension; b is the shape coefficient of the solenoid, which takes the value of 1; and t is a random number. Logarithmic spiral is shown in Fig. 10.

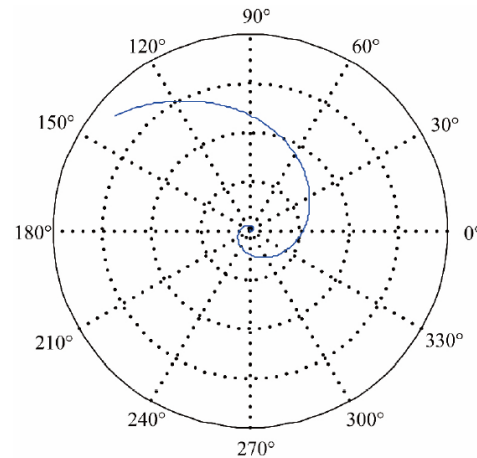


Fig. 10. Logarithmic spiral.

b) *Divergent approach*: The NOA algorithm uses the Levy flight strategy [17] as a model for dispersion, and the specific dispersion model is as follows:

$$P_i^{k+1} = \begin{cases} P_i^k + rL(d)(P_g^k - P_i^k) & k < N_{iter}/2 \\ P_i^k + r(L(d) \cdot P_i^k - P_i^k) & k \geq N_{iter}/2 \end{cases} \quad (5)$$

Where r is a constant, set to 0.7; $L(d)$ is the Levy flight step (Levy flight trajectory in 3D space is shown in Fig. 11); N_{iter} is the maximum number of iterations.

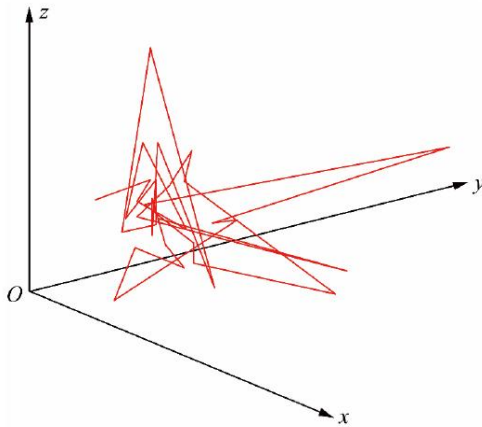


Fig. 11. Levy flight path in 3D space.

c) *Search cycle*: The search cycle is an important tool used by the NOA algorithm to alternate between "searching" and "utilising". If the number of search cycles is set too large, the number of iterations in each search cycle will be too small, and the navigators will not be able to fully diverge or converge; if it is set too small and similar to the idea of traditional intelligent algorithms of searching for excellence, the navigators will not be able to effectively "use" the "search". If the setting is too small, it is similar to the idea of traditional intelligent algorithms, and the "search" of navigators cannot be effectively "utilised". After testing, when the number of search cycles to take the maximum number of iterations $1/30 \sim 1/10$, you can achieve better results, the number of search cycles in this paper to take $1/20$ of the number of iterations.

d) *Transboundary processing techniques*: In order to avoid the navigator to enter outside the boundary, the NOA algorithm adopts an out-of-bounds processing technique. When the navigator crosses the boundary and $z < p$, its boundary crossing processing technique is as follows:

$$P_{ij} = \begin{cases} P_{j\max} & P_{ij} \geq P_{j\max} \\ P_{j\min} & P_{ij} < P_{j\min} \end{cases} \quad (6)$$

where P_{ij} is the position of the j th dimension of the i th navigator; $P_{j\max}$ and $P_{j\min}$ are the upper and lower bounds of

the j th dimension of the navigator, respectively; p is a constant set to 0.5; and z is a random number.

When a navigator crosses the border and $z \geq p$, his or her crossing is handled with the following technique:

$$P_{ij} = \begin{cases} P_{j\max} - Ce_1(P_{j\max} - P_{j\min}) & P_{ij} \geq P_{j\max} \\ P_{j\min} + Ce_1(P_{j\max} - P_{j\min}) & P_{ij} < P_{j\min} \end{cases} \quad (7)$$

Where C is a constant with a value of 0.01 and e_1 is a random number.

According to the optimisation strategy of the NOA algorithm, the pseudo-code of the NOA algorithm is shown in Table I with the following steps:

- Step 1: Initialise the number of navigators, dimensions and the number of search cycles to generate the initial position of the navigator;
- Step 2: Calculate the initial navigator fitness value and update the optimal fitness value;
- Step 3: At each odd search cycle, update the mariner position using convergence; at each even cycle, update the mariner position using divergence;
- Step 4: During each iteration, compare the navigator fitness value with the global optimum and update the global optimum;
- Step 5: Interpret whether the maximum number of iterations is reached, if so output the optimal solution, otherwise jump to step 3.

TABLE I. PSEUDO-CODE OF THE NOA ALGORITHM

Algorithm 1: NOA algorithm	
1	Initialize navigator number, dimension, search periods;
2	Generate navigator population;
3	Calculate fitness and output best fitness;
4	For t=1:Max iter
5	If t== odd search periods
6	Use convergence to update the navigator's position,
7	Else
8	Use divergence to update the navigator position;
9	End
10	Update navigator's position;
11	Bound position using upper and lower limits;
12	Update best navigator position;
13	End
14	Output best solution.

2) *LSSVM algorithm*: Least Squares Support Vector Machine (LSSVM) [18] A variant of Support Vector Machine

(SVM) [19]. Compared with the traditional SVM, LSSVM simplifies the computational process by solving the model parameters through the least squares method, which transforms the optimisation problem into the solution of a system of linear equations. LSSVM is not only suitable for classification problems, but also widely used in regression problems. It has high computational efficiency and good generalisation ability, which is especially suitable for dealing with large-scale datasets.

a) *LSSVM basic principles*: The core idea of LSSVM is to use the least squares method to solve the parameters of the SVM model by eliminating the Lagrange multipliers in the form of minimising the sum of squares of the errors and transforming the original convex quadratic programming problem into a system of linear equations. This system of linear equations can be solved by numerical methods (e.g., Cholesky decomposition, iterative methods, etc.) to obtain the parameters of the model. The output of the LSSVM model is obtained by a linear combination of the kernel functions in a high-dimensional space, as shown in Fig. 12.

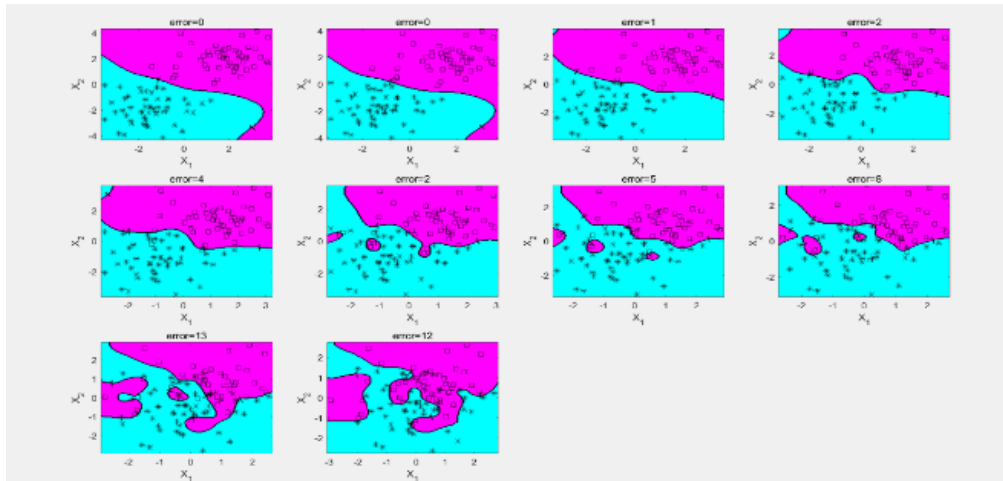


Fig. 12. Structure of LSSVM algorithm.

b) *Features of LSSVM*: The idea of LSSVM is characterized by the following features: a) great computing efficiency; b) resilience to noisy data; c) simplicity in finding parameter solutions; d) application to nonlinear problems; and e) absence of sparsity [20], as seen in Fig. 13.

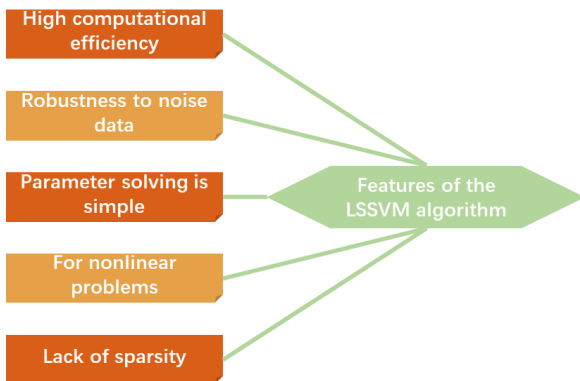


Fig. 13. Characteristics of the LSSVM algorithm.

c) *Utilization of least squares support vector machines (LSSVM)*: The Least Squares Support Vector Machine (LSSVM) is used in several domains (Fig. 14), such as financial market analysis, medical diagnosis, bioinformatics, image analysis, industrial control, and machine learning. Due to its high computing capacity and strong generalization abilities, it has become a very effective tool for tackling real-world issues [21].

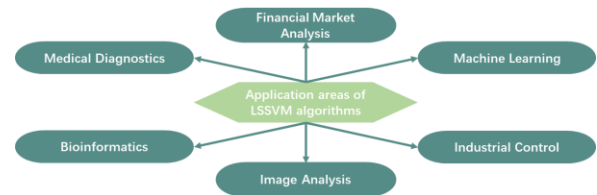


Fig. 14. Application areas of LSSVM algorithm.

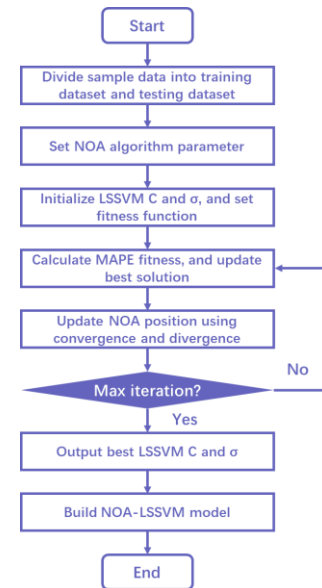


Fig. 15. Flowchart of NOA-LSSVM algorithm application analysis.

3) *NOA-LSSVM model*: This work employs the NOA algorithm to enhance the accuracy of application analysis in the LSSVM model. The algorithm optimizes the parameters of the LSSVM, namely the penalty coefficient and kernel bandwidth, using the MAPE as the fitness value function. The NOA algorithm is employed to search for optimization through both convergence and divergence modes. The specific flow chart of the application analysis using the NOA-LSSVM model is depicted in Fig. 15.

B. Application of NOA-LSSVM Model in Intelligent Analysis

Combined with NOA-LSSVM model, this paper proposes an intelligent analysis method of college physical bookstore indoor space design based on NOA-LSSVM model, and the specific application analysis diagram is presented in Fig. 16. Firstly, examine the relationship between the new media

environment and the indoor space of bookstores. Analyze the process of designing the indoor space of physical bookstores in colleges and develop a design scheme for the indoor space of physical bookstores in colleges. Extract the application analysis of the design of the indoor space of physical bookstores in colleges within the media environment. Secondly, collect the index data for the design of the indoor space of physical bookstores in colleges based on the established set of indices. Preprocess the input data, annotate it, and output the scores for the application analysis. Utilize the NOA-LSSVM algorithm to construct an intelligent analysis model for the interior space design of physical bookstores in colleges. Finally, use an example to verify the feasibility of the design scheme for the interior space of physical bookstores in colleges, as well as the efficiency of the intelligent analysis algorithm for interior space design based on the NOA-LSSVM model.

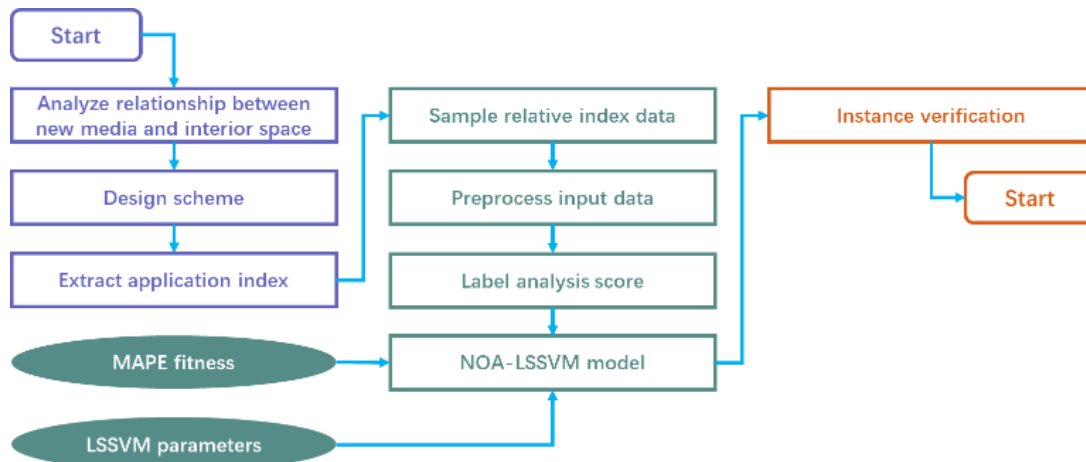


Fig. 16. Flow chart of intelligent analysis of interior space design for college physical bookstore based on NOA-LSSVM model.

IV. EXAMPLE ANALYSES

A. Presentation of the Case

In order to verify the feasibility of the interior space design scheme of college physical bookstore and the effectiveness and feasibility of the intelligent analysis algorithm of college physical bookstore interior space design based on NOA-LSSVM model, this paper takes the indoor space design of physical bookstore in the Jilin Jianzhu University as an example to be analysed.

The current layout is shooting to investigate the impact of the new media environment on college physical bookstore spaces. It focuses on the bookstore space design of Jilin Jianzhu University. The goal is to showcase the changes brought about by the integration of new media into various aspects of life, as well as the higher expectations of college teachers and students for campus bookstores. Following the principles of the era, experience, functionality, and humanization, the design aims to create a new type of college physical bookstore that combines book purchasing, cultural exchange, leisure, immersive reading, and experiential

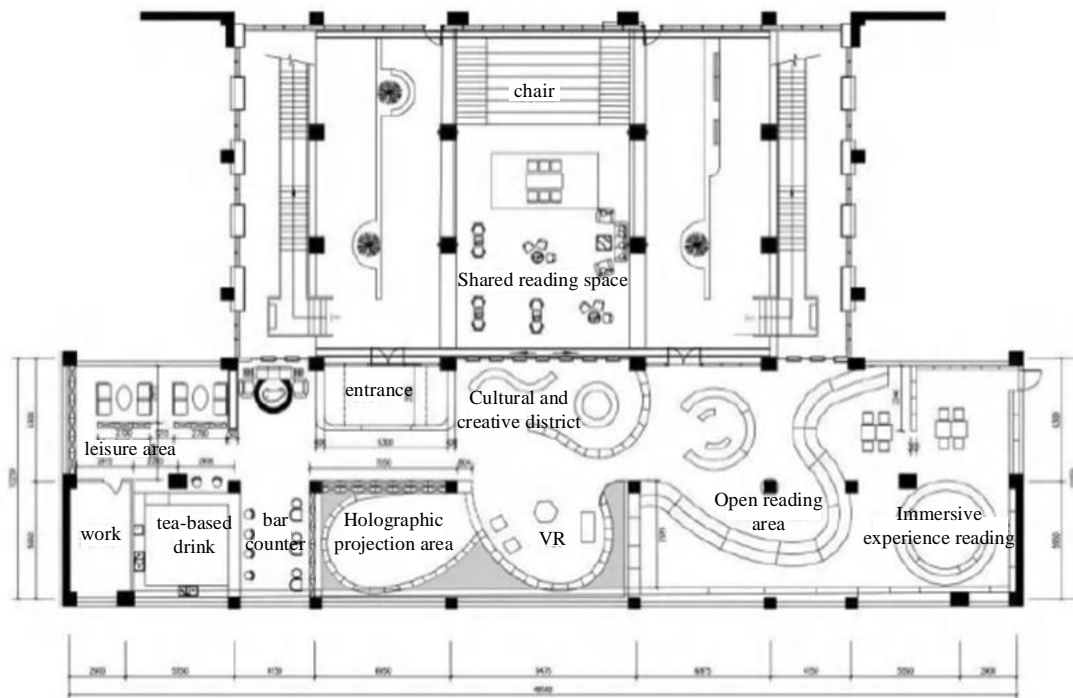
services. The bookstore is an innovative university physical shop that offers book purchasing, cultural interchange, leisure activities, immersive reading experiences, and experiential services.

B. Algorithm Configuration

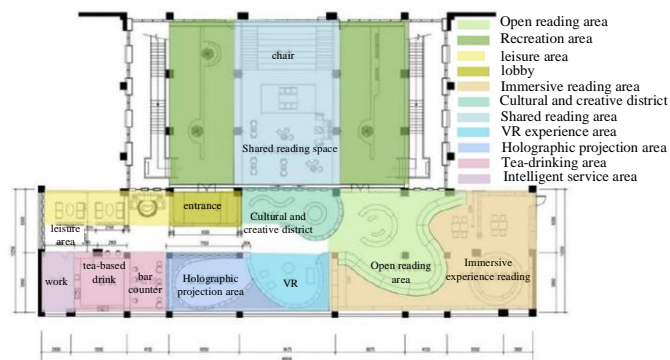
This work used the Flower Pollination Algorithm (FPA) [22], Whale Optimization Algorithm (WOA) [23], and Sine Cosine Algorithm (SCA) [24] as comparison algorithms to optimize the parameters of the LSSVM model. The optimization technique was iterated 500 times, with a population size of 50. The NOA algorithm underwent 25 search cycles, while the other parameters of the algorithm were established according to references [22-24].

C. Design Outcomes

According to the indoor space design method of college physical bookstore under the new media environment designed in this paper, this paper takes the physical bookstore of Jilin Jianzhu University as an example, and designs and analyses the indoor space plan layout and functional partition diagram, specifically as shown in Fig. 17.



(a) The arrangement and structure of an interior design plan.



(b) Interior design functional zoning plan.



(c) A top-down perspective

Fig. 17. Interior design.

The bookstore, depicted in Fig. 17, has a T-shaped layout that utilizes flexible curves and straight lines to separate the space. Curved bookshelves are strategically placed to direct the flow of the interior. The bookstore is divided into eleven functional zones, taking into account the balance between movement and stillness. The dynamic zone includes the

recreation area, shared reading area, tea area, and intelligent service area, while the other zones are designated as relatively quiet areas.

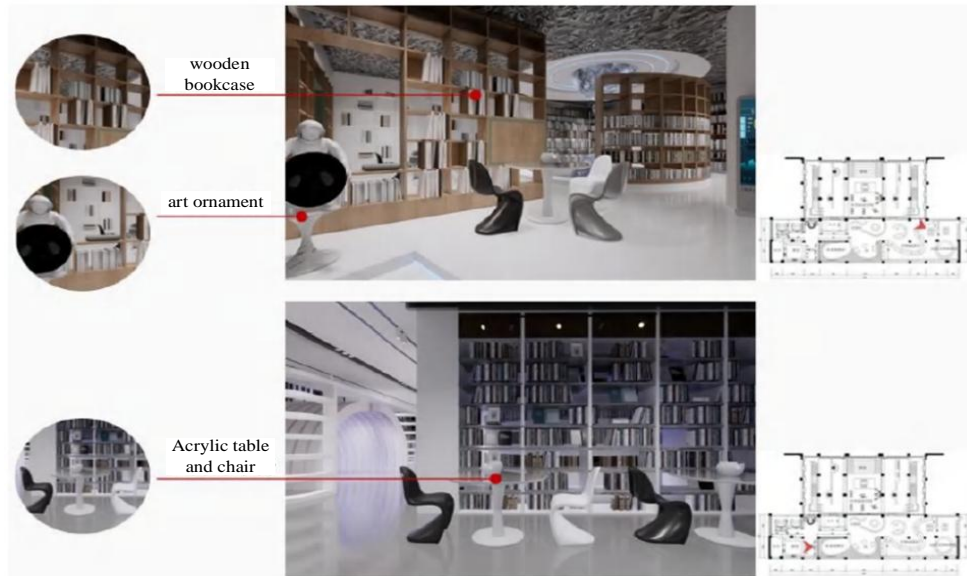
The physical bookstore's interior design is examined from the perspectives of reading spaces, shared reading spaces, open

reading spaces, virtual reality experiences, 3D holographic projection spaces, immersive reading spaces, foyer spaces, tea and beverage spaces, leisure spaces, cultural and creative spaces, and intelligent service spaces, as shown in Fig. 18.

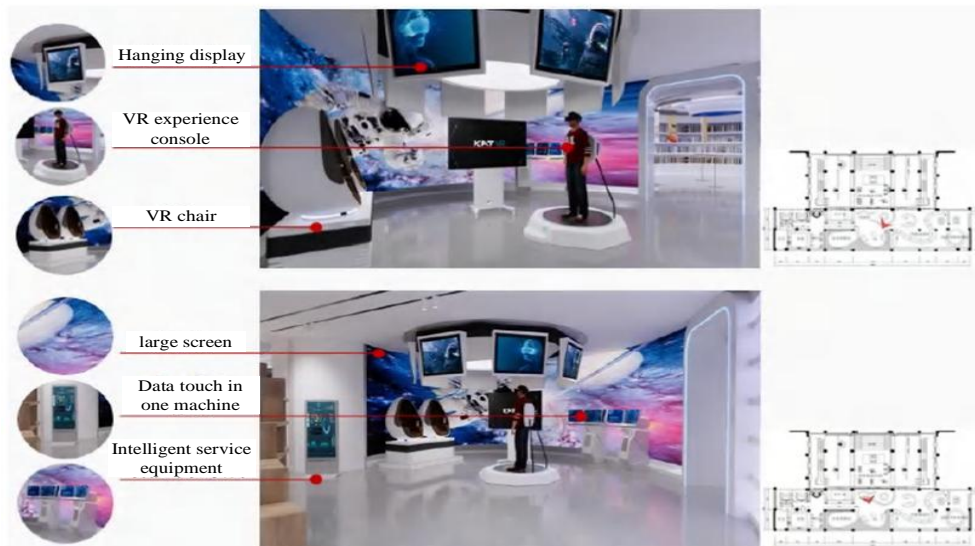
D. Evaluation of Algorithm Performance Outcomes

To evaluate the efficiency and superiority of the intelligent analysis algorithm for interior space design of college physical bookshops using the NOA-LSSVM model, this study randomly picks five test sets and presents the test results in Table II and Fig. 19.

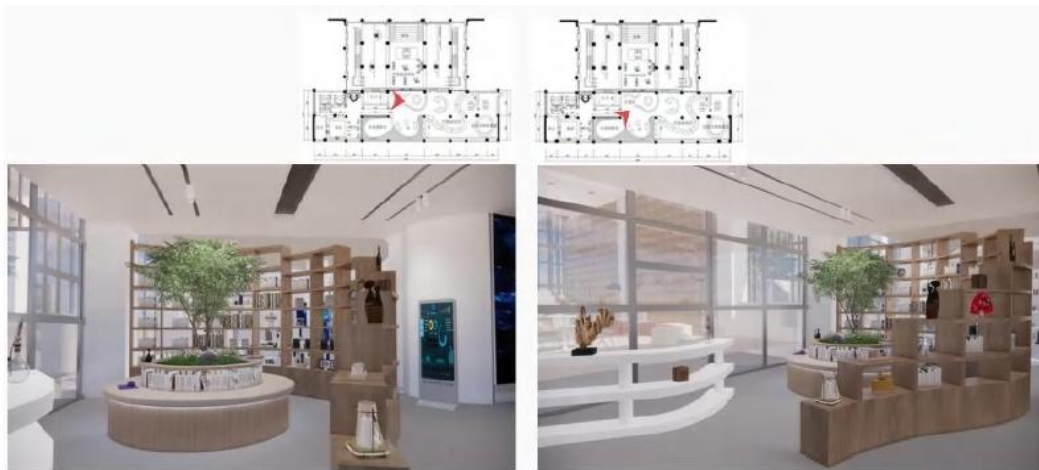
The outcomes of optimizing the LSSVM parameters of FPA, WOA, SCA, and NOA algorithms are shown in Table II, while the optimization curves of FPA, WOA, SCA, and NOA algorithms are displayed in Fig. 19. The figure demonstrates that the intelligent analysis algorithm for the interior space design of a college physical bookstore, based on the NOA-LSSVM model, exhibits faster convergence speed and superior convergence accuracy. The Mean Absolute Percentage Error (MAPE) achieved is approximately 2.9, which is lower than the MAPE values obtained by other algorithms such as FPA, WOA, and SCA.



(a) Reading space design.



(b) Designing the spatial experience for virtual reality.



(c) Designing cultural and creative venues.

Fig. 18. Interior design effect schematic diagram.

TABLE II. RESULTS OF DIFFERENT ALGORITHMS TO OPTIMISE LSSVM PARAMETERS

No.	LSSVM Parameters	FPA	WOA	SCA	NOA
1	C	102	200	160	120
2	σ	0.05	0.08	0.39	0.02

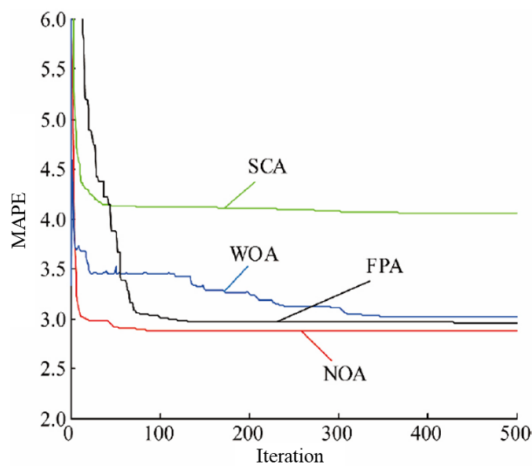


Fig. 19. Optimisation curves for different algorithms.

V. CONCLUSION

In this paper, we suggest a method for designing the interior space of a college physical bookstore that is based on the NOA-LSSVM model. This method involves analyzing the relationship between the new media environment and bookstore design, designing the interior space design scheme for the college physical bookstore under the new media environment, extracting the intelligent analysis indexes of the interior space design, combining the NOA algorithm to find the optimal parameters of the LSSVM, and establishing the NOA-LSSVM-based college physical bookstore indoor space design intelligent analysis model. The physical bookstore at Jilin Jianzhu University is used as a case study to analyze and compare other intelligent analysis models for interior space design. The results indicate that the NOA-LSSVM model has smaller MAPE results, higher analysis accuracy, and can

enhance the efficiency of interior space design for college physical bookstores in the new media environment.

The study demonstrates the effectiveness of the NOA-LSSVM model in optimizing indoor spatial design for college bookstores under the new media environment. However, it has some limitations. First, the model's generalizability remains uncertain as its application is only validated within a single case study, lacking tests in diverse bookstore types or cultural settings. Second, the research does not deeply explore user behavior and preferences, which are critical in designing user-centered spaces. Third, while emphasizing the importance of new media, the study provides limited details on integrating specific technologies such as AR/VR or social media into the design process. Future research should focus on extending the model to broader scenarios, integrating multi-source data like user behavior and new media interaction data for more comprehensive analyses, and exploring in-depth applications of advanced technologies to create interactive and immersive environments.

ACKNOWLEDGMENT

This work is supported by Research on the Construction of an Innovative Practical Teaching System for Environmental Design Major through Multidisciplinary Integration under the Background of 'Mass Entrepreneurship and Innovation'. Key project of the Jilin Provincial Education Science '14th Five-Year Plan' for 2023, Grant No.: ZD23009.

Key project of the Jilin Provincial Education Science '14th Five-Year Plan' for 2021: 'Research on the Pathways and Practices of Building a First-Class Major in Environmental Design under the Guidance of New Liberal Arts Development Concepts', Grant No.: ZD21035.

REFERENCES

- [1] Zhou Y. Discussion on the Application of VR Technology in Architectural Interior Design[J]. Architectural Engineering:Chinese-English Edition, 2022, 6(1):5-9.
- [2] Group M T. Interior design of EVs integrated with 3D fashion technology[J].MobileTex, 2023.

- [3] Enwin A, Ikiriko T D, Jonathan-Ihua G O. The Role of Colours in Interior Design of Liveable Spaces[J]. Sciences, 2023.
- [4] Shonk J. Your Space, Made Simple: Interior Design That's Approachable, Affordable, and Sustainable[J]. Library Journal, 2023.
- [5] Wang L, Wang Y. Research on the optimization of interior design of architectural space considering user perception[J]. Applied Mathematics and Nonlinear Sciences, 2024, 9(1).
- [6] Sari S M. Implementation of Interior Branding in Retail Interior Design[J]. GATR Journals, 2022.
- [7] Bae S, Asojo A O. Interior Environments in Long-Term Care Units From the Theory of Supportive Design:[J]. HERD: Health Environments Research & Design Journal, 2022, 15(2):233-247.
- [8] Soureshjani O K, Massumi A, Nouri G. Martian Buildings: design Loading[J]. Advances in Space Research, 2022.
- [9] Al-Ansari R A R, Alawad A, Hareri R. An Exploratory Study on the Effect of Indoor Lighting for Buildings on Light Pollution[J]. Art and Design Review., 2022.
- [10] Nyboer J. Critiquing contemporary interior design students[J]. International Journal of Technology and Design Education, 2024, 34(4):1579-1602.
- [11] Nowakowski P. Beauty and Utility in Architecture, Interior Design and in the New European Bauhaus Concepts[J]. buildings, 2024, 14(4).
- [12] Yu F, Liang B, Tang B W H. An Interactive Differential Evolution Algorithm Based on Backtracking Strategy Applied in Interior Layout Design[J]. algorithms, 2023, 16(6).
- [13] Tedjokoesoemo P E D. Interior Design Students' Perception on Interior Health and Comfort in Shop House Design for New Normal Era[J]. GATR Journals., 2022.
- [14] Zhang Z, Ban J. Aesthetic Evaluation of Interior Design Based on Visual Features[J]. International journal of mobile computing and multimedia communications, 2022.
- [15] Chestnut R, Zhang F, Jin B Y. Application of navigator optimisation algorithm in power system optimal tidal current calculation[J]. Power Construction, 2017, 38(06):7-14.
- [16] Liang J. Comparison of Price Prediction Based on LSTM, GRU, Random Forest, LSSVM and Linear Regression[J]. BCP Business & Management, 2023.
- [17] Sun C, Wang X, Jiang G Z. NSST image enhancement based on Levy-SOA adaptive threshold segmentation and improved bootstrap filtering[J]. Control Engineering, 2024, 31(07):1297-1304.
- [18] Song J H, Yue H. An investment estimation method for comprehensive pipeline corridor based on PCA-PSO-LSSVM[J]. Journal of Human University of Science and Technology (Natural Science Edition), 2024, 39(01):36-44.
- [19] Jia R Z. Research on the prediction of bending roll force of metallurgical plate cold continuous rolling based on PSO-SVM improved model[J]. Shanxi Metallurgy, 2023, 46(11):98-99.
- [20] Yin C L. Carbon emission measurement method for coal-fired generating units of thermal power plants based on LSSVM[J]. Journal of Chemical Industry, 2024, 38(03):5-8.
- [21] Dai Y J, Gao X G, Liu Z Y. A study on the improvement of the detection accuracy of Mn components in aluminium alloys by combining LASSO-LSSVM and laser-induced breakdown spectroscopy[J]. Spectroscopy and Spectral Analysis, 2024, 44(04):977-982.
- [22] Shi T, Xiong T, Zhao L Z. A review of flower pollination algorithm research[J]. Software Guide, 2023, 22(04):245-252.
- [23] Wu P. Soft measurement modelling method for penicillin fermentation process based on SPA-WOA-SVR[J]. Journal of Zhenjiang Higher Education, 2024, 37(03):88-93.
- [24] Chen Y G, Li G F, Chen Y G. Extraction of photovoltaic model parameters based on improved sine-cosine algorithm[J]. Journal of Xuchang College, 2024, 43(02):109-112.

Enhancing Customer Churn Prediction Across Industries: A Comparative Study of Ensemble Stacking and Traditional Classifiers

Nurul Nadzirah bt Adnan, Mohd Khalid Awang

Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, 22000 Tembil, Terengganu, Malaysia

Abstract—Predicting customer churn is essential in sectors such as banking, telecommunications, and retail, where retaining existing customers is more cost-effective than acquiring new ones. This paper proposes an enhanced ensemble stacking methodology to improve the prediction performance of ensemble methods. Classic ensemble classifiers and individual models are undergoing enhancements to enhance their sector-wide generalisation. The proposed ensemble stacking method is compared with well-known ensemble classifiers, including Random Forest, Gradient Boosting Machines (GBMs), AdaBoost, and CatBoost, alongside single classifiers such as Logistic Regression (LR), Decision Trees (DT), Naive Bayes (NB), Support Vector Machines (SVM), and Multi-Layer Perceptron. Performance evaluation employs accuracy, precision, recall, and AUC-ROC metrics, utilising datasets from telecom, retail, and banking sectors. This study highlights the importance of investigating ensemble stacking within these three business entities, given that each sector presents distinct challenges and data patterns related to customer churn prediction. According to the results, when compared to other ensemble approaches and single classifiers, the ensemble stacking method achieves better generality and accuracy. The stacking method uses a meta-learner in conjunction with numerous base classifiers to improve model performance and make it adaptable to new domains. This study proves that the ensemble stacking method can accurately anticipate customer turnover and can be used in different industries. It gives firms a great way to keep their clients.

Keywords—Customer churn; single classifier; ensemble classifier; stacking; accuracy

I. INTRODUCTION

Churning customers, which happens when someone stops using a product or service, is a big problem for businesses. This is especially true in fields that rely on steady streams of income. Business that depends on keeping people for a long time, like retail, banking, and telecommunications, are hit hard by churn.

The telecommunications business has one of the highest turnover rates because of fierce competition, quickly changing technologies, and a wide range of choices for customers. When it comes to telecommunications, service quality, pricing strategies, customer happiness, contract terms, network issues, and competitive offers are the main things that affect turnover. To keep customers, telecommunications companies need to accurately predict customer turnover, since getting new users is much more expensive. Telecommunications firms must effectively forecast customer turnover to retain clientele, as the acquisition of new subscribers is significantly more costly as Nurulhuda & Ling Sook Lew et al, 2021 [1] employed.

Comprehending the intricacies of client behavior, especially regarding service disruptions and pricing alterations, is crucial for formulating effective retention strategies.

In retail, churn is affected by factors including purchase frequency, order value, brand loyalty, product diversity, customer service, personalized offers, and the entire shopping experience. Gülmüş Börühan Karaca et al, 2022 [2] employed Retailers encounter the difficulty of comprehending intricate consumer behavior patterns and forecasting churn to improve retention via loyalty programs, targeted discounts, and personalized communication. Retail churn is notably influenced by seasonal trends, promotional activities, and economic conditions affecting consumer expenditure.

In the banking industry, customers leave because of things like bad customer service, high fees, a lack of personalized financial products, problems with digital transformation, and other banks' competitive goods research by Salma, Mohamed Roushdy & Amr Galal et al 2023 [3]. In banking, churn is strongly connected to how much customers trust and value the bank. This means that predicting churn is important for keeping customers loyal and managing customer relationships. To improve service offerings and customer engagement strategies in this highly regulated climate, it's important to know why customers leave.

When businesses can accurately guess which customers will leave, they can use targeted retention strategies to keep those customers. This increases profits by keeping valuable customers from leaving. Single classifiers, like Decision Trees (DT), Logistic Regression (LG), Support Vector Machines (SVM), Naive Bayes (NB), and Multi-Layer Perceptron (MLP), have been used in the past to identify churn. A lot of people use these methods because they are easy to understand. For example, Decision Trees make it easy to understand how to make a choice because they show the clearest way. Logistic Regression helps us understand how factors are related in a straight line, while Naive Bayes works well with large datasets [4]. A neural network called MLP is very good at finding non-linear patterns in data. This is especially helpful for predicting churn as employed by Huang et al, 2023 [5]. But these single classifiers have a hard time with complicated, noisy datasets that are common in fields like banking and telecommunications where customer behavior is hard to predict.

Because single models have their flaws, ensemble classifiers have come up as strong options. Ensemble methods [6], such as Random Forest (RF), Gradient Boosting (GBM), AdaBoost,

CatBoost, and Stacking, take the best parts of several algorithms and combine them to make predictions that are more accurate and reliable [7]. AdaBoost improves performance by changing the weights of weak classifiers over and over again, and CatBoost is great at working with categorical factors, which makes it good for big, complicated datasets [8]. Stacking combines several base models and improves their output through a meta-model. This gives better generalization and predictive power [9]. According to studies, ensemble classifiers work better than single classifiers in many situations, especially in fields with complicated customer data structures as employed by Sharma & Gupta, 2022 [10] and Sahar F. Sabbeh, 2018 [11].

Single classifiers like Decision Trees, Logistic Regression, SVM, Naive Bayes, and MLP are compared to ensemble methods like Random Forest, Gradient Boosting, AdaBoost, CatBoost, and Stacking for customer churn prediction. Ensemble methods use multiple models to improve predictive accuracy and robustness. Ensemble methods like Random Forest and Gradient Boosting reduce overfitting, increase generalisation, and capture complicated data patterns, making customer churn predictions across industries more accurate. The study examines these models with banking, retail, and telecoms data. Performance is measured by accuracy, precision, recall, and AUC-ROC. Ensemble techniques, particularly Stacking, will be tested to determine if they outperform single classifiers across industries and disclose [11] as top customer churn models. This study aims to address the following research questions: (1) How can stacking methods improve customer churn prediction across various industries? (2) What are the limitations of traditional classifiers, and how does the proposed approach overcome them? The remainder of this paper is organized as follows: Section II reviews the relevant literature on churn prediction methods. Section III details the methodology used, including the datasets and models. Section IV presents the results, while Section V discusses these findings. Finally, Section VI concludes with recommendations and future work.

II. LITERATURE REVIEW

Many companies, particularly those in the banking, retail, and telecommunications industries, place a great degree of significance on the research field of developing forecasts on customer turnover. This emphasis is particularly prevalent in the banking industry. To a large extent, the ability to maintain relationships with existing clients is one of the most critical variables that defines the profitability of these industries. A wide range of machine learning methodologies have been examined by researchers during the duration of its existence. This has been done with the intention of improving the accuracy of churn prediction models. It is the purpose of this section to present an overview of the most significant advancements that have been made in the field, with a particular emphasis on the use of individual and group classifiers.

A. Single Classifier

Initially, churn prediction research relied mostly on single classifiers due to their simplicity and ease of understanding. Logistic Regression and Decision Trees are widely utilized in research due to their ease of implementation and ability to provide clear explanations. Chang and Hall, 2024 [12] used

Logistic Regression to identify the elements that cause customer turnover in the telecommunications industry, emphasizing the importance of consumer demographics and usage habits. Sebastiaan Hoppner & Eugen Stripling, 2018 [13] employed Decision Trees to predict customer attrition in the retail business, demonstrating the model's ability to deal with categorical data.

Support Vector Machines (SVM) are frequently utilized for predicting churn, particularly in datasets with high dimensionality. Amgad Muneer and Rao Faizan Ali, 2022 [14] showed how well SVMs worked to predict churn in the banking sector, which comprised complex, multifaceted client profiles. Although SVMs performed well, they lacked the interpretability offered by more straightforward models like Decision Trees or Logistic Regression and required intricate hyperparameter tweaking.

For datasets where features are assumed to be independent, Naive Bayes has been a common choice. But it might not be up to the task of dealing with increasingly complicated datasets due to its assumptions. Yulianti et al. 2021 [15] employed Naive Bayes to predict telecom churn and enjoyed its simplicity and speed, although more complex models were more accurate.

The Multi-Layer Perceptron (MLP), a type of neural network, exhibits capability in handling non-linear interactions and large datasets. Abdullah et al. 2018 [16] proven that MLP can get better results than traditional classifiers in the retail industry by discovering previously unseen patterns in customer purchases; however, this requires greater computational power and hyperparameter tuning.

B. Ensemble Classifier

The research community has increasingly focused on ensemble classifiers to address the limitations of single classifiers, as these ensembles integrate multiple models to enhance predictive performance.

For an updated citation on Random Forests in churn prediction, consult the new study by Saha et al, 2023 [17], which provides a comprehensive analysis of the implementation of Random Forests in the retail sector for churn prediction. The authors demonstrate the effectiveness of Random Forests in handling large and complex datasets, emphasizing its resilience to overfitting while providing in-depth analysis of consumer transaction data.

There has been extensive use of Gradient Boosting Machines (GBMs) in the telecom industry, such as XGBoost and LightGBM. Khanna et al. 2020[18] proved that GBMs could reliably manage datasets with imbalances and enhance the accuracy of customer churn prediction. John Ogbonna et al. 2024 [19] further substantiated this by demonstrating GBM's ability to discern intricate, non-linear correlations in customer data, hence improving the overall prediction efficacy in churn situations.

Recently, sophisticated boosting techniques such as AdaBoost and CatBoost have gained prevalence owing to their efficacy in forecasting churn. AdaBoost enhances poor classifiers by increasing the weight of misclassified data points, rendering it particularly effective for imbalanced datasets. Liu et

al. 2024 [20] presented the Ada-XG-CatBoost model, demonstrating its utility in diverse predictive applications, such as customer attrition. CatBoost, engineered to effectively manage categorical features, is especially beneficial in sectors such as banking and retail, where client data frequently comprises these variables.

Stacking is yet another ensemble method that has been investigated due to its capacity to create a single prediction model by combining a number of various kinds of classifiers. Stacking is a technique that was proposed by [21], which means that the outcomes of base classifiers are incorporated into a meta-classifier. Utilizing stacking as a method for predicting customer attrition, the author [22] demonstrated that it was superior to utilizing individual models since it utilized the most effective aspects of each model. Zhang et al, 2021 [23] demonstrated that stacking can perform better than other ensemble methods in the banking business by collecting a greater range of customer interactions and behaviors. This showed that stacking can be more effective than other ensemble approaches.

C. Industry-Specific Applications of Churn Prediction

Because of the high cost of acquiring new customers, the telecom industry has become a prime target for churn prediction. Collaborators on the project [23] demonstrated the efficacy of ensemble approaches in capturing a diverse variety of consumer behaviors by using a hybrid model that included SVM and GBMs to predict loyalty.

Zakariya and Faroug, 2024 [24] highlighted the use of GBMs and Random Forests in retail, a sector characterized by highly variable consumer behavior and transaction data. Ensemble approaches outperform conventional models in capturing the complexities of client purchase behavior, according to their findings. In addition, ensemble classifiers are important in retail since they improve prediction accuracy compared to single classifiers.

Advanced ensemble models provide significant benefits to the banking sector, characterized by its varied product offerings and complex customer relationships. Kimura, et al. 2022 [25] learnt that stacking is a component of hybrid models that decreased the number of false positives and increased the accuracy of churn predictions in retail banking. Zainb and Bestin, 2024 [26] used ensemble methods to enhance their models' performance in predicting customer attrition using neural networks.

D. Ensemble Stacking for Churn Prediction

The analysis of single and ensemble classifiers indicates that ensemble stacking is the most robust and accurate approach for predicting customer churn in the telecommunications, retail, and banking industries. Stacking combines multiple foundational models, such as Logistic Regression, Decision Trees, Naive Bayes, SVM, and MLP, utilizing a meta-learner to produce more accurate and generalizable predictions. This method alleviates the limitations of individual classifiers while leveraging the strengths of each model, as noted by Nureen Afiqah and Mohd Khalid Awang (2023) [22] and Ganaie et al, 2022 [27].

The foundational layer comprises classifiers such as Logistic Regression, Naive Bayes, Decision Trees, SVM, and MLP, with

each one targeting distinct aspects of the data. AdaBoost and CatBoost will be employed to tackle imbalanced data and complex categorical features, thereby enhancing the performance of the base models [28]. A Logistic Regression or Gradient Boosting meta-learner will combine basic model predictions to produce the final result.

Ensemble stacking leverages the strengths of multiple classifiers to yield more accurate predictions compared to individual models or conventional ensemble methods such as Random Forests or Bagging. Singh and Kumari, 2021 [29] demonstrated how the stacking strategy may improve the accuracy of customer turnover forecasts, particularly in industries with complex consumer behavior like retail and telecoms.

Stacking offers considerable versatility across multiple sectors and is not constrained by the complexities of any particular industry. Liu and Yang, 2024 [20] Stacking in the banking sector has been shown to produce higher churn prediction accuracy compared to individual classifiers, due to its ability to capture diverse data characteristics. Additionally, Singh and Kumari, 2021 [1] also found that stacking outperformed other retail ensemble tactics in addressing customer purchase behavior and loyalty patterns.

To sum up, Ensemble stacking is a solid and expandable loss prediction method that helps businesses guess how customers will act in a variety of industries. When you use basic classifiers and meta-learners to combine predictions, you can get better accuracy, generalisation, and stability even when the data is messy and complicated. Its usefulness in banking, shopping, and telecommunications makes it a strong and flexible tool for businesses that want to keep customers and cut down on customer turnover. Ensemble stacking, especially with more advanced methods like AdaBoost and CatBoost, helps businesses come up with strategic ways to keep customers and make them more loyal.

III. METHODOLOGY

A. Datasets

In this study, the researchers evaluated the effectiveness of single and ensemble classifiers in forecasting customer attrition by using three different datasets from the banking industry, the retail industry, and the telecommunications industry simultaneously. Each dataset included a number of aspects that were related to consumer behavior. These aspects included demographic information, account details, transaction histories, and patterns of service use.

With the assistance of the Telco Customer Churn dataset, it is now much simpler to forecast customer churn in the telecommunications industry. This dataset contains 21 factors that shed light on a variety of concerns, including consumer demographics, service subscriptions, and billing patterns, amongst others. Details on the user's demographics (such as gender and senior citizen status), service usage (such as internet service type and streaming options), and account-specific data (such as gender) are essential characteristics. Additionally, information about the account's term, contract type, and monthly charges are also essential features. The dependent variable of interest is the churn status, which is an indicator of whether or

not a customer has discontinued their membership. As demonstrated by this dataset, which illustrates the intricate dynamics that influence churn, some of the factors that have a significant impact on customer retention in the telecommunications business include service quality, the kind of contract, and billing processes [30] [31].

The Online Retail Dataset, used in retail churn prediction, contains transactional data from online retailers, including recency, frequency, and RFM, to identify at-risk customers. Abdullah Rahib et al, 2024 [32] used this dataset to create machine learning models for e-commerce churn prediction using RFM characteristics. Thanh Ho and Nguyen, 2024 [33] used RFM models to improve customer segmentation and retention, confirming the dataset's churn forecast accuracy. Machine learning was used to predict retail turnover using client purchasing behavior [32]. Transaction characteristics such as invoice numbers, customer IDs, and purchase history were important.

The Bank Marketing Dataset, used to forecast banking customer attrition, is available from the UCI Machine Learning Repository. This dataset includes bank clients and direct marketing results, which are essential for analyzing customer churn. Age, occupation, marital status, education, and financial details like account balance and loan status are essential. This dataset relies on interaction data like contact time and kind. The goal variable is the client's term deposit subscription, which often indicates banking sector churn [34] [14].

B. Preprocessing

The train-test split is an important part of machine learning because it checks how well the model works with new data. Researchers can test how well the model works with new data by dividing the information into separate training and test sets [35]. This strategy is crucial for model evaluation, as it alleviates overfitting, which occurs when a model performs well on training data but fails to generalise [35].

Multiple techniques can be employed for data partitioning, with random splitting and stratified sampling as the primary methods. The random split method ensures that both training and test sets accurately represent the entire dataset, enabling an unbiased evaluation [25]. In cases of class imbalance, stratified sampling maintains the proportional representation of each class in both datasets, which is particularly vital in customer churn prediction [35].

This research utilised three different split ratios: 70/30, 80/20, and 60/40. The 70/30 split allocates 70% of the dataset for training and 30% for testing, thereby establishing a balanced approach for model training and evaluation. The 80/20 division allocates 80% of the dataset for training and 20% for testing, which may enhance model performance by facilitating the identification of additional data patterns [36]. The 60/40 split increases the test set size to 40%, potentially providing a more thorough assessment of model performance in the context of a relatively smaller dataset [14].

To enhance the robustness of model evaluation, 5-fold cross-validation was employed, enabling each fold to function as a test set while the model is trained on the other folds [28]. This approach improves the dependability of performance

estimations by diminishing variance in evaluation metrics. Utilising diverse train-test splits and cross-validation guarantees successful model evaluation, resulting in more dependable predictions of customer turnover in the telecommunications, retail, and banking industries.

C. Ensemble Stacking

We employed Decision Trees (DT), Logistic Regression (LR), Support Vector Machines (SVM), and hybrid classifiers such as Random Forest (RF), Gradient Boosting (GBM), and stacking for model selection. Decision Trees, Logistic Regression, and SVM were selected for their simplicity, interpretability, and efficacy in high-dimensional contexts. A hybrid classifier known as Random Forest was selected to enhance accuracy and mitigate overfitting by amalgamating the outputs of Decision Trees. Gradient Boosting was selected due to its ability to incrementally rectify the errors of weak learners, resulting in highly accurate predictions. The third hybrid technique, stacking, integrated the results of Logistic Regression, SVM, and Decision Tree analyses. Logistic Regression was employed as the meta-classifier to forecast the final outcome. Fig. 1 below illustrates the ensemble stacking model.

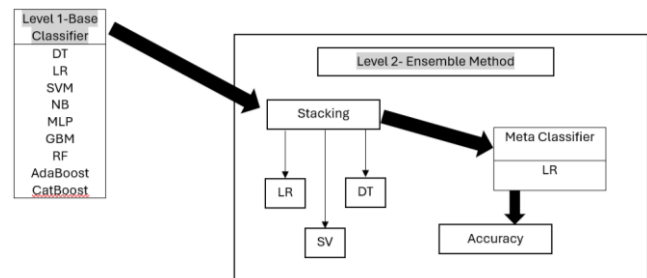


Fig. 1. Model of ensemble stacking.

D. Evaluation Metrics

The models were assessed through various performance metrics, including accuracy, precision, recall, F1 score, and area under the receiver operating characteristic curve (AUC-ROC), to deliver a thorough evaluation of their effectiveness. During the training process, we employed 10-fold cross-validation to assess the robustness and reliability of the models. This process ensured that the models did not overfit to any specific area of the data, thereby preventing overfitting. Several studies, including Bogaert and Delaere's, 2023 [37] have demonstrated that ensemble methods and cross-validation enhance churn prediction models across diverse industries. Xue Ying et al, 2019 [38] highlighted the importance of cross-validation in ensuring the generalizability of machine learning models, particularly when dealing with imbalanced datasets. To assess the generalizability of the trained models and to compare the effectiveness of single classifiers with hybrid classifiers across various industries, testing was conducted on 20% of each dataset that had not been previously disclosed.

IV. RESULT AND DISCUSSION

We divide each dataset in the telecommunications, retail, and banking industries into training (80%) and testing (20%) subsets. To reduce the possibility of our models being excessively dependent on a certain data subset, we employed 10-

fold cross-validation throughout the training phase. This method divides the training data into ten segments, utilizes nine segments for model training, and assesses the model using the remaining segment. This method is performed a total of ten times. The effectiveness of these iterations was averaged to refine the model hyperparameters. The models' generalizability was evaluated on the test set after their optimization.

A. Performance Comparison

1) *The following tables*, provide a concise overview of the performance parameters, including accuracy, precision, recall, F1 score, and AUC-ROC, for each classifier across the three datasets. The results emphasize the disparities in prediction ability between individual classifiers and hybrid classifiers.

B. Discussion of Result

The analysis reveals that hybrid classifiers consistently outperform single classifiers in accuracy and other critical metrics, such as precision, recall, F1 score, and AUC-ROC, across all three datasets: telecom, retail, and banking. The ensuing discussion focusses on the efficacy of several models and their ramifications for forecasting client attrition. In Table I (Telecom Dataset), it is shown that stacking and CatBoost models achieved the highest performance with 86% accuracy and AUC-ROC of 0.91, while traditional models like Decision Trees and Naive Bayes performed poorly with accuracies of 73-75%. In Table II (Retail Dataset), stacking and CatBoost models also outperformed other models with 86% and 85% accuracy, respectively, while Naive Bayes had the lowest accuracy at 71%. Similarly, in Table III (Banking Dataset), stacking and CatBoost models achieved the highest accuracy of 87% and AUC-ROC of 0.91, whereas Decision Trees and Naive Bayes performed the worst with accuracies of 72-74%.

1) *Single classifier*: Decision Trees (DT), although interpretable, exhibited only modest performance, achieving accuracies ranging from 72% to 75% throughout the datasets. This supports findings that decision tree models frequently struggle with complex data patterns, as research highlights their susceptibility to overfitting in high-dimensional datasets, particularly in churn prediction tasks. Despite their simplicity, decision trees remain relevant due to their clarity; yet, they are generally outperformed by more sophisticated models.

TABLE I. PERFORMANCE PARAMETER OF TELECOM DATASET

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
DT	0.75	0.73	0.78	0.75	0.81
LR	0.78	0.76	0.80	0.78	0.84
SVM	0.80	0.78	0.82	0.80	0.85
NB	0.73	0.71	0.77	0.73	0.80
MLP	0.82	0.80	0.83	0.81	0.8
RF	0.83	0.81	0.85	0.83	0.88
GBM	0.85	0.83	0.87	0.85	0.90
AdaBoost	0.84	0.82	0.86	0.84	0.88
CatBoost	0.86	0.84	0.88	0.86	0.90
Stacking	0.86	0.84	0.88	0.86	0.91

TABLE II. PERFORMANCE PARAMETER OF RETAIL DATASET

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
DT	0.72	0.70	0.74	0.72	0.79
LR	0.76	0.74	0.77	0.75	0.82
SVM	0.78	0.76	0.79	0.77	0.83
NB	0.71	0.69	0.73	0.71	0.78
MLP	0.80	0.78	0.81	0.79	0.84
RF	0.82	0.80	0.84	0.82	0.86
GBM	0.84	0.82	0.86	0.84	0.88
AdaBoost	0.83	0.81	0.85	0.83	0.87
CatBoost	0.85	0.83	0.87	0.85	0.89
Stacking	0.86	0.83	0.87	0.85	0.89

TABLE III. PERFORMANCE PARAMETER OF BANKING DATASET

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC
DT	0.74	0.72	0.76	0.74	0.80
LR	0.77	0.75	0.78	0.76	0.83
SVM	0.79	0.77	0.80	0.78	0.84
NB	0.72	0.70	0.75	0.72	0.79
MLP	0.81	0.79	0.82	0.80	0.85
RF	0.84	0.82	0.86	0.84	0.88
GBM	0.86	0.84	0.88	0.86	0.90
AdaBoost	0.85	0.83	0.87	0.85	0.89
CatBoost	0.87	0.85	0.89	0.87	0.91
Stacking	0.87	0.85	0.89	0.87	0.91

Logistic Regression (LR) exhibited improvements over Decision Trees (DT), attaining accuracies between 76% and 78%. Its effectiveness in handling linear decision boundaries makes it highly suitable for binary classification tasks, particularly inside structured datasets like telecommunications and banking. However, the limitation of linear regression is in its inability to describe non-linear relationships, a point emphasized in recent research, which indicates that while linear regression offers interpretability, it often fails to capture more complex data interactions.

Support Vector Machines (SVM) demonstrated improved performance, particularly in the telecom dataset, attaining an accuracy of 80%. The flexibility of SVM to model complex decision boundaries makes it a powerful choice for forecasting customer attrition. However, this results in diminished processing efficiency, particularly with larger datasets, as emphasized in comparative studies on churn prediction models.

Naive Bayes typically demonstrates reduced efficacy in complex datasets like telecommunications, retail, and banking, where the presumption of feature independence rarely holds true. Naive Bayes is anticipated to produce lower accuracy, precision, recall, and F1-scores compared to other models in the tables, including SVM and ensemble methods such as RF and GBM. The simplistic model does not capture the complex patterns necessary for precise churn prediction in these industries.

The Multilayer Perceptron (MLP), a type of neural network, is more proficient at handling nonlinearities in datasets than

linear models like Linear Regression (LR). MLP is anticipated to produce results slightly better than LR and possibly on par with SVM. However, because of its tendency to overfit on small datasets and requiring significant tuning, it may not outperform ensemble methods. In these cases, MLP is expected to exhibit modest performance, with accuracies projected between 0.78 and 0.80.

2) *Ensemble classifier*: Ensemble approaches, including Random Forest (RF) and Gradient Boosting Machines (GBM), demonstrated enhanced performance, particularly in the banking dataset, attaining accuracies of 84% and 86%, respectively. This aligns with research highlighting the effectiveness of ensemble methods in aggregating weak learners to discern complex patterns in churn datasets. Both RF and GBM consistently achieved high AUC-ROC scores, indicating strong discriminatory capacity between churners and non-churners.

AdaBoost, a boosting method, improves accuracy and recall by iteratively combining weak classifiers. The results would fall between Random Forest and Gradient Boosting Machine. Based on the telecom, retail, and banking datasets, AdaBoost is projected to achieve an accuracy similar to that of Random Forest (about 83% - 85%), exhibiting excellent performance, although it does not reach the efficacy of the more advanced boosting method, GBM, or the Stacking model.

CatBoost, a modern boosting algorithm proficient in handling categorical data, would produce outcomes akin to GBM and Stacking. CatBoost is expected to attain accuracies ranging from 0.86 to 0.87, making it equivalent to the Stacking model. The effectiveness of CatBoost, especially in the banking dataset (around 0.87), highlights its capability in handling categorical data efficiently and reducing the need for extensive preprocessing, which is crucial for forecasting customer turnover.

3) *Ensemble stacking*: The stacking ensemble technique demonstrates effectiveness in predicting customer churn in the telecom, retail, and banking sectors, achieving accuracy rates between 0.86 and 0.87 in the examined datasets. Stacking is effective due to its capacity to integrate multiple foundational models, including Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting Machines (GBM), into a meta-learner that exhibits improved generalization across diverse datasets.

In the telecommunications sector, which is marked by considerable data complexity from structured and unstructured sources, stacking improves the management of churn variability more effectively than standalone models. Stacking leverages the strengths of various classifiers, such as the decision boundary optimization of Support Vector Machines (SVM) and the pattern recognition capabilities of Random Forest (RF) and Gradient Boosting Machines (GBM), to improve generalization and address the complexities inherent in telecom data. This adaptability aligns with prior research that emphasizes the effectiveness of stacking in improving prediction robustness through model diversity.

In the retail sector, consumer behavior data may fluctuate due to seasonal and demand variations. Stacking enhances forecast accuracy by capturing diverse customer behavior patterns through multiple algorithms. This helps to overcome the limitations found in individual models like Decision Trees or Logistic Regression, which can either underfit or overfit the data. Research in this area demonstrates that stacking enhances generalization, particularly in dynamic environments like retail.

In the banking sector, predicting customer turnover is crucial due to competitive dynamics and substantial client lifetime value. Stacking offers an advantage by integrating robust classifiers. Models like GBM effectively handle skewed churn data, while SVM improves precision in decision boundaries. Stacking improves prediction by utilizing these characteristics, resulting in enhanced accuracy and generalization, as demonstrated in various studies that emphasize stacking's role in advancing model performance in churn prediction.

C. Implications for Customer Prediction

Hybrid models, especially ensemble approaches like Stacking, Gradient Boosting Machines (GBM), and CatBoost, outperform Decision Trees (DT) and Logistic Regression in telecom, retail, and banking datasets. These models perform better and more accurately across industries. Stacking had the highest accuracy of 0.87 in banking, while CatBoost excelled in categorical data with 0.86 in telecom and banking [37][39].

For better churn prediction, our findings suggest hybrid models. Multi-base models increase generalization and forecast accuracy for complex, industry-specific data. Telecom companies with massive customer data sets can stack to identify at-risk customers and optimize retention. CatBoost models enable quick categorical data handling and personalized engagement in retail, including seasonal client behavior [39].

Hybrid models increase high-value client churn detection in banking, because revenue directly affects retention. GBM and Stacking, with strong AUC-ROC ratings (up to 0.91), enable organizations adjust retention efforts. The accuracy of these models helps organizations allocate resources, create focused solutions, and reduce customer churn while adapting to industry trends.

Recent research show hybrid models are used more in real-world churn prediction. Sector-wide churn prediction is more accurate, scalable, and flexible when many algorithms are used. Hybrid models can handle many datasets and complicated consumer behavior patterns, making them excellent for customer relationship management and churn reduction across industries.

V. CONCLUSION

This research examined both single and hybrid classifiers for predicting customer attrition in the telecommunications, retail, and banking sectors. Hybrid models, particularly ensemble methods such as Stacking, Gradient Boosting Machines (GBM), and CatBoost, surpassed individual classifiers including Decision Trees (DT), Logistic Regression (LR), and Naive Bayes (NB) in terms of accuracy, precision, recall, F1-score, and AUC-ROC. The hybrid models Stacking and CatBoost exhibited superior performance across all datasets, with an

accuracy of 86%–87%. Despite being more straightforward and accessible, individual classifiers failed to achieve the performance of ensembles, with Decision Trees yielding subpar accuracy in high-dimensional and complex datasets.

Hybrid models proficiently manage structured and unstructured data by integrating the strengths of classifiers, as demonstrated by the telecom dataset. Stacking demonstrated enhanced generalization in retail, when consumer behavior fluctuated with seasonal demands. Hybrid models were particularly effective in forecasting high-value client attrition in banking, where misclassification could result in significant financial losses.

VI. RECOMMENDATION AND FUTURE WORKS

The study is flawed. First, hybrid models outperformed single classifiers but required more computer resources and tuning. This may limit time-sensitive real-time churn prediction systems. Second, this study's industry-representative datasets may not cover all industrial circumstances. Generalizability may be limited by ignoring client demographics, product diversity, and regional differences. Most of this research used structured datasets with few category characteristics, which helped CatBoost. Customer turnover prediction increasingly relies on unstructured data like text and social media interactions, hence these hybrid models should be examined. Models were tested using 10-fold cross-validation. Retailers could study time-series cross-validation as customer behaviour changes. LSTM or Transformer-based deep learning models can reveal long-term customer behaviour dependencies. These tactics may help telecoms and finance organisations retain customers.

Hybrid models significantly enhance consumer churn prediction across various industries. To develop more practical and comprehensive churn prediction systems, it is crucial to focus on reducing computational complexity, improving real-time processing, and exploring new data sources and validation methods. However, one challenge with the proposed stacking model is the increased computational complexity due to training multiple base models and a meta-learner. Furthermore, the model's performance can vary across different datasets, as certain algorithms are more sensitive to data quality and preprocessing steps.

ACKNOWLEDGMENT

This work is supported by Fundamental Research Grant Scheme (FRGS/1/2023/ICT02/UNISZA/02/1) under the Ministry of Higher Education (MOHE) and University Sultan Zainal Abidin (UniSZA), Malaysia.

REFERENCES

- [1] N. Mustafa and L. S. Ling, "Customer churn prediction for telecommunication industry : A Malaysian Case Study [version 1 ; peer review : awaiting peer review]," 2021.
- [2] "İzmir İktisat Dergisi Churn Customer Management in Retail Industry : A Case Study," vol. 37, pp. 0–3, 2022, doi: 10.24988/ije.
- [3] R. Of and H. I. N. The, "P Rediction of C Ustomer C Hurn in the B Anking S Ector :"
- [4] B. R. Agasti and S. Satpathy, "Predicting customer churn in telecommunication sector using Naïve Bayes algorithm," vol. 35, no. 3, pp. 1610–1617, 2024, doi: 10.11591/ijeecs.v35.i3.pp1610-1617.
- [5] O. Adwan, H. Faris, O. Harfoushi, and N. Ghatasheh, "Predicting Customer Churn in Telecom Industry using Multilayer Preceptron Neural Networks : Modeling and Analysis," no. March, 2014.
- [6] P. Doctor and A. Sciences, "Machines against malaria : Artificial intelligence classification models to advance antimalarial drug discovery Ashleigh van Heerden," no. December, 2023.
- [7] I. Alshourbaji, N. Helian, Y. Sun, and A. G. Hussien, "An efficient churn prediction model using gradient boosting machine and metaheuristic optimization," *Sci. Rep.*, pp. 1–19, 2023, doi: 10.1038/s41598-023-41093-6.
- [8] A. V. Dorogush, V. Ershov, and A. Gulin, "CatBoost : gradient boosting with categorical features support," pp. 1–7.
- [9] T. Y. Lin et al., "Journal of Engineering Technology and Applied Physics Stacking Ensemble Approach for Churn Prediction : Integrating CNN and Machine Learning Models with CatBoost Meta-Learner," vol. 5, no. 2, pp. 99–107, 2023.
- [10] M. Z. Alotaibi, "Customer Churn Prediction for Telecommunication Companies using Machine Learning and Ensemble Methods," no. June, pp. 1–8, 2024, doi: 10.48084/etasr.7480.
- [11] S. F. Sabbeh, "Machine-Learning Techniques for Customer Retention : A Comparative Study," vol. 9, no. 2, pp. 273–281, 2018.
- [12] V. Chang, K. Hall, Q. A. Xu, F. O. Amao, M. A. Ganatra, and V. Benson, "Prediction of Customer Churn Behavior in the Telecommunication Industry Using Machine Learning Models," 2024.
- [13] E. Stripling and B. Baesens, "Profit Driven Decision Trees for Churn Prediction," no. December 2017, 2018, doi: 10.1016/j.ejor.2018.11.072.
- [14] A. Muneer, R. F. Ali, A. Alghamdi, S. M. Taib, and A. Almaghthawi, "Predicting customers churning in banking industry : A machine learning approach," no. March, pp. 539–549, 2022, doi: 10.11591/ijeecs.v26.i1.pp539-549.
- [15] I. O. P. C. Series and M. Science, "Sequential Feature Selection in Customer Churn Prediction Based on Naive Bayes Sequential Feature Selection in Customer Churn Prediction Based on Naive Bayes," 2020, doi: 10.1088/1757-899X/879/1/012090.
- [16] H. Faris, "A Hybrid Swarm Intelligent Neural Network Model for Customer Churn Prediction and Identifying the Influencing Factors," pp. 1–18, 2018, doi: 10.3390/info9110288.
- [17] V. Vu, "Predict customer churn using combination deep learning networks model," *Neural Comput. Appl.*, vol. 36, no. 9, pp. 4867–4883, 2024, doi: 10.1007/s00521-023-09327-w.
- [18] G. Fatima, S. Khan, F. Aadil, and D. H. Kim, "An autonomous mixed data oversampling method for AIOT-based churn recognition and personalized recommendations using behavioral segmentation," pp. 1–32, 2024, doi: 10.7717/peerj-cs.1756.
- [19] O. J. Ogbonna and G. I. O. Aimufua, "Churn Prediction in Telecommunication Industry : A Comparative Analysis of Boosting Algorithms," vol. 10, no. 1, pp. 331–349, 2024.
- [20] Y. Liu, T. Yang, L. Tian, B. Huang, J. Yang, and Z. Zeng, "Ada-XG-CatBoost : A Combined Forecasting Model for Gross Ecosystem Product (GEP) Prediction," 2024.
- [21] N. N. Adnan and M. K. Awang, "A Review on Classification Algorithm for Customer Churn Classification," vol. 3878, no. X, pp. 1–15, 2023, doi: 10.35940/ijrte.
- [22] N. Afiqah, M. Zaini, and M. K. Awang, "Hybrid Feature Selection Algorithm and Ensemble Stacking for Heart Disease Prediction," vol. 14, no. 2, pp. 158–165, 2023.
- [23] L. C. C. By et al., "Customer Churn Prediction on E-Commerce Data using Stacking Classifier Customer Churn Prediction on E-Commerce Data using Stacking Classifier," pp. 0–10, 2022, doi: 10.36227/techrxiv.20291694.v1.
- [24] Z. M. S. Mohammed, F. A. Abdalla, M. Salih, A. F. A. Mahmoud, and A. Satty, "A Comprehensive Bibliometric Analysis of Churn Prediction Research : An Essay on Trends , Key Contributors and Global Participation in the Field .," vol. 36, no. 4, pp. 219–228, 2024.

- [25] T. Kimura, "Customer Churn Prediction With Hybrid," no. January, 2022.
- [26] B. Baby, Z. Dawod, W. Elmedany, and M. S. Sharif, "Customer Churn Prediction Model Using Artificial Neural Networks (ANN): A Case Study in Banking".
- [27] M. Ganaie, M. Hu, A. K. Malik, and M. Tanveer, "Ensemble deep learning: A review," no. October, 2022, doi: 10.1016/j.engappai.2022.105151.
- [28] H. Tran, N. Le, and V. Nguyen, "Customer Churn Prediction in the Banking Sector Using Machine Learning-Based," no. February, 2023, doi: 10.28945/5086.
- [29] M. S. Devi, J. Saharia, S. Kumar, A. Chansoriya, and P. Yadav, "Machine Learning Based Suspicion of Customer Detention in Banking with Diverse Solver Neighbors and Kernels," vol. 3878, no. 4, pp. 3244–3249, 2019, doi: 10.35940/ijrte.D8043.118419.
- [30] V. Hariharan, I. Khan, P. Katkade, and P. S. D., "Customer Churn analysis in Telecom Industry," pp. 932–936, 2020.
- [31] M. Odusami, O. Abayomi-alli, S. Misra, A. Abayomi-alli, and M. M. Sharma, "A Hybrid Machine Learning Model for Predicting Customer Churn in the Telecommunication Industry for Predicting Customer Churn," no. November 2023. Springer International Publishing, 2021. doi: 10.1007/978-3-030-73603-3.
- [32] A. Al Rahib, N. Saha, R. Mia, and A. Sattar, "Customer data prediction and analysis in e-commerce using machine learning Customer data prediction and analysis in e-commerce using machine learning," no. August, 2024, doi: 10.11591/eei.v13i4.6420.
- [33] T. Ho, S. Nguyen, H. Nguyen, N. Nguyen, and D. Man, "An Extended RFM Model for Customer Behaviour and Demographic Analysis in Retail Industry," vol. 14, no. 1, pp. 26–53, 2023.
- [34] E. Dindigul- and E. Dindigul-, "Bank Customer Retention Prediction and Customer," vol. 5, no. 9, pp. 444–449, 2020.
- [35] M. Rai, B. Ghimire, N. Uprety, and R. Shrestha, "A Comparative Analysis of Data Balancing Techniques: SMOTE and ADASYN in Machine Learning for Enhancing Bank Loan Default Risk Predictions," vol. 155, no. 1, pp. 159–168, 2024.
- [36] A. Manzoor, M. A. Qureshi, E. Kidney, and L. Longo, "A Review on Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners," IEEE Access, vol. PP, p. 1, 2024, doi: 10.1109/ACCESS.2024.3402092.
- [37] M. Bogaert, "Ensemble Methods in Customer Churn Prediction: A Comparative Analysis of the State-of-the-Art," 2023.
- [38] C. Series, "An Overview of Overfitting and its Solutions An Overview of Overfitting and its Solutions," 2019, doi: 10.1088/1742-6596/1168/2/022022.
- [39] M. Imani, "Hyperparameter Optimization and Combined Data Sampling Techniques in Machine Learning for Customer Churn Prediction: A Comparative Analysis," 2023.

Hotspots and Insights on Quality Evaluation of Study Tours: Visual Analysis Based on Bibliometric Methodology

Meihua Deng 

School of International Communication, Hunan Mass Media Vocational and Technical College, Changsha, 410100, Hunan, China

Abstract—In this paper, taking 474 articles about quality evaluation of study tours in Web of Science (WOS) database as the research object, quantitatively analyze them with the help of CiteSpace 6.3.R1 software and excel data statistics, and analyze the impact of the literature data, authors' cooperation network, issuing institutions, journal distribution, and keywords' co-occurrence, clustering, and emergence factors, combined with time interval in-depth analysis and prediction, so as to present the research results in the form of visualized knowledge map. The results of the study show that the field of quality evaluation of research and study tourism an interdisciplinary field involving innovative research with multidisciplinary integration. During the decade of 2015-2024, it has experienced three stages of starting and exploration (2015-2018), rapid growth and diversification (2019-2021), and adjustment and maturity (2022-2024). From the viewpoint of authors and issuing organizations, authors are mostly independent research and have not yet formed a clustering research network. Research hotspots from the theoretical system construction and model development, empirical analysis, gradually shifted to user behavior analysis and recommendation system research. The future tends to research on research and learning integration intelligent decision-making, research and learning industry economy, environmental tourism practice and risk management.

Keywords—Research tourism; tourism quality evaluation; visualization analysis

I. INTRODUCTION

With the deep integration of the education sector and the tourism industry, study tours, as an innovative form of experiential activity, have begun to receive widespread attention worldwide [1]. This combination of educational purposes and tourism experiences not only provides participants with an opportunity to learn and explore in a real-world environment, enabling them to acquire knowledge and skills while traveling, but for tourist destinations, study tourism has also become an effective economic development tool [2-3]. By attracting learners of different ages and backgrounds, it increases the attractiveness of the destination, extends the length of stay of tourists, and drives the development of local catering, accommodation, transportation and other related industries, thus bringing significant economic benefits to the destination [4]. Therefore, study tourism not only enriches the connotation of education, but also injects new vitality into the tourism industry, and this win-win characteristic makes study tourism a hot area of common concern for both education and tourism [5].

However, due to the lack of uniform evaluation standards, quality evaluation of study tours has become a challenge that requires urgent attention from both academics and practitioners. The quality evaluation of study tours is crucial for enhancing the tourism experience and educational effectiveness. The development of this form of tourism not only enriches educational resources, improves the satisfaction of learners participating in study tours, and promotes the enhancement of knowledge and skills. At the same time, it injects new vitality into the tourism industry, helps the local tourism industry to build a richer and multi-level tourism brand image, and attracts more tourists. Therefore, an in-depth discussion of the quality evaluation of study tours is of great theoretical and practical significance for promoting the development of the educational tourism industry.

The framework of this paper is described as follows: Section II, "Research methodology" elaborates on the specific methods used for bibliometric analysis with CiteSpace software and Excel, including data sources and the data cleaning process. Section III, "Analysis process and findings," deeply analyzes the trends in literature publication, including the annual number of publications and publication curve trends in "III (A) Analysis of literature releases" and discusses three developmental stages in two subsections: "III (A) (1) Annual number of communications" and "III (A) (2) Trends in the issuance curve", the initial exploration phase, the rapid growth and diversification phase, and the adjustment and maturation phase. Subsequently, the "III (B) Analysis of literature authors" section explores the authors of the literature and the institutions publishing them, including "III (B) (1) Authors and issuing organizations" and "3.2.2 Author collaboration network" in two subsections. The "III (C) Distribution analysis of journals" section analyzes the distribution of journals, including "III (C) (1) Analysis of core journals" and "III (C) (2) Analysis of cited journals" in two subsections. Section IV, "Relevant analysis based on the field" reveals the research hotspots and the dynamic evolution trends of hot fields through keyword co-occurrence, clustering, and emergence analysis, including three subsections: "IV (A) Hot topic analysis," "IV (B) Analysis of hot areas" and "IV (C) Trend analysis of dynamic evolution" Finally, Section V, "Conclusion," summarizes the main findings of the research, discusses the limitations of the study, and provides an outlook on future research directions.

II. RESEARCH METHODOLOGY

A. Research Tools

CiteSpace is a multivariate dynamic visualization and analysis software developed based on Java language, which is capable of handling a large amount of transcription information and performing various analyses such as collaborative network, co-citation, keyword co-occurrence, and keyword clustering [6]. It also provides three advanced clustering analysis methods of Latent Semantic Analysis (LSI), Log Likelihood Ratio (LLR), and Mutual Information Algorithm (MI), which help users to identify potential themes and trends in research [7]. In this study, with the help of CiteSpace 6.3.R1 software application, this paper transforms the literature data in this field during the period of 2015-2024 into a knowledge graph so as to visualize the current status of research, research hotspots, and future trends in this research field.

In order to increase the accuracy of the study, this study quantifies the research scholars' research result publishing behavior, the interaction behavior between research scholars, and between research scholars and institutions in the research field, and uses standardized EXCEL data forms to summarize the statistics, and then uses the empirical data to analyze the state of research in the field.

B. Data Sources

Literature from the Web of Science (WOS) database, which is recognized by authors worldwide for its authoritative academic citation index, was selected as the primary data source for this study. A set of keywords was carefully designed to ensure a comprehensive coverage of the relevant studies in the literature search. The main keywords include "educational tour", "tourism quality evaluation" and "study tour", which cover this study. ", covering the concept of this research field, quality evaluation dimensions, evaluation index system, educational effect and other aspects of the research [8]. Boolean logic operators such as "AND" and "OR" are also used to optimize the search strategy and improve the relevance and accuracy of the search results [9]. Considering that in-depth research needs to be supported by a sufficient amount of literature data and closely related to the development of research and study tourism, the search time range was set as 2015-2024, and 2,221 pieces of related literature were initially obtained by combining the search strategies of databases, types of literature, and language ranges.

C. Data Cleansing

In order to ensure the objectivity and authenticity of the results of this study, a rigorous data cleaning process was carried out in this study before using the data. Literature with no direct relevance, low relevance, lack of keywords, authors, and other lack of key elements, as well as non-research literature such as duplicated, unreviewed, and scrapped manuscripts, news commentaries, interview reports, and so on, were eliminated [10]. After several rounds of data cleaning, 474 valid literatures with high relevance were finally obtained. For the screened valid literature, detailed data records were made in this paper, including information such as title, author, publication year, journal name, keywords and so on. This information will provide basic data support for the subsequent bibliometric analysis and help us to

show the research hotspots and development trends in this research field.

III. ANALYSIS PROCESS AND FINDINGS

A. Analysis of Literature Releases

1) *Annual number of communications:* In this study, a bar chart of literature publication was produced based on the year of publication and the annual publication quantity of 474 valid literatures. In terms of the number of annual publications, the research in this field shows a relatively obvious growth trend. In 2015, the number of relevant literature published was only 18, showing that the research in this field is still in its infancy. In the following years, the number of publications increased steadily, with 24 and 28 publications in 2016 and 2017, respectively, indicating that authors began to gradually focus on this field. Entering 2018, the number of publications decreased slightly to 24, with authors exploring new research methods or waiting for more empirical data. Starting from 2019, the number of publications increased significantly to 41, and this growth trend peaked in 2020 and 2021, with 71 and 80 publications, respectively, and the surge in the number of publications was related to the attention, exploration, and practice of research in this field in the field of global education. From 2022 onwards, the annual number of publications, although decreasing slightly from 76 to 46, remained overall at a high level [11]. This indicates that the heat of research in this field is not decreasing, and academic research authors are shifting to more in-depth empirical research on the existing research results, waiting for more practice cases to accumulate before summarizing and analyzing them. As shown in Fig. 1.

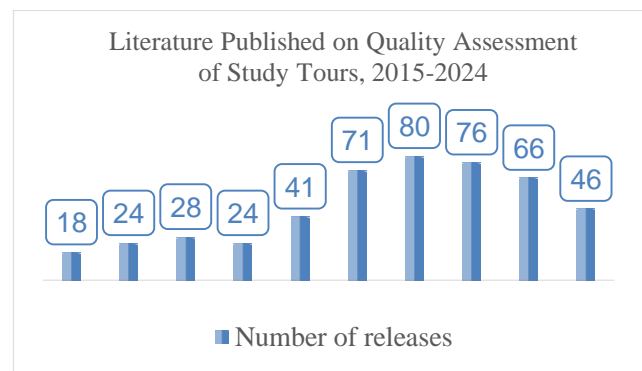


Fig. 1. Literature published, 2015-2024.

2) *Trends in the issuance curve:* The annual change in the number of literature publications shows the trend of research dynamics in this research area, which has shown a curvilinear growth over the past decade. In this paper, this period is broadly categorized into three phases: start and exploration, rapid growth and diversification, and adjustment and maturity.

Start-up and exploration phase: 2015-2018

In the early period of research in this field, i.e., 2015-2018, the annual number of literature publications increased from 18 to 24, showing a slow but steady growth trend, marking the

beginning and exploratory period of this field of research [12]. Research authors conducted preliminary discussions on the basic concepts, theoretical frameworks, and potential value of research tourism in the field of education during this period, with much of the research focusing on defining the connotations of research tourism, evaluating its educational efficacy, and exploring its comprehensive impact on learners [13]. These preliminary studies have laid a solid foundation for subsequent in-depth exploration, and despite the small number of publications, each of them is of great significance to the construction of a body of knowledge in this field.

Rapid Growth and Diversification Phase: 2019-2021

Entering 2019, research in this field began to receive wider attention, with a significant increase in the number of literature publications, reaching 41 in 2019 and climbing to a peak of 71 and 80 in 2020 and 2021, respectively. The rapid growth in this phase reflects the fact that the issue of quality evaluation of research tourism, as a new form of combining education and tourism, has begun to become a focus of attention in both academia and practice. Research scholars researchers began to explore diversified evaluation models and index systems, trying to assess the quality of study tours from different angles and levels [14]. The studies at this stage not only increase in number, but also present diversified characteristics in research methodology and theoretical depth, providing rich perspectives and profound insights for the research in this field.

Phase III: Adjustment and maturity phase (2022-2024)

From 2022 onwards, the number of publications within this research area decreased, with 76, 66 and 46 publications, respectively. This trend suggests that research scholars are beginning to consolidate and reflect on existing research findings after experiencing rapid growth in the previous period [15]. During this period, research may focus more on quality than quantity, and academic scholars may be seeking more effective research methods or waiting for more empirical data to support their research hypotheses. The fallback in this stage represents a sign that research in this field has entered a mature period, and the focus of research may shift from broad exploration to specific problem solving and application practice, reflecting the researchers' concern and thinking about the deeper issues of quality evaluation of research and study tourism.

Overall, the change in the amount of literature published in this research field demonstrates the dynamic development process of research in the field, from the start and exploration to rapid growth and diversification, and then to adjustment and maturity, with each stage contributing to the accumulation of knowledge and academic deepening in the field [16]. As the research in the field continues to deepen, more high-quality research results will emerge, providing more scientific and systematic theoretical support and practical guidance for the practice of study tours.

B. Analysis of Literature Authors

1) *Authors and issuing organizations:* The depth of cooperation among the authors of the literature and the academic influence of the core authors are key indicators for

assessing the maturity of research in this field. By extracting the first author information of 671 documents, a total of 619 authors were identified, of which 52 had more than 2 publications, accounting for 8.4% of the total number of authors. This data indicates that the number of scholars who have been deeply engaged in researching this field for a long period of time is relatively limited, but they have a high influence and academic contribution to the research of this field, which not only enriches the academic discussion on the evaluation of the quality of research and study tours but also provides important studies and inspirations for the subsequent research [17]. They not only enrich the academic discussion on quality evaluation of study tours, but also provide important references and inspirations for subsequent studies. According to Price's law, the number of publications by the core authors calculated in this study is only 1.03, which is inconsistent with the expectation of Price's law, proving that the research in this field is not yet mature, and the leadership role of the high-producing authors has not yet been emphasized [18]. In view of this, this paper defines scholars with two or more publications as high-producing authors in order to more accurately identify researchers who have made significant contributions to the field of this research. Statistically, there are 52 high-producing authors with a total of 131 publications, accounting for 15.5% of the total number of publications. This percentage indicates that despite the relatively small number of high-producing authors, they have played an important role in advancing research in the field as shown in Table I.

In addition to individual research scholars, the contribution of research institutions in this research field cannot be ignored. These institutions have provided substantial research funding, resource support and academic environment support for research in this field, thus facilitating collaboration and knowledge sharing among scholars in this research field [19]. After counting the first author's institution, it was found that university institutions published 325 documents and research institutes published 163 documents. The percentage of publications from university institutions is about 48.40% and the percentage of publications from research institutes is about 24.30%. This data shows that university institutions play a leading role in the research in this field, and research institutes have also made significant contributions [20]. In addition, of the 46 universities with more than two publications, 67% are in the computer technology application category and about 22% are in the teacher training category. This finding indicates that computer-based universities are more prominent in terms of attention and research results in this field of study, and they are more active and rich in this field of study compared to other types of institutions.

In addition, this paper also statistically analyzes the number of publications and citations of the authors to assess their research contributions and influence. Highly prolific authors usually have a high number of citations, indicating that their research results have been widely noticed and recognized by all circles and disciplines, and the research results of these core authors have not only promoted the theoretical development of the field, but also provided theoretical guidance for practice.

TABLE I. LIST OF STATISTICS ON THE NUMBER OF PUBLICATIONS BY CORE AUTHORS

Author	Number of communications	Author	Number of communications
Alkhamees, Nora	2	Li, Qing	2
Aloud, Monira Essa	2	Li, Shaoshuai	2
Ammirato, Salvatore	2	Liu, Chichang	2
Balland, Pierre-Alexandre	2	Liu, Hao	2
Bhattacharya, Pronaya	2	Liu, Weihua	2
Bodendorf, Frank	2	Liu, Xiaolei	2
Broekel, Tom	2	Liu, Zonghua	2
Cao, Jie	2	Long, Shangsong	2
Cerna, Fernando V	2	Lu, S-Y	2
Chen, Chien-Ming	2	Ma, Qiongxiu	2
Chen, Ruey-Shun	2	O'clery, Neave	2
Chen, Yeh-Cheng	2	Rabelo, Ricardo A L	2
Contreras, Javier	2	Raso, Cinzia	2
Deng, Shangkun	2	Rigby, David	2
Dincer, Hasan	2	Rodrigues, Joel J P C	2
Diodato, Dario	2	Sofo, Francesco	2
Felicetti, Alberto Michele	2	Tanwar, Sudeep	2
Franke, Joerg	2	Tian, Guixian	2
Giuliani, Elisa	2	Wang, Fei-Yue	2
Guo, Naicheng	2	Wang, Shuai	2
Guo, Xiaobo	2	Xiao, Yingyuan	2
Hausmann, Ricardo	2	Xiong, Naixue	2
Hsu, Ching-Hsien	2	Yuksel, Serhat	2
Huang, Szu-Hao	2	Zhang, Wenyuan	2
Li, Jing	2	Zheng, Wenguang	2
Zhu, Yingke	2	Zhou, MengChu	2

Based on the results of the authors' analysis of the literature, it is suggested that future research should focus more on interdisciplinary and international cooperation. Knowledge exchange and innovation can be promoted by strengthening cooperation between researchers from different fields and different regions [21]. At the same time, emerging research institutions and young scholars are encouraged to participate in research on quality assessment of study tours in order to increase the diversity and vitality of research.

2) *Author collaboration networks*: In this study, the collaborative network of scholars in this research area was carefully analyzed through CiteSpace 6.3.R1 software. The co-occurrence threshold was set to 2, so that the connectivity between two scholars would only be visible in the network if they had co-authored at least 2 papers. This setting helps to capture the main collaboration patterns of authors in the field, while filtering out episodic collaborations and ensuring that the network mapping is somewhat stable and substantial [22]. The author collaboration network mapping consists of 622 nodes and 937 connectors, with nodes representing independent research authors and connectors indicating collaborative

relationships between them. Although the network density is only 0.0049, showing that the overall structure is relatively loose and the author collaboration network has not yet formed a highly dense cluster. However, the gradual increase in the frequency of collaboration among research scholars in the field compared to historical data signals the potential for future development of research collaboration in the field. As shown in Fig. 2.

The mapping of author collaboration networks, as shown in Fig. 2, not only demonstrates the current research collaboration dynamics, but also foretells the possible development direction of future collaboration networks. It is worth noting that the distribution of high-producing authors is relatively concentrated, and the research results are bit prominent in specific years, which may be related to the research hotspots, financial support, or specific research projects at that time [23]. For example, author Ammirato-Salvatore's high output in 2019 reflects a concentrated burst of demand in that research area in that year. Meanwhile, Balland-Pierre-Alexandre's multiple research outputs published in 2022 may be related to the financial support and policy impetus of related research projects in that year.

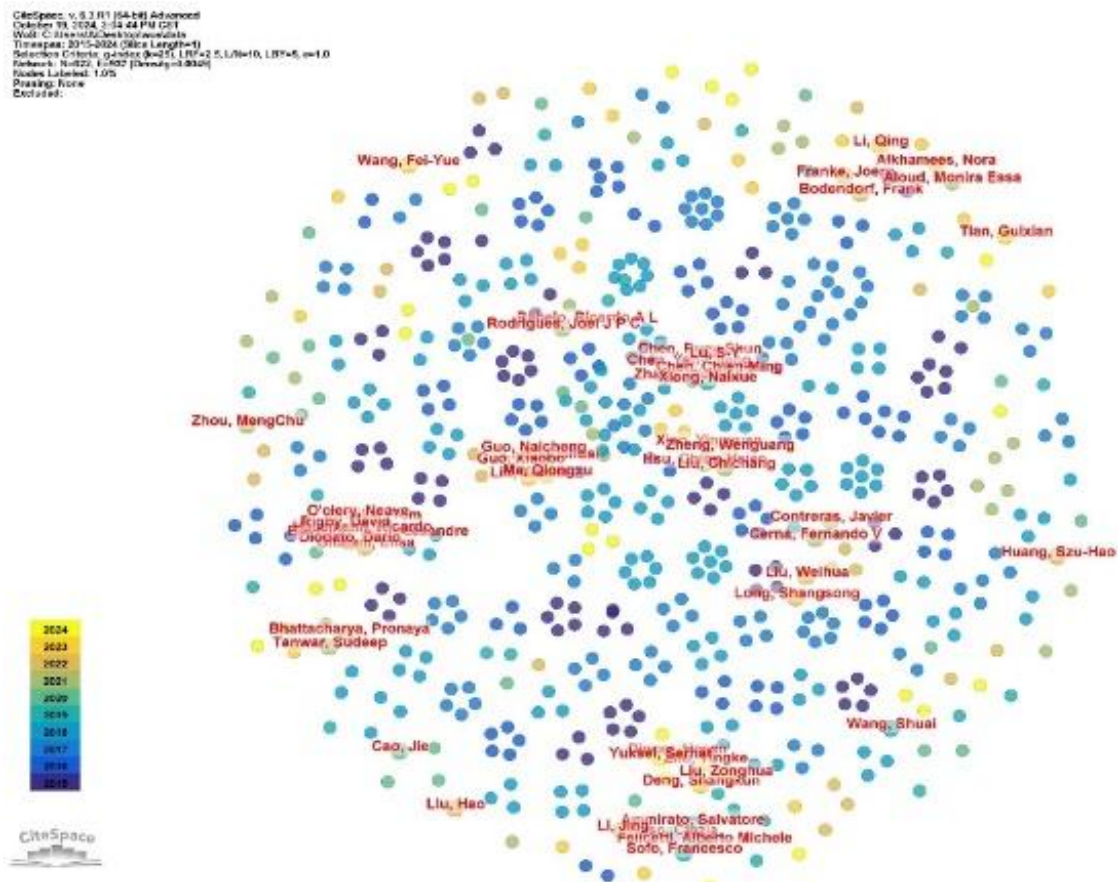


Fig. 2. Mapping of author collaboration networks.

The evolutionary trend of the authors' collaborative network provides a visual observation of the researchers' activity and participation patterns in the field. Some authors consistently publish research successes under the same topic in the same field, showing their long-term research and in-depth exploration of the research area. While other authors publish only 1 research result, their participation increases the diversity of the author collaboration network. This diversity in the author collaboration network reflects the broad appeal and flexibility of the research field, providing a rich variety of perspectives and methodologies for research in this area. Over time, it is expected that more new researchers will join this field of research, further enriching the structure of the author collaboration network [24]. As the collaboration deepens and expands, it is expected that research collaboration in this field will become more intense and systematic, which will not only promote the accumulation of knowledge and innovation, but also the development of interdisciplinary research. The development of such collaborative networks heralds a more active and diverse future for the research field of research and study tourism quality evaluation.

C. Distribution Analysis of Journals

1) *Analysis of core journals:* After counting the number of journal articles published in this research field, a number of journals with high influence in this field were found. These journals not only provide a platform for the publication of

research results in this field, but also reflect the popularity and academic attention of different research directions [25]. The top three journals are Expert Syst Appl, IEEE Access and Eur J Oper Res, with 227, 178 and 149 articles respectively. Among them, Expert Syst Appl focuses on original papers in the application field, with 96.25% of research articles, which has a significant academic influence in this research field, IEEE Access focuses on interdisciplinary research, and Eur J Oper Res prefers operations research methods and decision-making practices. Inform Sciences and Sustainability-Base, which follow closely in the ranking, are biased with information science and sustainability research [26]. The specialized nature of these journals is enough to show that this field of study is an interdisciplinary field that involves multidisciplinary integration of innovative research. As shown in Table II.

2) *Analysis of cited journals:* To further understand the interactions between core academic journals in this research area and the potential correlations between research topics in this area. In this paper, a network graph of journal co-citation relationships was reconstructed for the journals described above using CiteSpace 6.3.R1 software. The spectrogram consists of 473 nodes and 2613 connecting lines, with a network density of 0.0049 and a relatively loose overall network structure. As shown in Fig. 3.

TABLE II. LIST OF STATISTICS ON THE NUMBER OF ARTICLES IN CORE JOURNALS

PERIODICALS	VOLUME OF LITERATURE	PERIODICALS	VOLUME OF LITERATURE
Express Syst Appl	227	Appl Energ	63
IEEE Access	178	Eng Appl Artif Intel	61
Eur J Oper Res	149	Energies	60
Lect Notes Comput Sc	118	Technol Forecast Soc	59
J Clean Prod	116	J Bus Res	57
Inform Sciences	109	Energy	52
Sustainability-Basel	109	Arxiv	52
Decis Support Syst	108	Int J Inform Manage	49
Appl Soft Comput	106	Appl Sci-Basel	49
Manage Sci	93	Commun Acm	48
Int J Prod Econ	91	Ieee T Intell Transp	48
Knowl-Based System	91	Soft Comput	48
Int J Prod Res	89	Omega-Int J Manage S	47
Future Gener Comp Sy	83	Comput Oper Res	46
Ieee T Ind Inform	80	Ann Oper Res	45
Neurocomputing	77	Int J Elec Power	44
Comput Ind Eng	74	Energy Policy	44
Procedia Comput Sci	72	J Bank Financ	42
Ieee Internet Things	72	Plos One	42
Renew Sust Energy Rev	70	J Intell Fuzzy Syst	42
Sensors-Basel	69	Adv Neur In	41
Ieee T Knowl Data En	68	Lect Notes Artif Int	41
J Financ	67	Ieee Commun Surv Tut	40
Neural Comput Appl	65		

CiteSpace v. 5.3.R1 [64-bit Advanced]
October 30, 2024, 10:50:59 AM CST
Work: C:\Users\user\Documents\IJACSA
Preset: g=0.1, z=0.01, w=0.1, q=0.1, r=1, p=0.01, m=10, n=10, l=0
Maximum Q: 0.95, Maximum Z: 0.95, Modularity Q: 0.95, Mean Silhouette S: 0.95
Weighted Mean Silhouette: 0.95, Weighted Mean Modularity: 0.95
Number of Nodes: 100, Number of Edges: 100
Pruning: None
Execution:

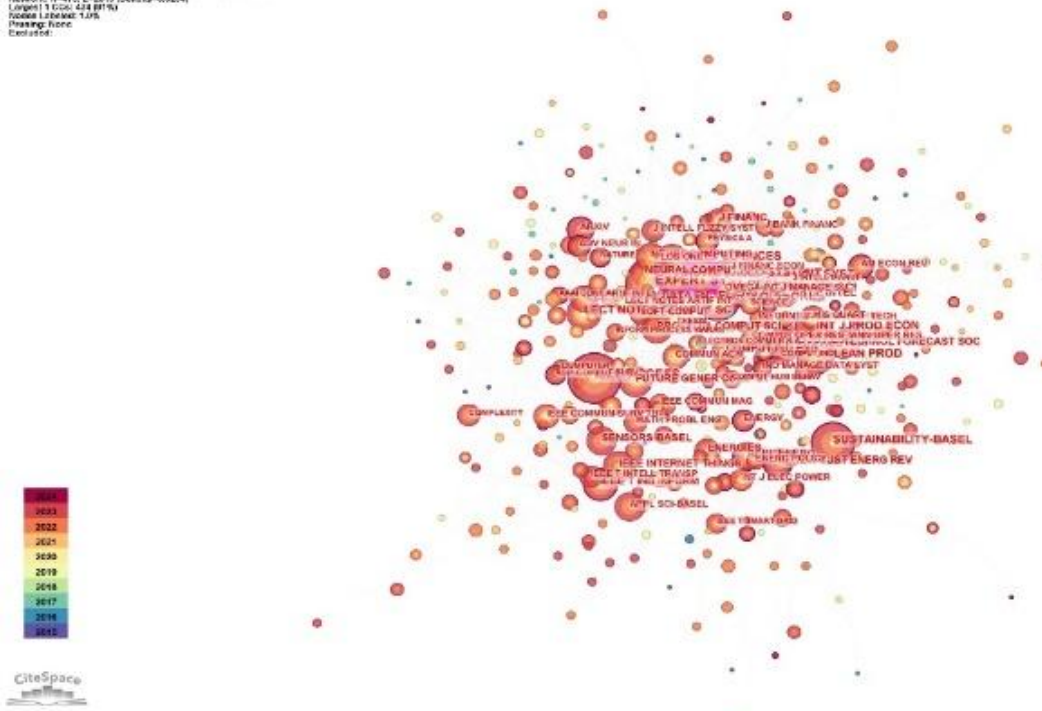


Fig. 3. Co-citation network mapping of core journals.

In the analysis, it was found that core journals such as Knowl-Based Syst, Appl Soft Comput, and Inform Sciences had high co-citations, which reflected the concentration of attention and publication of high-quality research results on the main research themes in these core journals. For example, Knowl-Based Syst's highly cited literature in 2016 points to the application of knowledge systems in research in this field [27]. The citations of Inform Sciences are related to the role of information science in the analysis of tourism data software. IEEE Commun Mag's journals' citations have increased significantly in 2016, showing that they are contributions in the field of communication technology and management science.

By analyzing the co-cited journals over time, it is possible to observe the trend of research themes within the research field. For example, with the development of technology, some emerging research themes such as big data and artificial intelligence have begun to occupy an important position in the journal co-citation network [28]. These trends indicate that research in this field is gradually moving in a technology-driven

direction, while reflecting the high level of academic interest in the practical application of emerging technologies in this research area.

IV. RELEVANT ANALYSIS BASED ON THE FIELD

A. Hot Topic Analysis

In this study, the following core keyword co-occurrence mapping was constructed by CiteSpace 6.3.R1 software with the time slice parameter set to 1 year and the keyword occurrence frequency selection threshold parameter set to 10. The total number of nodes $N=342$, links $E=1269$, network density value 0.0218, through the nodes and links can show the research focus of the field, and the social network connection between the research topic. The larger the node, the more the keyword is proved to be hot, and the more times it co-occurs in the literature [29]. The thicker the linkage, the stronger the connection between the keywords and the deeper the influence is proved. As shown in Fig. 4.

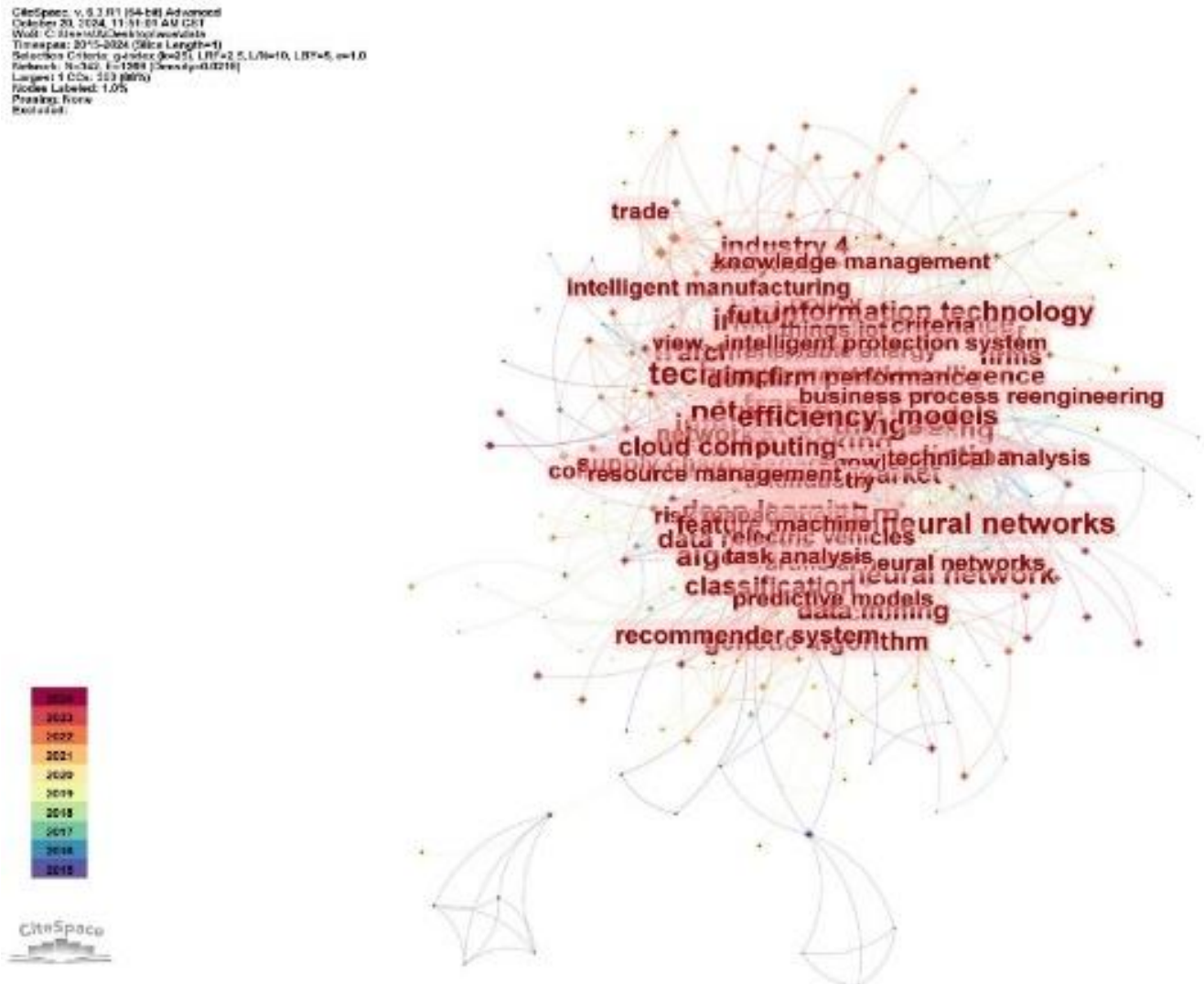


Fig. 4. Core keyword co-occurrence network.

Combined with the core keyword co-occurrence mapping, the keywords with larger nodes and more frequent occurrences can be clearly seen. These core keywords basically cover the hot topics in the research in this field. For example, Model and System, as core keywords, appeared 74 times in 2016 and 40 times each in 2015 and 2016, respectively, showing that scholars have sustained research interests in the construction and assessment of evaluation models in this field [30]. These studies may involve the construction of theoretical models, analysis of system dynamics, and empirical testing of models. Research scholars have attempted to use these models to explain and predict changes in the quality of research and learning activities within the field and their impact on educational outcomes. Artificial Intelligence appeared 43 times in 2015, while Machine Learning appeared 51 times in 2017, which demonstrates that the use of intelligent technologies in the practical applications in this field of study are beginning to gain traction [31]. These studies focus on the use of AI and Machine Learning algorithms to analyze tourism data, predict tourism trends, and enhance the tourism experience in terms of software applications. Management (Management) appeared 50 times in 2017, and the hotspot of research is beginning to shift from theoretical models and software applications to research and study tourism management practices [32]. These studies focus on developing and validating evaluation metrics, as well as exploring the impact of different factors on the field [33]. Internet (Internet) appears 30 times in 2019, while Deep Learning (Deep Learning) appears 24 times in 2020. It shows that in recent years, the hotspots in this research field have been influenced by Internet technology, and research scholars have begun to explore new

research tourism models, Deep Learning to improve the research experience, and evaluation methods [34].

Through the interpretation of the core keyword co-occurrence, it can be observed that the hot topics of research in this field are gradually shifting from the construction of theoretical models to empirical analysis and technological applications, and academic scholars are utilizing intelligent technologies and data analysis to enhance the quality and effectiveness of research and learning in this field. The keyword co-occurrence analysis not only demonstrates the hot topics in this field, but also provides directional guidance for us to deeply understand the research dynamics in this field.

B. Analysis of Hot Areas

In order to further understand the hot field of this research, this paper clusters the core keywords and draws the core keyword clustering map. Among them, Modularity $Q=0.4796$, Silhouette $S=0.7824$. It is not difficult to see from the two queer values that the clustering structure of the hot area constructed in this study is obvious, the internal module similarity is extremely high, and the mapping has high and significant confidence [35]. Using the software LSI algorithm to automatically calculate, filtering out the classification of keyword class group members less than 10, and finally obtaining the hotspot domain clustering of 8 major categories. In order, 00# deep learning algorithm, 01# intelligent manufacturing, 02# assisting investor, 03# Chinese logistics companies, 04# economic complexity, the 05#using reinforcement learning, 06#blockchain technology, 07#sustainable m-commerce as given in Fig. 5.

CiteSpace v. 5.7.R1 (64-bit Advanced)
October 20, 2024, 2:38:33 PM CST
Node: C:\Users\user\Documents\workspace
Workspace: C:\Users\user\Documents\workspace
Selection Criteria: g-index (k=2), LRF=2.0, LRF=10, LRF=6, m=1.0
Network: N=142, E=1289 (Density=0.9216)
Pruning: LRF=1.0
Modularity: Q=0.4796
Weighted Mean Silhouette: S=0.7824
Mean Silhouette: S=0.6341
Classified

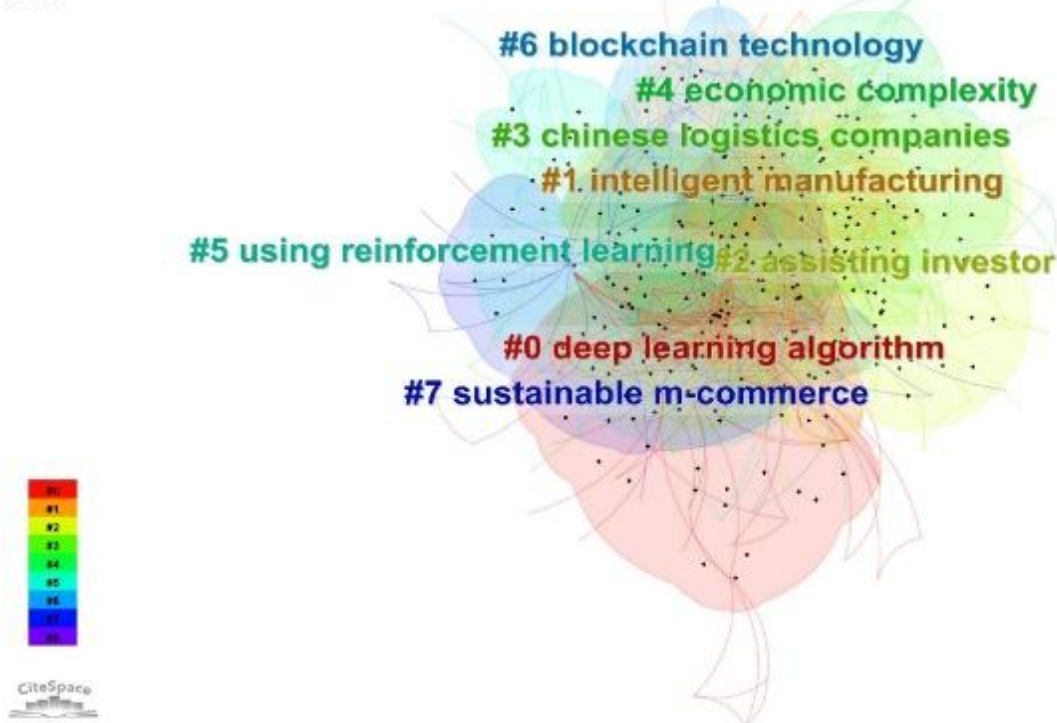


Fig. 5. Core keyword clustering network mapping.

According to the content and topic relevance of keyword clustering, this paper obtained three major research hotspot areas with high similarity or high impact relationship by clustering the eight hotspot areas again. They are technology and algorithmic innovation, industry and economic development, and education and social impact areas.

Hot Area 1: Technology and Algorithmic Innovation

This clustering covers 00# Deep Learning Algorithm, 05# Using Reinforcement Learning and 06# Blockchain Technology. The application of these technologies in this research area demonstrates the strong academic interest in utilizing advanced technologies to enhance tourism experiences and evaluation methods. Deep learning algorithms show great potential in handling tourism big data analytics, personalized recommendation systems and intelligent decision support systems. Reinforcement learning, on the other hand, plays a role in dynamically optimizing tourism strategies and enhancing user interaction experience [36]. Blockchain technology, on the other hand, focuses on improving the security and transparency of tourism transactions, especially in tourism supply chain management and traceability of tourism products, and technological and algorithmic innovations have driven the rapid development of this research area.

Hot Area 2: Industry and Economic Development

01# Intelligent Manufacturing, 03# Chinese Logistics Companies and 4# Economic Complexity constitute the field of industry and economic development, which focuses on the economic impact of study tours and how to promote industrial upgrading and economic development through study tours. The research in this area focuses on the impact of study tours on the economy and how to promote industrial upgrading and economic development through study tours. The co-occurrence of keywords in the clusters intuitively demonstrates the close connection between study tourism and industrial development and economic dynamics [37]. The application of smart manufacturing technologies promotes innovation in the tourism industry by improving the quality and productivity of tourism products [38]. The involvement of logistics companies highlights the importance of efficient logistics in safeguarding the tourism experience and improving the quality of tourism services. The study of economic complexity focuses on the impact of the macroeconomic environment on the development of study tourism and how to maintain the stable growth of the tourism industry under complex and changing economic conditions.

Hot Area 3: Education and Social Impact Clustering

The third hot area of research, to be composed of the clustering of 02# Assisting Investor, 07# Sustainable M-Commerce and 08# Innovation, demonstrates the importance of this area of research both at the educational level and at the societal level. Research in assisting investors focuses on attracting investment through research and study tourism programs and optimizing tourism products and services to enhance return on investment [39]. Research in sustainable m-commerce focuses on the use of mobile technology in

environmental tourism practices and research and study tourism sustainability [40]. The application of Internet technology plays a key role in facilitating tourism information sharing, enhancing tourism experience and improving the efficiency of tourism services.

In summary, the core keyword clustering analysis reveals both the research hotspots in this research field and reflects the intrinsic connection between the hotspots. The clustering of these research areas provides a direction for the research in this field and provides theoretical support for scholars when practicing their work. With technological advances and changes in the global economic environment, these research hotspots change accordingly, bringing new research opportunities and challenges to the research field.

C. Trend Analysis of Dynamic Evolution

In the in-depth analysis of this research field, keyword time mapping provides a unique perspective, prompting research scholars to observe the dynamic evolutionary trends of research hotspots within the field. In this paper, we construct a map of the dynamic evolutionary trend of research hotspots from 2015 to 2024 by using the time zone mapping function of CiteSpace 6.3.R1 software. Each node represents a keyword, and the keyword is fixed in the year of its first appearance. The larger the node, the higher the frequency of the keyword; the longer the arc and the darker the color, the higher the attention of the keyword research and the longer the duration. Along the evolutionary pulse of the keywords, the temporal development of the research hotspots in the field is excavated, so as to explore the future development trend of the research field. As shown in Fig. 6.

From the time mapping, it is easy to find that most of the arcs lasted between 5-7 years, and some keywords lasted for 2-3 years. Among them, Models (models) appeared with higher frequency and core values, reaching a peak in 2016, research scholars in this field began to focus on constructing research and study tourism evaluation models, which may be related to the exploration of evaluation methods and tools at that time, but this concern quickly disappeared in 2017, and this disappearance may be that the construction of models has already matured or that the research in this field was replaced by emerging hotspots. Systems emerged as a research hotspot gradually from 2017 and lasted for two years. During this period, research scholars regarded research tourism as a complete system, and began to pay attention to the interaction and overall optimization between its internal elements, and the systematic way of thinking promoted the development of the whole research field. By 2018, the keyword Behavior appeared, pointing to the research scholars' in-depth and focused research on the behavioral patterns of tourists, trying to interpret their impact on the quality of research and study tourism from the direction of tourists' needs, preferences and behaviors [41]. In 2020, two different sub-directions of research emerged, namely Recommender Systems and Risk Management, with some researchers focusing on the application and impact of recommender systems on personalized services in study tours, while others began to turn to the potential risks and risk management in the tourism process.

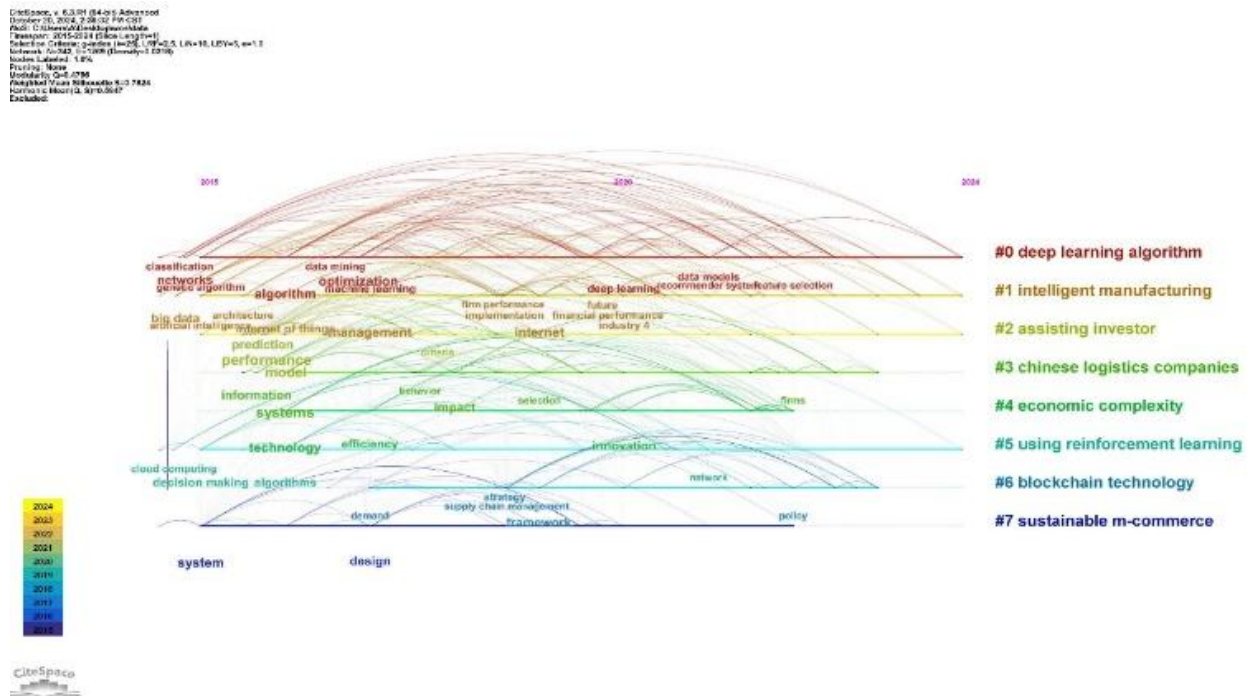


Fig. 6. Time mapping of core keywords.

In summary, the core keywords are distributed across the literature studies in the middle of the decade 2015-2024, and their durations provide a window for researchers to understand the dynamics of research in the field as keywords appear and disappear, as well as the potential connections between different research directions within the research field. By analyzing the temporal changes of these keywords, changes in the research trends in the field are inferred, and subsequently possible future research hotspots are predicted [42]. For example, the prominence of recommender systems may signal that personalized tourism services will become an important direction in the evaluation of study tours, while the prominence of risk management may point to the growing importance of tourism safety and stability evaluation.

V. CONCLUSION

A. Conclusions of the Study

This study provides an in-depth visual analysis of the literature in the field of quality assessment in research tourism through CiteSpace 6.3.R1 software, using knowledge mapping and data statistics to reveal several important aspects of the research dynamics in this field.

1) By statistically summarizing and analyzing the number of literature releases during the decade of 2015-2024, it is found that the research in this field has experienced three distinct development phases: starting and exploration, rapid growth and diversification, and adjustment and maturity. In the starting and exploring stage, the research mainly focuses on building the basic framework and model development. With the development of technology, the research hotspots in this field have proliferated and deepened, and the research aspects have started to cover a wider range of topics, such as user behavior

analysis and recommender systems. At the stage of adjustment and maturity, the research focus is further concentrated and deepened, showing the trend that the research field is gradually developing in a deeper direction.

2) In the statistical analysis of the authors of the literature in this field, the author noticed that there are relatively few highly productive authors, only 52 authors out of 619 authors have published two pieces of literature. Through the author collaboration mapping, it is intuitively observed that the core authors have fewer communication links with each other and have not yet formed a clear core group of authors. Although the research participation in this field is extremely broad, there is a lack of sustained research output and in-depth academic collaboration. This finding suggests that we need to strengthen cooperation and communication among scholars in future research to promote knowledge accumulation and academic innovation.

3) From the statistics of the number of journal articles and cited journals in this field, it is easy to see that this research field is an interdisciplinary field that involves the integration of multiple disciplines in innovative research. The research results in this field are widely distributed in various types of journals, including specialized journals in the fields of tourism, education, management and information technology. This interdisciplinary nature provides a wealth of perspectives and methodologies for research in the field, but it also poses the challenge of research integration and knowledge sharing.

4) Through the keyword sharing, clustering and emergence analysis, it is found that the research hotspots in this field show a dynamic evolutionary trend change over time, from system construction and model development in the early stage to user

behavior analysis and recommender system research in the later stage, which reflects the impact of technological advancement and social development on the research in this field. In particular, the application of emerging technologies such as deep learning, reinforcement learning, blockchain technology and artificial intelligence has provided new research tools and methods for this research field.

B. Research Limitations

In this study, despite the comprehensive visualization and analysis of the literature in the research area through CiteSpace software, there are still some limitations that may have had an impact on the comprehensiveness of the findings and the depth of the study. First, the selection of the study sample was limited to the literature included in the Web of Science database, which may mean that the source data failed to cover all relevant studies, especially those published in regional or specialized journals, and this choice may have resulted in the analysis results not being fully representative of the current state of research in the whole research area. Second, the keyword co-occurrence analysis, which is mainly based on the keyword fields of the literature, although it can show the research hotspots and research trends, fails to fully capture the depth and diversity of the literature content, and some important research topics may not be covered by the keywords or fully reflected in the keyword fields in the titles of the literature. In addition, the time mapping analysis, while showing the evolution of hotspots in the research field, does not delve into the social, economic and policy factors behind these changes. These environmental factors may have a significant impact on the research hotspots and trends in the field, and the analysis of the current study fails to adequately consider these external variables.

C. Future Prospects

In response to these limitations, future research will go back to expanding the scope of literature samples to include literature data from multilingual and multiregional countries in order to gain a more comprehensive research perspective. In terms of research methodology, the theoretical models, methodologies and empirical studies proposed in the literature will be explored in depth in conjunction with the content analysis methodology in order to more accurately understand the research hotspots and trends. External variables such as social, economic and policy contextual factors will be added to analyze their impact on the research hotspots and trends in the field, so as to find deeper research motivations.

REFERENCES

- [1] Leong, L.-Y., Hew, T.-S., Tan, G. W.-H., Ooi, K.-B., & Lee, V.-H. Tourism research progress – a bibliometric analysis of tourism review publications[J]. *Tourism Review*, 2021, 76(1), 1–26. <https://doi.org/10.1108/TR-11-2019-0449>
- [2] Streimikiene, D., Svagzdiene, B., Jasinskas, E., & Simanavicius, A. Sustainable tourism development and competitiveness: The systematic literature review[J]. *Sustainable Development*, 2021, 29(1), 259–271. <https://doi.org/10.1002/sd.2133>
- [3] Mariani, M., & Baggio, R. Big data and analytics in hospitality and tourism: A systematic literature review[J]. *International Journal of Contemporary Hospitality Management*, 2022, 34(1), 231–278. <https://doi.org/10.1108/IJCHM-03-2021-0301>
- [4] Salmela, T., Nevala, H., Nousiainen, M., & Rantala, O. Proximity tourism: A thematic literature review[J]. *Matkailututkimus*, 2022, 17(1), 46–63. <https://doi.org/10.33351/mt.107997>
- [5] Qiao, G., Ding, L., Zhang, L., & Yan, H. Accessible tourism: A bibliometric review (2008–2020)[J]. *Tourism Review*, 2022, 77(3), 713–730. <https://doi.org/10.1108/TR-12-2020-0619>
- [6] Adamus-Matuszyńska, A., Dzik, P., Michnik, J., & Polok, G. Visual Component of Destination Brands as a Tool for Communicating Sustainable Tourism Offers[J]. *Sustainability*, 2021, 13(2), Article 2. <https://doi.org/10.3390/su13020731>
- [7] Gelter, J., Lexhagen, M., & Fuchs, M. A meta-narrative analysis of smart tourism destinations: Implications for tourism destination management[J]. *Current Issues in Tourism*, 2021, 24(20), 2860 – 2874. <https://doi.org/10.1080/13683500.2020.1849048>
- [8] Gelter, J., Fuchs, M., & Lexhagen, M. Making sense of smart tourism destinations: A qualitative text analysis from Sweden[J]. *Journal of Destination Marketing & Management*, 2022, 23, 100690. <https://doi.org/10.1016/j.jdmm.2022.100690>
- [9] Karyatun, S., Efendi, S., H. Demolingo, R., Wiweka, K., & Putri, A. P. Between Instagrammable Attraction and Selfie Tourist: Characteristic and Behavior[J]. *South Asian Journal of Social Studies and Economics*, 2021, 12(4), Article 4. <https://doi.org/10.9734/sajsse/2021/v12i430338>
- [10] Li, J., Weng, G., Pan, Y., Li, C., & Wang, N. A scientometric review of tourism carrying capacity research: Cooperation, hotspots, and prospect[J]. *Journal of Cleaner Production*, 2021, 325, 129278. <https://doi.org/10.1016/j.jclepro.2021.129278>
- [11] Moral-Cuadra, S., Solano-Sánchez, M. Á., Menor-Campos, A., & López-Guzmán, T. Discovering gastronomic tourists' profiles through artificial neural networks: Analysis, opinions and attitudes[J]. *Tourism Recreation Research*, 2022, 47(3), 347 – 358. <https://doi.org/10.1080/02508281.2021.2002630>
- [12] Prawira, N. G., Susanto, E., & Prawira, M. F. A. Visual Branding on Indonesian Tourism Destinations: Does it Affect Tourists? [J] *ABAC Journal*, 2023, 43(1), Article 1. <https://doi.org/10.14456/abacj.2023.4>
- [13] Wei, Y., & Wu, T. Visual representation of a linear tourist destination based on social network photos: A comparative analysis of cross-cultural perspectives[J]. *Journal of Tourism and Cultural Change*, 2021, 19(6), 781–804. <https://doi.org/10.1080/14766825.2020.1849239>
- [14] Sun, Y., & Hou, G. Analysis on the Spatial-Temporal Evolution Characteristics and Spatial Network Structure of Tourism Eco-Efficiency in the Yangtze River Delta Urban Agglomeration[J]. *International Journal of Environmental Research and Public Health*, 2021, 18(5), Article 5. <https://doi.org/10.3390/ijerph18052577>
- [15] Wang, J., & Lv, W. Tourism poverty alleviation hotspots in China: Topic evolution and sustainable development[J]. *Sustainable Development*, 2023, 31(3), 1902–1920. <https://doi.org/10.1002/sd.2492>
- [16] Yin, L. J., Zhang, N., & Chang, Z. Y. Study on the impact of tourism quality perception on tourists' environmentally responsible behaviour in rural tourism areas[J]. *IOP Conference Series: Earth and Environmental Science*, 2021, 626(1), 012015. <https://doi.org/10.1088/1755-1315/626/1/012015>
- [17] Bakker, M., van der Duim, R., Peters, K., & Klomp, J. Tourism and Inclusive Growth: Evaluating a Diagnostic Framework[J]. *Tourism Planning & Development*, 2023, 20(3), 416 – 439. <https://doi.org/10.1080/21568316.2020.1850517>
- [18] Qiu, Q., & Zhang, M. Using Content Analysis to Probe the Cognitive Image of Intangible Cultural Heritage Tourism: An Exploration of Chinese Social Media[J]. *ISPRS International Journal of Geo-Information*, 2021, 10(4), Article 4. <https://doi.org/10.3390/ijgi10040240>
- [19] Işık, C., Aydın, E., Dogru, T., Rehman, A., Sirakaya-Turk, E., & Karagöz, D. Innovation Research in Tourism and Hospitality Field: A Bibliometric and Visualization Analysis[J]. *Sustainability*, 2022, 14(13), Article 13. <https://doi.org/10.3390/su14137889>
- [20] Sánchez-Franco, M. J., & Rey-Tienda, S. The role of user-generated content in tourism decision-making: An exemplary study of Andalusia, Spain[J]. *Management Decision*, 2024, 62(7), 2292–2328. <https://doi.org/10.1108/MD-06-2023-0966>

- [21] Jia, Y., Ouyang, J., & Guo, Q. When rich pictorial information backfires: The interactive effects of pictures and psychological distance on evaluations of tourism products[J]. *Tourism Management*, 2021, 85, 104315. <https://doi.org/10.1016/j.tourman.2021.104315>
- [22] Ülker, P., Ülker, M., & Karamustafa, K. Bibliometric analysis of bibliometric studies in the field of tourism and hospitality[J]. *Journal of Hospitality and Tourism Insights*, 2023, 6(2), 797–818. <https://doi.org/10.1108/JHTI-10-2021-0291>
- [23] Le Busque, B., Mingoia, J., & Litchfield, C. Slow tourism on Instagram: An image content and geotag analysis[J]. *Tourism Recreation Research*, 2022, 47(5–6), 623–630. <https://doi.org/10.1080/02508281.2021.1927566>
- [24] Farokhi, S., Namamian, F., Asghari Sarem, A., & Ghobadi Lamuki, T. Explaining the visual attention model in impulsive purchase behavior of tourism industry customers by theme analysis method[J]. *Journal of Islamic Marketing*, 2024, 15(1), 279–292. <https://doi.org/10.1108/JIMA-05-2022-0136>
- [25] Huang, Z., Weng, L., & Bao, J. How do visitors respond to sustainable tourism interpretations? A further investigation into content and media format[J]. *Tourism Management*, 2022, 92, 104535. <https://doi.org/10.1016/j.tourman.2022.104535>
- [26] Leiras, A., & Eusebio, C. Perceived image of accessible tourism destinations: A data mining analysis of Google Maps reviews[J]. *Current Issues in Tourism*, 2024, 27(16), 2584 – 2602. <https://doi.org/10.1080/13683500.2023.2230338>
- [27] Meng, S., Li, H., & Wu, X. International cruise research advances and hotspots: Based on literature big data[J]. *Frontiers in Marine Science*, 2023, 10. <https://doi.org/10.3389/fmars.2023.1135274>
- [28] Pelit, E., & Katircioglu, E. Human resource management studies in hospitality and tourism domain: A bibliometric analysis[J]. *International Journal of Contemporary Hospitality Management*, 2022, 34(3), 1106–1134. <https://doi.org/10.1108/IJCHM-06-2021-0722>
- [29] Wijaya, I. N. C. Exploring the Evolution and Prospects of Gastronomy Tourism Development in Tista Tourism Village, Tabanan: A Comprehensive Analysis[J]. *International Journal of Global Tourism*, 2023, 4(3), 235–244. <https://doi.org/10.58982/injogt.v4i3.498>
- [30] Wang, M., Liu, S., & Wang, C. Spatial distribution and influencing factors of high-quality tourist attractions in Shandong Province, China[J]. *PLOS ONE*, 2023, 18(7), e0288472. <https://doi.org/10.1371/journal.pone.0288472>
- [31] Volo, S. The experience of emotion: Directions for tourism design[J]. *Annals of Tourism Research*, 2021, 86, 103097. <https://doi.org/10.1016/j.annals.2020.103097>
- [32] Liu, R., Huang, Z., Yu, R., Bao, J., & Mo, Y. The impact of red tourism on national identity of tourists[J]. *Journal of Natural Resources*, 2021, 36(7), 1673–1683. <https://doi.org/10.31497/zrzyxb.20210704>
- [33] Shang, Y., Wen, C., Bai, Y., & Hou, D. A novel framework for exploring the spatial characteristics of leisure tourism using multisource data: A case study of Qingdao, China[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2022, 15, 6259 – 6271. <https://doi.org/10.1109/JSTARS.2022.3196002>
- [34] Zuo, Y., Chen, H., Pan, J., Si, Y., Law, R., & Zhang, M. Spatial distribution pattern and influencing factors of sports tourism resources in China[J]. *ISPRS International Journal of Geo-Information*, 2021, 10(7), 428. <https://doi.org/10.3390/ijgi10070428>
- [35] Jin-wei, W., Guo-quan, W., Yi, L., Ting, L., Jie, S., & Xin, W. Spatio-temporal distribution and network structure of red tourism flow in Jinggangshan[J]. *Journal of Natural Resources*, 2021, 36(7), 1777–1791. <https://doi.org/10.31497/zrzyxb.20210711>
- [36] Yapici, O. O. Bibliometric Analysis of Smart Cities and Tourism Studies with Visual Mapping Technique. *Revista Rosa Dos Ventos - Turismo e Hospitalidade*, 2022, 14(3), Article 3.
- [37] Ye, C., Zheng, R., & Li, L. The effect of visual and interactive features of tourism live streaming on tourism consumers' willingness to participate[J]. *Asia Pacific Journal of Tourism Research*, 2022, 27(5), 506–525. <https://doi.org/10.1080/10941665.2022.2091940>
- [38] Liu, J., Wei, W., Zhong, M., Cui, Y., Yang, S., & Li, H. A bibliometric and visual analysis of hospitality and tourism marketing research from 2000–2020[J]. *Journal of Hospitality and Tourism Insights*, 2023, 6(2), 735–753. <https://doi.org/10.1108/JHTI-10-2021-0277>
- [39] Zhang, J., Xiong, K., Liu, Z., & He, L. Research progress and knowledge system of world heritage tourism: A bibliometric analysis[J]. *Heritage Science*, 2022, 10(1), 42. <https://doi.org/10.1186/s40494-022-00654-0>
- [40] Liu, X., Zeng, Y., He, J., & Li, Z. Value cocreation research in tourism and hospitality: A comparative bibliometric analysis[J]. *International Journal of Contemporary Hospitality Management*, 2022, 34(2), 663–686. <https://doi.org/10.1108/IJCHM-05-2021-0666>
- [41] Qiao, G., Xu, J., Ding, L., & Chen, Q. The impact of volunteer interaction on the tourism experience of people with visual impairment based on a mixed approach[J]. *Current Issues in Tourism*, 2023, 26(17), 2794–2811. <https://doi.org/10.1080/13683500.2022.2098093>
- [42] Atabay, E., & Güzeller, C. O. A Bibliometric Study on Eye-Tracking Research in Tourism[J]. *Tourism: An International Interdisciplinary Journal*, 2021, 69(4), 595–610. <https://doi.org/10.37741/t.69.4.8>

Internet of Things (IoT) Driven Logistics Supply Chain Management Coordinated Response Mechanism

Chong Li

School of Economics and Management, Beihua University, Jilin 132000, Jilin, China

Abstract—This study explores the development of an IoT-driven logistics supply chain coordination and response mechanism aimed at achieving real-time information sharing, precise forecasting, and rapid decision-making among supply chain nodes. By employing a hierarchical system construction method, SQL database techniques for data management, and an evaluation model combining AHP and entropy methods, the study proposes a robust framework for improving supply chain efficiency and adaptability. The results demonstrate that IoT technology significantly enhances supply chain transparency, resource allocation, and operational efficiency while reducing risks and costs. The proposed mechanism facilitates dynamic adjustments to market changes and unexpected disruptions, fostering a resilient and collaborative supply chain network. This research provides a foundational basis for the integration of IoT in modern supply chains and offers insights into advancing intelligent logistics systems, with implications for improving global competitiveness in the evolving digital economy.

Keywords—IoT; logistics supply chain; management coordination; response mechanism

I. INTRODUCTION

In the new post-epidemic normal, the manufacturing industry urgently seeks to accelerate its digitization process and smart warehouse and logistics upgrades in light of the continued climb in market demand for smart warehousing and logistics [1]. In the ecology of logistics and supply chain, the efficiency of the warehouse and logistics system, service quality, and operating costs constitute the core considerations. By optimizing the intelligent warehousing and logistics system, enterprises can effectively speed up the logistics process, ensure efficient resource deployment and management, and then cut costs and improve overall performance [2]. In the construction of intelligent warehousing, promote the level of informatization of warehousing and logistics to enhance the irreversible development trend in the field of intelligent manufacturing.

In recent years, with the deep penetration of Internet of Things (IoT) technology in the industrial sector, the Industry 4.0 era has witnessed an increasing convergence of technologies [3]. This cutting-edge concept is gradually attracting the attention of the industry, which is dedicated to the comprehensive digital mapping of physical entities, empowering the intelligent design, manufacturing, commissioning, and full lifecycle management of physical equipment through data access and model-driven. Specifically, the application of digital technology in the field of warehouse logistics, embodied in the construction of digital

subsystems to provide intuitive monitoring and predictive maintenance services, at the same time, it also helps the simulation and pre-commissioning of industrial equipment in the virtual environment, to provide a virtual test platform for the development of the production plan, to effectively assess the feasibility of the production process, the initial cost can be reduced, and production efficiency has been significantly improved [4]. On the other hand, in real production environments, manufacturers often need to integrate different industrial equipment from multiple suppliers, and the diverse communication protocols and interfaces between these devices often become obstacles to the efficient collection, transmission, and processing of data, which exacerbates the difficulty of data sharing [5]. This phenomenon, the so-called "knowledge silo", is gradually evolving into one of the challenges in promoting the digital and intelligent transformation of production [6].

This research aims to explore and construct a set of logistics supply chain coordination response mechanisms based on IoT test the constructed platform to a certain extent, and evaluate the IoT platform according to the test results and subjective and objective factors to explore the level of logistics supply chain management coordination response in China.

II. BACKGROUND OF THE STUDY

As a key pillar of the national economy, the logistics industry is of strategic importance in shaping the framework of the modern cycle, driving high-quality development, and building a modernized economic system [7]. In 2022, the General Office of the Government issued a five-year blueprint for the development of modern logistics, which is the first five-year planning document for the development of modern logistics in China [8]. The blueprint specifies the core tasks for the next five years: to enhance the innovation driving force and market competitiveness of logistics enterprises, to optimize the quality and efficiency of logistics services, and to build a hub network operation system and a modern logistics and distribution network to flexibly respond to changes in domestic and international supply and demand [9]. At present, China has jumped to the top of the global logistics market, carrying more than half of the world's express parcel volume. However, despite the huge scale of logistics, its comprehensive strength has yet to be strengthened. Therefore, accelerating the construction of a modern and efficient logistics system, and realizing the leap from "big" to "strong" has become a key mission to improve logistics quality, cut costs, and enhance efficiency, which is particularly urgent in the five-year planning period [10].

Logistics companies need to have a forward-looking vision, advance insight into industry trends, carefully plan the future path, and continue to innovate and optimize, to better meet the challenges and achieve sound management and long-term development [11].

The continuous prosperity of the logistics industry has prompted intelligent logistics to become an increasingly critical growth trend. This trend not only can effectively deal with the information asymmetry, lack of transparency inefficiency, and other constraints in the logistics industry, but also greatly improves the efficiency of logistics operations and service quality, accurately meeting the expectations of both the logistics industry and consumers, and gives the logistics experience better, faster and more convenient characteristics [12]. In addition, intelligent logistics is accelerating the digital transformation of the logistics industry, helping the logistics supply chain synergy and fine-tuned operation, and realizing the double benefits of economic and social growth [13]. Given the rapid progress of artificial intelligence technology and its in-depth penetration in various fields, the vision of "comprehensive monitoring, seamless links, intelligent supply" is gradually becoming a reality. To achieve integrated management and

intelligent strategy development in logistics operations, IoT technology and data analysis means that the nodes in the logistics planning network can instantly capture, receive, and transmit real-time data to the core command system, realizing a seamless flow of data. This has undoubtedly strengthened the foundation of intelligent logistics and significantly enhanced its market competitiveness [14]. It is worth noting that road transportation dominates China's cargo transportation pattern, accounting for 73.8% in 2020, so it is urgent to build an IoT-driven supply chain to serve road transportation.

Therefore, to improve the quality of logistics services and transportation efficiency, in the layout of intelligent logistics, it is necessary to make use of advanced tools such as Internet of Things (IoT) technology, data analysis means, and optimization algorithms to solve a series of key technical difficulties [15]. The application of these technologies shows great potential in enhancing the real-time monitoring capability of logistics services, intelligent identification accuracy, and promoting the synergistic optimization of various links [16]. Therefore, for logistics planning and decision-making, in-depth exploration and application of these technologies have extremely important research significance and practical value.

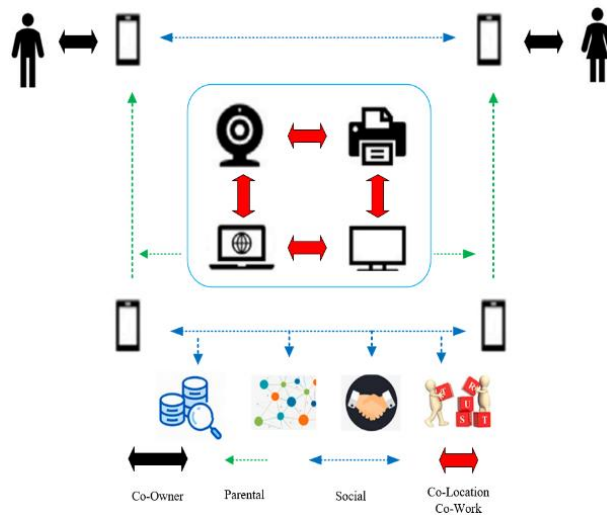


Fig. 1. Correlation of traditional IoT technologies.

III. RESEARCH METHODOLOGY

A. IoT Theory and Technology Foundations

At the heart of smart logistics is the efficient use of cutting-edge technology, with IoT dominating the logistics sector and being widely used. IoT, as a unique and recognizable network of "things", is centered on the identification and aggregation of personalized information about "things" through sensors/controllers connected to the Internet. IoT, as a unique and recognizable network of "things", is centered on the personalized identification and aggregation of "things" through sensors/controllers connected to the Internet, a process that integrates electronic devices, Internet connectivity, and multiple devices such as sensors [17]. These integrated devices can interact with infrastructure such as cloud servers and respond and operate quickly to dynamically changing environments and situations. The IoT architecture can be subdivided into four

layers: sensing, transmission, processing, and application layers. The sensing layer, as the core layer of IoT devices, is responsible for capturing and quantifying the physical characteristics of various types of objects and is composed of diverse sensors, RFID tags, and other sensing networks [18]. This layer captures data from the sensors directly associated with the objects and subsequently converts this raw data into digital signals and passes them to the transmission layer [19]. The transmission layer, on the other hand, plays the role of a bridge, using a variety of digital communication methods such as Bluetooth, ZigBee, 5G, etc., to receive and transmit these data signals from the sensing layer. The traditional IoT technology is known as SIoT technology and its IoT technology correlation is shown in Fig. 1.

IoT system integration not only covers the functions at the basic data level but also provides data processing services and support for specific application requirements. Based on the

organizational structure of the enterprise, it flexibly provides diversified business support services [20]. Its building blocks include hardware platforms such as cloud services, big data technologies, artificial intelligence, and advanced algorithms, as well as a series of middleware (e.g., sensor network gateways, sensor network security middleware, embedded M2M middleware, etc.), which together weave a powerful technology network [21]. For example, the mid-tier capabilities of intelligent computing are fully utilized, while cloud services are transformed into a hub for remote data storage and processing, greatly facilitating the process of accessing, storing, and processing data.

B. Key Technologies

In the field of identification technology for LoT, Radio Frequency Identification (RFID) technology occupies a dominant position. This technology relies on radio waves for data collection and identification and constitutes a unique identification system. The system consists of an RFID tag, a reading device equipped with a transmitter, and an interface to the target system. Specifically, the RFID tag has a built-in microchip and antenna assembly; when the tag enters a specific magnetic field, it receives a frequency signal from the reading device. The microchip on the tag serves as the core of data storage, carrying relevant information about the target object [22]. The antenna acts as a data transmission medium, enabling the microchip to pass information about the object to the reading device. The reading device then converts the RFID identification data into a format that can be easily processed by a computer, allowing the user to easily identify a specific object or individual.

Io Big Data enables flexible access to a diverse set of computing resources - servers, network architectures, storage, applications, and services - that are widely accessible and on demand. Such resources can be rapidly provisioned into place with minimal control and interaction between the user and the service provider, aiming to maximize data processing efficiency and capacity [23]. This model contributes significantly to the efficiency of data processing and computing. On the other hand, the big data technology system focuses on the all-round processing and insight of large-scale data sets, covering data collection, storage, refined processing, information extraction, and visualization. This process empowers users to dig deeper into the value of massive data, revealing the valuable information hidden behind the data. Through in-depth analysis and application exploration of big data technology, we can more thoroughly understand big data-related devices and their practical scenarios, which in turn will promote the performance leap and continue to optimize the user's interactive experience.

The technique of processing and analyzing real-time large-scale event streams can be called integrated event processing [24]. This process involves searching and organizing basic events to construct more complex and advanced events. Streaming analytics enables users to instantly monitor and analyze input data to dissect the causal chain of events and derive conclusions based on specific complex events. Low-level events are either manifested as a series of activities within a timeframe or as a bridge between events from different sources, and these elements are synthesized from a variety of data sources to shape complex event models that can be used to predict,

manage, and regulate potential events, scenarios, and biases. At the core of CEP (Complex Event Processing) technology is the uninterrupted processing and profiling of massive streams of high-speed data (e.g., RFID data), which are combined with a variety of data streams, and are analyzed uninterruptedly. The core of CEP (Complex Event Processing) technology lies in the uninterrupted processing and analysis of massive high-speed data streams (such as RFID data), which are fused with decentralized data to enable immediate monitoring and response to critical business scenarios[25]. The efficient Complex Event Processor can react quickly to hidden patterns, correlations, and data abstractions between apparently unconnected events based on predefined rules. It can be viewed as a continuously operating intelligent application designed to maximize the overall value of challenging events, provide immediate decision support for diverse event scenarios, and enhance overall situational awareness and understanding.

C. IoT Heuristic Exact Algorithms

The solution of numerous complex optimization challenges often relies on metaheuristic algorithms that are viewed as efficient tools. These algorithms can be strategically differentiated into individual solutions for a single case and generalized solutions that are universally applicable. Among them, innovative practices in individual solving encompass unique strategies such as large-network search (LNS) and adaptive large-network search. Population-based metaheuristics, on the other hand, have the core goal of breeding novel solutions by integrating and adapting existing solution sets, or by promoting synergies between solutions in the learning process [26]. Among the many heuristic algorithms inspired by natural biological processes, genetic algorithms stand out for their generality. In each iteration, the algorithm selects two pairs of parents from the population based on the principle of fitness and generates new solutions (i.e., "offspring") by merging their characteristics through a mechanism that mimics the "crossover" mechanism of biological interbreeding. In addition, to ensure the diversity of solutions in the population, the genetic algorithm introduces a "mutation" procedure as an effective means of facilitating the algorithm's exploration of the unknown solution space.

Eventually, the selected optimal solution is set as the dominant strategy in the next iteration loop. The family of swarm intelligence algorithms includes two members, Ant Colony Optimization (ACO) and Particle Swarm Optimization. Within the framework of these algorithms, ACO algorithms are often complemented with local optimization strategies to solve problems based on a predefined graph and a probabilistic mechanism for deciding their routes. On the other hand, particle swarm optimization algorithms envision a population of many "particles", each of which migrates from one place to another during the exploration process [27]. In this algorithm, the record of the optimal positions of individual particles and the optimal solution for the whole swarm of particles act as guiding factors that influence and shape the trajectory and direction of each particle.

A meta-heuristic local search strategy examines the current state of a solution and promotes the transition from the current solution to a neighboring new solution that shows potential [28]. Taboo search (TS), a widely used algorithm, is rooted in the

strategy of continuous exploration, which does not stop searching even when a local optimum has been reached. Even if the objective function is relaxed, the rheology guarantees the validity of the current solution. Another strategy is the Large Neighborhood Search (LNS), which is known for its ambitious search landscape and shows remarkable search patterns by partially dismantling existing solutions through the removal operator and then restoring them with the reorganization operator [29]. Adaptive SPS strategies are similar but are unique in their ability to dynamically adapt and select the most effective operators in each iteration according to the search process. To avoid the pitfalls of local optimality, guided local search (GLS) broadens the search boundary by imposing constraints on the objective function, thus expanding the exploration domain. In addition, meta-heuristic techniques such as Variable Neighborhood Search (VNS), Stochastic Greedy Adaptive Search Process (GRASP), Simulated Annealing (SA), and Iterative Local Search (ILS) are also focused on the optimization of local search, aiming to jump out of the limitations of local optima [30]. These algorithmic architectures not only serve the direct solution of the problem but also play an important role in constructing the initial solution of the heuristic algorithms, which lays a solid foundation for finding more optimal solutions.

Actuarial algorithm generation process for IoT:

$$\min \sum_{j \in J} c_j x_j \quad (1)$$

$$s.t. \sum_{j \in J} a_{ij} x_j \geq b \quad (2)$$

$$x_j \geq 0 \quad (3)$$

In this case, both Eq. (1) $c_j x_j$ unknown x and constant c are considered to be summed. Eq. (2) $\sum_{j \in J} a_{ij} x_j \geq b$ are constraints, and Eq. (3) both determine the positive and negative values of each variable of the function of x .

$$\overline{c_j} \rightarrow = \min c_j - \sum_{i \in I} \pi_i a_{ij} \quad (4)$$

$$E_i = e(w_{id}, r_{id}, t) \quad (5)$$

In Eq. (4), consider the variable in Eq. (1) as a constant 1 for $\min c_j$, and consider π as a constant variable; in Eq. (5) consider w , r , and t as natural variables.

To assess the positive expectations of the current solution, perform an accounting of the price subcategories and ensure that at least the negatively corresponding columns are included in the updated RMP solution set. An optimal solution is considered to have been reached only if the price analysis does not reveal any unfavorable test results. For RMP problems, the new variables

introduced during each iteration are intended to optimize the current set of variables and their corresponding pairwise optimal solutions, thus ensuring the optimization of the overall solution.

D. Definitional Approach to IoT

The most common means to solve global design challenges and many combinatorial optimization problems is with the help of branching strategies and their associated algorithms. Branching algorithms traverse the entire search domain, identify potential solutions, and select the optimal solution. This process relies on the structure of the search tree for partitioning and definition. A solution tree covering all potential solution paths is constructed, where the root node summarizes the global search pattern and pre-lists possible initial solutions. Each subroutine corresponds to a node presentation in the search tree. The decentralization technique, on the other hand, partitions the solution space into smaller blocks that can be recursively refined, aiming at generating children of unexplored nodes and eliminating suboptimal search patterns that may be confirmed during the partitioning process [31]. After completing the scrutiny of the entire tree structure, the searched optimal solution is fed back to the solution endpoint.

IoT delimitation was originally designed to meet the challenges of various types of constraints or variables. It incorporates a combination of different branching strategies, related algorithms, and column-generation techniques. Similar to the linear relaxation branching method, the original problem is depicted by a Dwolfe distribution. This process first splits the primal problem into a master system with several sub-systems, followed by a consistent path to solving the currently limited master system problem and passing the corresponding binary variables to each sub-category. For each subcategory, a column generation technique is then employed to solve its linear relaxation constraint set. It is worth noting that although the column generation algorithm plays a key role in estimating the node lower bounds, the practical application of the branching strategy is carried out after the linear relaxation solutions are obtained. In addition, it is possible to divide a series of potential primal solution candidate sets into subsets and apply recursive constraints with branching to each subset. In the branch-and-bound framework, the relaxation problem at each node is solved directly while the branch-and-bound pricing process is optimized using a column generation algorithm. By solving local problems iteratively, a series of problem sets optimized in the sense of linear relaxation can be identified. If a linear relaxation problem reaches an optimal state, the optimal solution can be further verified to satisfy the condition of an integer solution. During branching iterations, each new branch introduces unique constraints, resulting in the derivation of new subcategories that are again solved with the help of the column generation algorithm. This process continues, introducing new columns (i.e., variables) into the linear relaxation model until a globally optimal solution to the original problem is found.

The IoT delimitation method is carried out utilizing the branch pricing algorithm, in addition to the main process, where column and integer solutions are examined at the nodes respectively, as shown in Fig. 2.

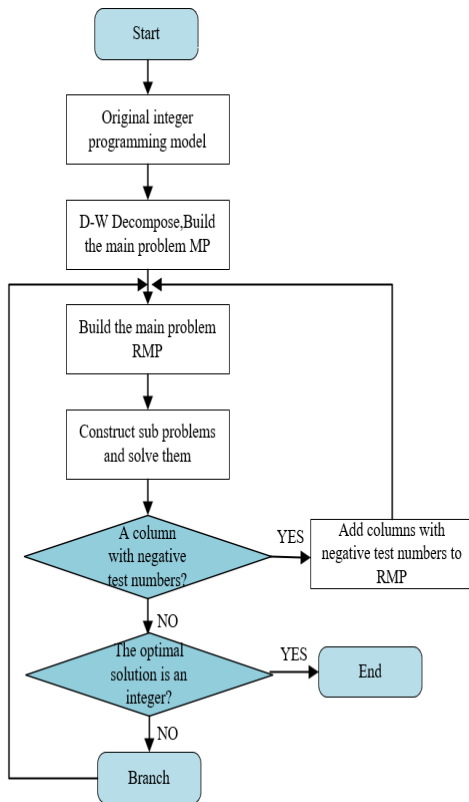


Fig. 2. Branch pricing algorithm flow.

The set of cut strategies constitutes a solution methodology for universal design challenges. The core idea is that multiple linear programming paradigms can be nested within the planning framework, which together portray multiple faces of the same set of integer solutions through a system of transformed linear inequalities. When dealing with a linear relaxation problem and obtaining a fractional, and then additional solution, one realizes that the current scheme is not rigorous enough linear inequality constraints need to be added, aiming at eliminating unrealistic floating-point or generalized solutions, and progressively approximating the globally optimal integer solution. This process is a way of pinpointing the poles of the optimal numerical solution that can be easily solved outside of the widened linear relaxation domain.

The dyadic principle of linear programming, on the other hand, breaks down the complex overall design problem into a series of asymptotically accurate local linear programming subproblems, with each round of solution pushing us closer to the ultimate solution. In the execution of the tangent algorithm, the first step is to solve the base problem with boundaries on all original variables, and then move to the relaxation problem [32]. The algorithm is terminated if a no-solution situation is encountered in the process, or if the best solution presents a generic characterization as non-integer. Instead, an additional linear boundary called a "cut" is introduced, which is designed

to precisely cut out the non-integer solution space that has not yet been covered and integrate it into the current relaxation model to exclude the validity of the current solution, which is then resolved based on the updated model. This process is repeated until all solutions for all basic variables meet the integer requirements.

The delimitation of IoT should also consider the interaction flow between the front and back end, to consider the request to send Axios to start, through the control, service, data, and entity four levels to reach the end of data persistence, as shown in Fig. 3.

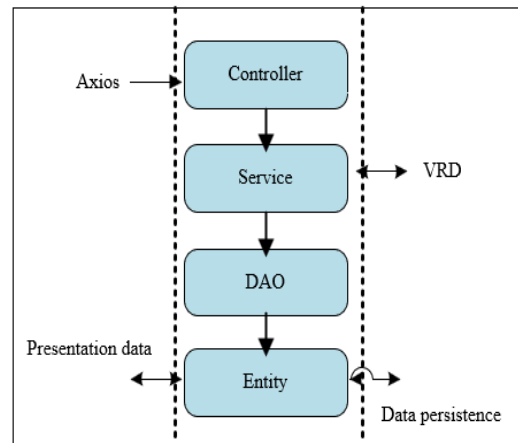


Fig. 3. IoT front and back-end interaction flow.

IV. RESULTS AND DISCUSSION

A. IoT System Architecture

Using the system structure hierarchical construction method to build an IoT platform and commissioning system, the initial intention of building an industrial IoT platform and virtual commissioning system is to rely on a cloud platform to realize instant remote monitoring of the status of each device in the industrial production chain [33]. In addition, the cloud data-driven remote storage line virtual debugging mechanism effectively shortens the on-site troubleshooting cycle and significantly improves the maintenance efficiency and performance during the various stages of production line debugging.

The construction of the IoT platform and debugging system is divided into five layers from the "virtual device layer" to the "virtual simulation debugging layer", as shown in Table I.

The experiments for IoT were divided into 30 sessions, and the number of debugging system fluctuations for the IoT platform is shown as a folded line in Fig. 4 (right axis), and the IoT latency test points are shown as squares in Fig. 4 (left axis) as the summation of the values on the major latency nodes (3, 5, 6, 15, 20, 22, 23, 25, 27, 28, 29, and 30). The IoT platform debugging system fluctuation and delay test results are shown in Fig. 4.

TABLE I. CONSTRUCTION OF IOT PLATFORM AND COMMISSIONING SYSTEM

Virtual Device Layer		Equipment communication control layer	Data Storage Layer	IoT cloud platform layer	Virtual simulation debugging layer
ABB industrial robots	PC-SDK	Data acquisition	Data management	Data monitoring	Data Mapping
Youao Industrial Robot	TCP/IP		Data subscription	Command parsing	Motion driven
Siemens PLC	S7	Data parsing	OPC UA	OPC UA Client	collision detection
KEBA PLC	TCP/IP	Data display	Real-time	MOTT	Visualization
AVG	REST	Data control	HISTORICAL DATA	data display	
RFID	Serial port		MySQL	Command Control	Unity 3D

The communication management layer architecture adopts a separated front and back-end design mode, and its core system module relies on Java and Spring Boot framework to build and realize efficient data interaction and persistence with the database by integrating a hybrid programming paradigm and MyBatis technology. This layer focuses on data processing logic planning and communication protocol management at the virtual device level. The front-end display layer is crafted using the Vue.js framework, focusing on the smoothness of user interface interaction and the flexibility of data configuration. In contrast to traditional project architectures, the UI code is often tightly coupled with Java Server Pages (JSPs), which are located in the backend of the server. In this model, the UI browser needs to retrieve HTML, CSS, and JavaScript resources from the JSP for data visualization and page presentation. This process is accompanied by a large amount of interactive data transfer, resulting in a lengthy and inefficient analysis and processing process, which in turn hinders the long-term maintenance and iterative development of the project. On the contrary, after the implementation of the design principle of separating the UI from the back-end logic, the UI layer actively disengages from the role of direct control of the browser page and focuses on the serialization of data (e.g., JSON, XML, form data, etc.) and recovery work. Front-end pages are loaded independently, which significantly reduces the number of communications between the front and back ends, greatly reduces the complexity of data interaction, effectively reduces the business processing pressure on the back-end server, and thus realizes a significant improvement in the overall performance of the system[34]. Therefore, the specific embodiment of the robustness of the Internet of Things logistics for the first high, then low, and then high, its horizontal axis is the number of iterations, the vertical axis is the robustness, that is, the performance of the performance due to the increase in the number of communications to deteriorate, but also because of the iteration and the enhancement of the process of the number of Internet of Things iterations and the linkage of the robustness, specifically as shown in Fig. 5.

The IoT experiment is divided into 30 times. The fluctuation frequency of the IoT platform's debugging system is shown by the line in Figure 4 (right axis), and the IoT delay test integral is shown by the block in the figure (left axis), which is the sum of the values on the key nodes (3, 5, 6, 15, 20, 22, 23, 25, 27, 28, 29, 30).

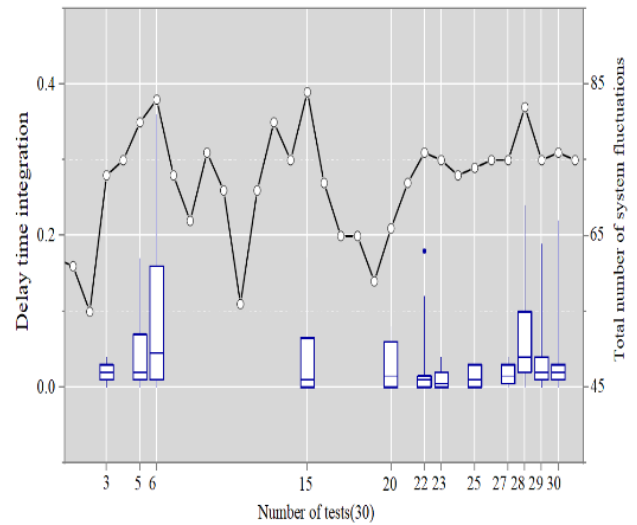


Fig. 4. IoT platform debugging results.

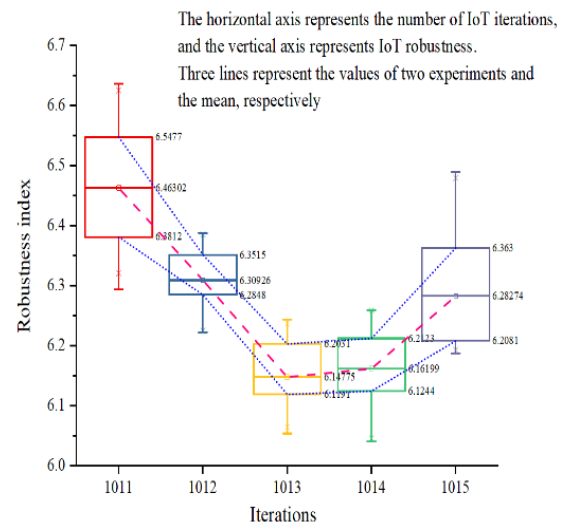


Fig. 5. IoT iteration count and robustness linkage.

In the heuristic algorithm mentioned in 3.3 above, the heuristic actuarial algorithm columns are generated in such a way that firstly the first row generates a column Ω_1 , followed by the algorithmic disaggregation using Do and while as shown in Table II.

TABLE II. HEURISTIC ACTUARIAL ALGORITHM COLUMN GENERATION

column generation
Generate an initial set of columns Ω_1
Do
Calculate the answer MP
Γ : New columns obtained from sub-problems
$\Omega_1 \cup \Gamma$
While $\Gamma \neq \theta$

The core concept of microservices architecture is to refine the huge software system into a series of independent, different functions of the service unit, the collaboration between these units does not interfere with the integrity of the business logic. Spring Cloud is the Spring ecosystem for the development of microservices framework, and support for the integration of Spring Boot Starter to enhance the scalability of the system. It provides a complete set of components, covering service registration and discovery, service consumption, maintenance, disaster recovery, API gateway, distributed tracking and monitoring, distributed configuration management, and other key aspects.

The development of communication layer systems and devices is rooted in the Java programming language, which is widely used in web server construction and big data processing because of its reliability, security, cross-platform compatibility, and superior performance. In industry, the construction of hardware operating systems and external communication libraries often relies on languages such as C++ and Python. Similarly, industrial simulation devices utilize virtual device technology on diverse simulation software platforms, relying on their core systems to provide communication protocols and library support. To achieve robust communication and efficient data processing, the debugging system has to cross the Java boundary and adopt hybrid programming and integrated programming strategies, which invariably exacerbates the complexity of the communication challenges among different programming languages. To address this challenge, the Hybrid Programming course will focus on exploring new ways to skillfully blend the strengths of each language through advanced integration techniques. Specifically, the extension session aims to integrate functional modules from languages other than Java, and realize seamless collaboration between modules through careful design and development; while the integration process involves configuring the parsers of other languages to execute in the Java Runtime Environment (JRE), thus expanding the capabilities of the system and achieving a leap in functionality.

In the construction of the communication layer system of IoT, it is found that the bands of centralized interactive data transmission are mainly concentrated in the following four bands "57~61", "73~76", "81~83", and "08~12"., "08~12", therefore, the evaluation of the four bands of the communication

layer system, column thickness indicates the level of expectations, the ability of interactive data transmission for the curve, which is concentrated in the range of 2.5T~13.5T, as shown in Fig. 6.

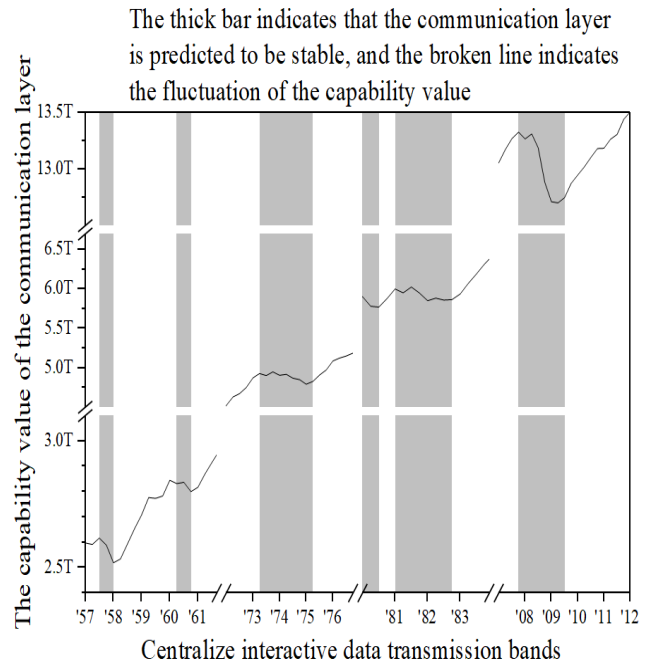


Fig. 6. Communication layer assessment for centralized interactive data transmission bands.

In the IoT communication architecture, the system integrates the Java Native Access (JNA) library, which is seen as an optimization and enhancement of the Java Native Interface (JNI) mechanism. When it comes to Java's Dynamic Link Library (DLL) calls to C++, JNA provides a more straightforward approach: it is assumed that you have already done the corresponding adaptation work, which constitutes one of the application scenarios. It is worth noting that the calling mechanism of the generic DLL follows the data structure specification defined by Sun, rather than directly adopting the data structure of the C language, to achieve access to functions within the established DLL. Ultimately, the process consists of downloading the shared Java library files and embedding them in the linked library system as function proxy service components.

B. Platform Construction and Testing for IoT

The database used to retrieve the data for this study is MySQL, which as a relational database management system has significant advantages in terms of its openness (open source), superior computational performance, and broad support for multiple platforms and applications [35]. In contrast, MyBatis is known for its lightweight, as an open-source database interaction manager, it is embedded in the essence of traditional JDBC (Java Database Connectivity) technology. Through the integration of global configuration data and mapping files, MyBatis can flexibly map the database table structure to the system-level classes and property blocks, this process cleverly circumvents the database driver registration, connection setup, and centralized SQL management of cumbersome

configuration. Integrating MyBatis into the Spring Boot framework not only simplifies dependency management but also makes it more straightforward and efficient to write query statements and manage database tables at the system data access object (DAO) level. This integration strategy promotes development efficiency and enhances application maintainability and scalability.

In the selection of MySQL to be centralized interactive data transmission band database validation, the band stability structure determination found that the purple for MyBatis, green for MySQL, and orange for the traditional other vb databases, found that the green MySQL peak band is similar to the interactive data that is, in the "57 ~ 61", "73~76", "81~83" all have better stability, most in line with the idea of this study, therefore, MySQL was selected as the tuning database, Fig. 7 horizontal axis is the band, the vertical axis is the stability. Specifically shown in Fig. 7.

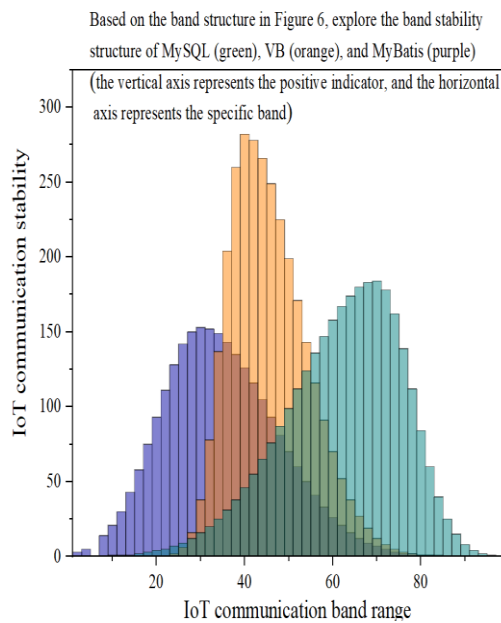


Fig. 7. Database selection test.

In the construction of the database, the IoT model is used to construct the data fields, the specific types of fields required such as demand, time, order details, storage information, etc., as shown in Table III.

TABLE III. FIELD TYPES REQUIRED FOR ORDERS

Field Name	Field type	Can it be left blank	explain
Bill_id	Int	NO	Order number, self increment primary key
Cargo_name	Int	NO	Name of goods
Demand_num	Int	NO	Quantity demanded
Demand_time	Int	NO	Requirement time
Cargo_id	varchar	NO	Customer ID
Creat_time	data	NO	Table record creation time
Description	varchar	NO	Order Description

The virtual commissioning system not only builds a bridge for communication and connecting information but also successfully meets the challenges of integrated management and unified storage of heterogeneous data from industrial equipment. In the face of complex industrial environments containing massive parallel, heterogeneous, and multi-source data, OPC UA technology is regarded as a powerful assistant for industrial IoT data management, helping the debugging work of virtual modeling. In this framework, Kepware, the preferred OPC UA server, is widely used as an industrial communication server software for controlling industrial automation equipment and associated industrial control programs. The process of initializing the server involves configuring communication addresses and ports, deploying channels on the KEPServer EX platform for the virtual system device hierarchy, adding analog devices, designing tag groups and tags based on the database table structure, and creating the corresponding nodes in the address space of the OPC UA server. Subsequently, the server control configuration table is optimized, and the OPC UA client service is constructed based on Spring Cloud microservice architecture, focusing on communication and hardware management functions.

In addition, the client can use the push mechanism to track the data changes from the server node, once the data is updated, the server will quickly analyze the host data and instantly send the host ID and attribute data back to the client, which effectively avoids the resource consumption of the client due to frequent polling of the server. Therefore, the three major systems in cloud services, i.e., IoT, Cloud Computing, and Edge Computing are evaluated for Data Return Resource Consumption where the results of the W1, W2, W4, W6, and W8 phases are shown in the following figure, and it is found that the average Data Return Resource Consumption rate of IoT is lower than that of Cloud Computing and Edge Computing, which is shown in Fig. 8.

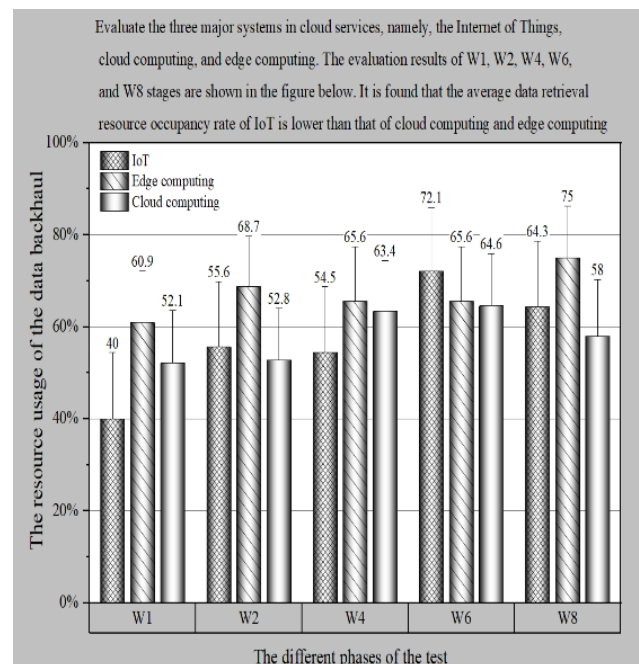


Fig. 8. Evaluation of data backhaul resource usage for IoT, cloud computing, and edge computing.

In the constructed IoT database, the logistic capacity data is retrieved from SQL, and the information table used, which restricts the data fields, only uses two types of data, int, and Varchar, as shown in Table IV.

TABLE IV. SQL CAPACITY INFORMATION RETRIEVAL TABLE

Field Name	Field type	Can it be left blank	explain
ID	Int	NO	Self-increment primary key
Warehouse_ID	Int	NO	Warehouse number
Vehicle status	Int	NO	Vehicle status
Vehicle id	Int	NO	Vehicle number
Vehicle position	Varchar	NO	Vehicle position
Description	Varchar	YES	Order details

To accurately replicate real-life IoT scenarios, the dimensions and construction of the simulation model need to follow the exact scale of the actual device. The process starts with the precise dimensioning and construction of the model according to the device manufacturer's detailed specifications, followed by an in-depth analysis of the entire simulation framework. The component modules are carefully constructed and finally brought together into a unified model system through a series of integration steps using uplink technology. To facilitate the seamless integration of the model across different software, the model templates are exported to STL format files, which greatly facilitates their importation into 3ds Max, thus allowing the user to adjust the mapping and optimize the surface details. To further enhance the realism of the simulation, the powerful rendering engine of 3ds Max is fully utilized, and its on-screen rendering function effectively enhances the realistic texture of the model. In terms of material and color selection, users can carefully choose from a rich library of materials and color samples to ensure that the surfaces of the model's components accurately reflect the texture and hue of the actual materials. In addition, to bring the design closer to real-world application scenarios, users can also subtly incorporate labeling elements, such as unique texture pattern files and brand logos, which undoubtedly add a sense of realism and professionalism to the overall design.

The logistics of IoT are finally warehoused, and the goods in the warehouse are still accessed using SQL's database, and the specific access and querying are done using three numeric types: int, varchar, and date, as shown in Table V.

TABLE V. SQL INBOUND ORDER FIELD QUERIES

Field Name	Field type	Can it be left blank	explain
ID	Int	NO	Self increment primary key
Creator_id	Int	NO	Creator ID
Bill_type	Int	NO	Order type
Start_time	Date	NO	Process start time
Real_time	Date	NO	Actual process time
State	Varchar	NO	State
Description	Varchar	YES	Order details

C. AHP and Entropy Weight Method for Logistics Supply Chain Evaluation

AHP and entropy weight method of logistics supply chain evaluation system is divided into four steps, firstly, to determine the indicator system of the evaluation object, secondly, to determine the AHP weights, again to determine the weights of entropy weight method, and finally to utilize the comprehensive weights for comprehensive analysis. The calculation of the evaluation object index system determination and objective assignment method-entropy weight method is shown in Table VI.

TABLE VI. EVALUATION OF INTERNET OF THINGS INDICATOR SYSTEM AND ENTROPY WEIGHT METHOD

Name	Weight	Difference Coe	Information entropy
Mobile electronic equipment	0.1198	0.043	0.953
Sensor market size	0.1231	0.001	0.864
Internet Popular rate	0.2981	0.089	0.943
Fixed broadband terminal	0.0471	0.012	0.896
IPv6 size	0.1871	0.074	0.798
R and D	0.2001	0.051	0.453
Technological personnel	0.1143	0.011	0.976
IoT personnel	0.0178	0.053	0.768
Number of patents	0.1841	0.095	0.989
GDP	0.0231	0.043	0.742

In the evaluation, the normal distribution function of the efficiency loss of the IoT is to be considered, and there are three peaks of the specific normal distribution, which are shown centrally in this paper, it turns out that the loss point is between [-1, 1] and [500, 600], and therefore, does not affect the results of this paper, and the specific normal function, as shown in Fig. 9.

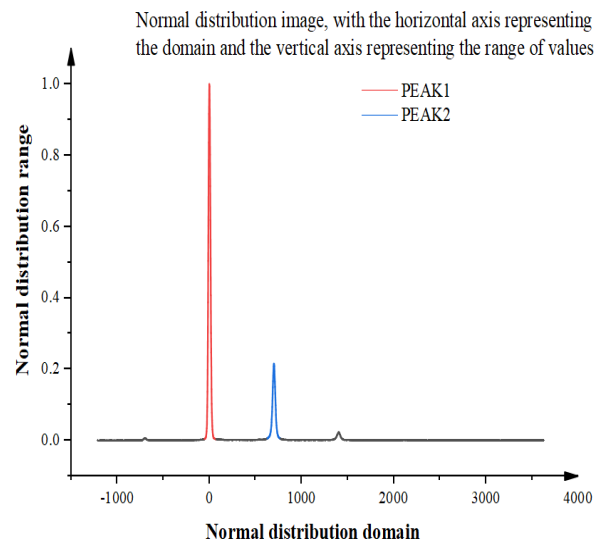


Fig. 9. Normal distribution of IoT efficiency loss.

After using the objective assignment method entropy weight method to determine the weights, but also to use the AHP method of IoT evaluation of subjective weight determination, the difference between it and entropy weight method is that one belongs to the objective assignment method, one belongs to the

subjective assignment method, the combination of the two to eliminate the defects of subjective and objective assignment of IoT, using the advantages of both, the following B1 ~ B10 represent 1 ~ 10 in Table VII respectively. Specific methods are as follows:

TABLE VII. ASSIGNMENT OF AHP FOR IOT EVALUATION

B\A	A1 0.322	A2 0.3331	A3 0.4331	Weight
B1	0.0423			0.0294
B2	0.3513			0.1321
B3	0.4091			0.0431
B4	0.0121			0.0213
B5	0.0311			0.0741
B6		0.1812		0.2913
B7		0.4121		0.0123
B8		0.3941		0.0478
B9			0.6412	0.1239
B10			0.3586	0.2312

This evaluation study is divided into three parts: comprehensive weights (i.e., a subjective and objective combination of weights), consistency test results, and IoT development index. The consistency test is a test of the AHP method, placed here the more intuitive expression of the accuracy of the empirical results, the IoT development index also shows an upward trend, as shown in Table VIII.

TABLE VIII. COMPOSITE WEIGHTS, CONSISTENCY TEST, AND IOT INDUSTRY DEVELOPMENT INDEX FOR IOT

Secondary indicators	Weight	Consistency	Level
B1	0.071	0.0784	0.6131
B2	0.124	0.0423	0.741
B3	0.232	0.0741	0.764
B4	0.041	0.0913	0.898
B5	0.021	0.0871	0.912
B6	0.251	0.0214	1.009
B7	0.031	0.0871	1.031
B8	0.071	0.0912	1.423
B9	0.012	0.0172	1.632
B10	0.129	0.0842	1.762

V. CONCLUSION AND FUTURE WORKS

This study deeply explores the core role and significant effect of IoT technology in the coordinated response mechanism of logistics supply chain management, which provides strong theoretical support and practical guidance for the intelligent transformation of the logistics industry. Through systematic analysis and practical verification, we have clarified that IoT, as a representative of the new generation of information technology, lays a solid foundation for the transparency, intelligence, and efficiency of the logistics supply chain with its

powerful data sensing, transmission, and processing capabilities. The coordination and response mechanism of the logistics supply chain has realized a qualitative leap. IoT technology not only realizes real-time and accurate information sharing among nodes of the supply chain and eliminates the barrier of information asymmetry, but also improves the sensitivity and response speed of the supply chain to market changes through intelligent analysis and prediction. This data-based decision-making support makes the supply chain more accurate and efficient in resource allocation, inventory management, logistics scheduling, etc., effectively reducing operating costs and risks. More importantly, the application of IoT technology promotes the in-depth integration and synergy of all links in the supply chain, forming a closer and more flexible supply chain network. In the face of unexpected events or market demand fluctuations, the IoT-driven coordination and response mechanism can quickly adjust strategies and optimize resource allocation to ensure the stability and resilience of the supply chain. This ability is of great significance for enhancing the overall competitiveness of the logistics industry and coping with the complex and volatile market environment.

The study has several limitations that warrant further exploration. Firstly, the proposed IoT-driven logistics supply chain coordination mechanism may lack generalizability across industries with diverse operational needs. Secondly, challenges such as network latency, data inconsistency, and device interoperability affecting real-time data accuracy remain inadequately addressed. Thirdly, the scalability of the framework for larger, more complex supply chains is not thoroughly evaluated. Lastly, the economic feasibility of implementing the proposed technologies, especially for small- and medium-sized enterprises, is insufficiently analyzed. Future research could focus on adapting the framework to different industries, integrating advanced technologies like blockchain, AI, and edge computing to enhance system reliability, exploring sustainable IoT practices to reduce environmental impact, and

conducting detailed cost-benefit analyses to assess the economic viability of such systems.

ACKNOWLEDGMENT

This research is supported by Social Science Research Planning Program of Jilin Provincial Education Department “Study on the Path of Integrated Development of Logistics Industry and Agriculture in Jilin Province under the Rural Revitalization Strategy” Grant No.: JJKH20230090SK.

REFERENCES

- [1] R. Huo et al., “A comprehensive survey on blockchain in industrial internet of things: Motivations, research progresses, and future challenges,” *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 88–122, 2022, doi: 10.1109/COMST.2022.3141490.
- [2] S. Yadav, S. Luthra, and D. Garg, “Internet of things (IoT) based coordination system in Agri-food supply chain: development of an efficient framework using DEMATEL-ISM,” *Operations management research*, vol. 15, no. 1, pp. 1–27, 2022, doi: 10.1007/s12063-020-00164-x.
- [3] K. Sallam, M. Mohamed, and A. W. Mohamed, “Internet of Things (IoT) in supply chain management: challenges, opportunities, and best practices,” *Sustainable Machine Intelligence Journal*, vol. 2, pp. 3–1, 2023, doi: 10.61185/SMIJ.2023.22103.
- [4] M. Ben-Daya, E. Hassini, and Z. Bahroun, “A conceptual framework for understanding the impact of the Internet of things on supply chain management,” *Operations and Supply Chain Management: An International Journal*, vol. 15, no. 2, pp. 251–268, 2022, doi: 10.31387/oscm0490345.
- [5] K. L. Keung, C. K. Lee, and P. Ji, “Industrial Internet of things-driven storage location assignment and order picking in a resource synchronization and sharing-based robotic mobile fulfillment system,” *Advanced Engineering Informatics*, vol. 52, p. 101540, 2022, doi: 10.1016/j.aei.2022.101540.
- [6] X. Chen, C. He, Y. Chen, and Z. Xie, “Internet of Things (IoT)—blockchain-enabled pharmaceutical supply chain resilience in the post-pandemic era,” *Frontiers of Engineering Management*, vol. 10, no. 1, pp. 82–95, 2023, doi: 10.1007/s42524-022-0233-1.
- [7] R. Mishra, R. K. Singh, T. U. Daim, S. F. Wamba, and M. Song, “Integrated usage of artificial intelligence, blockchain and the internet of things in logistics for decarbonization through paradox lens,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 189, p. 103684, 2024, doi: 10.1016/j.tre.2024.103684.
- [8] L. Liu, W. Song, and Y. Liu, “Leveraging digital capabilities toward a circular economy: Reinforcing sustainable supply chain management with Industry 4.0 technologies,” *Computers & Industrial Engineering*, vol. 178, p. 109113, 2023, doi: 10.1016/j.cie.2023.109113.
- [9] P. Kumar and S. Aziz, “Managing Supply Chain Risk with the Integration of Internet of Things in the Manufacturing Sector of Pakistan,” *Dutch Journal of Finance and Management*, vol. 5, no. 2, p. 22405, 2023, doi: 10.55267/djfm/13676.
- [10] Y. Mashayekhy, A. Babaei, X.-M. Yuan, and A. Xue, “Impact of Internet of Things (IoT) on inventory management: A literature survey,” *Logistics*, vol. 6, no. 2, p. 33, 2022, doi: 10.3390/logistics6020033.
- [11] M. Song, X. Ma, X. Zhao, and L. Zhang, “How to enhance supply chain resilience: a logistics approach,” *The International Journal of Logistics Management*, vol. 33, no. 4, pp. 1408–1436, 2022, doi: 10.1108/IJLM-04-2021-0211.
- [12] M. Rajabzadeh and H. Fatorachian, “Modelling factors influencing IoT adoption: With a focus on agricultural logistics operations,” *Smart Cities*, vol. 6, no. 6, pp. 3266–3296, 2023, doi: 10.3390/smartcities6060145.
- [13] M. M. Billah, S. S. Alam, M. Masukujjaman, M. H. Ali, Z. K. M. Makhbul, and M. F. M. Salleh, “Effects of Internet of Things, supply chain collaboration and ethical sensitivity on sustainable performance: moderating effect of supply chain dynamism,” *Journal of Enterprise Information Management*, vol. 36, no. 5, pp. 1270–1295, 2023, doi: 10.1108/JEIM-06-2022-0213.
- [14] W. Liu, S. Wei, S. Wang, M. K. Lim, and Y. Wang, “Problem identification model of agricultural precision management based on smart supply chains: An exploratory study from China,” *Journal of Cleaner Production*, vol. 352, p. 131622, 2022, doi: 10.1016/j.jclepro.2022.131622.
- [15] W. C. Tan and M. S. Sidhu, “Review of RFID and IoT integration in supply chain management,” *Operations Research Perspectives*, vol. 9, p. 100229, 2022, doi: 10.1016/j.orp.2022.100229.
- [16] D. Kumar, R. K. Singh, R. Mishra, and T. U. Daim, “Roadmap for integrating blockchain with Internet of Things (IoT) for sustainable and secured operations in logistics and supply chains: Decision-making framework with case illustration,” *Technological Forecasting and Social Change*, vol. 196, p. 122837, 2023, doi: 10.1016/j.techfore.2023.122837.
- [17] R. Kumar, S. Rani, and M. A. Awadh, “Exploring the application sphere of the Internet of things in industry 4.0: a review, bibliometric and content analysis,” *Sensors*, vol. 22, no. 11, p. 4276, 2022, doi: 10.3390/s22114276.
- [18] Y. Liu, C. Yang, K. Huang, W. Gui, and S. Hu, “A systematic procurement supply chain optimization technique based on industrial Internet of things and application,” *IEEE Internet of Things Journal*, vol. 10, no. 8, pp. 7272–7292, 2022, doi: 10.1109/JIOT.2022.3228736.
- [19] F. Ye, K. Liu, L. Li, K.-H. Lai, Y. Zhan, and A. Kumar, “Digital supply chain management in the COVID-19 crisis: An asset orchestration perspective,” *International Journal of Production Economics*, vol. 245, p. 108396, 2022, doi: 10.1016/j.ijpe.2021.108396.
- [20] P. Kumar and R. K. Singh, “Application of Industry 4.0 technologies for effective coordination in humanitarian supply chains: a strategic approach,” *Annals of Operations Research*, vol. 319, no. 1, pp. 379–411, 2022, doi: 10.1007/s10479-020-03898-w.
- [21] I. Vlachos, R. M. Pascuzzi, M. Ntosis, K. Spanaki, S. Despoudi, and P. Repoussis, “Smart and flexible manufacturing systems using autonomous guided vehicles (AGVs) and the Internet of things (IoT),” *International Journal of Production Research*, vol. 62, no. 15, pp. 5574–5595, 2024, doi: 10.1080/00207543.2022.2136282.
- [22] D. Kumar, R. K. Singh, R. Mishra, and S. F. Wamba, “Applications of the Internet of things for optimizing warehousing and logistics operations: A systematic literature review and future research directions,” *Computers & Industrial Engineering*, vol. 171, p. 108455, 2022, doi: 10.1016/j.cie.2022.108455.
- [23] G. Zhang, Y. Yang, and G. Yang, “Smart supply chain management in Industry 4.0: the review, research agenda and strategies in North America,” *Annals of Operations Research*, vol. 322, no. 2, pp. 1075–1117, 2023, doi: 10.1007/s10479-022-04689-1.
- [24] W. Liu, S. Long, and S. Wei, “Correlation mechanism between smart technology and smart supply chain innovation performance: A multi-case study from China’s companies with Physical Internet,” *International Journal of Production Economics*, vol. 245, p. 108394, 2022, doi: 10.1016/j.ijpe.2021.108394.
- [25] X. Mu and M. F. Antwi-Afari, “The applications of Internet of Things (IoT) in industrial management: a science mapping review,” *International Journal of Production Research*, vol. 62, no. 5, pp. 1928–1952, 2024, doi: 10.1080/00207543.2023.2290229.
- [26] P. Kumar, D. Sharma, and P. Pandey, “Coordination mechanisms for digital and sustainable textile supply chain,” *International Journal of Productivity and Performance Management*, vol. 72, no. 6, pp. 1533–1559, 2023, doi: 10.1108/IJPPM-11-2020-0615.
- [27] S. Al-Ayed and A. Al-Tit, “The effect of supply chain risk management on supply chain resilience: The intervening part of Internet-of-Things,” *Uncertain Supply Chain Management*, vol. 11, no. 1, pp. 179–186, 2023, doi: 10.5267/j.uscm.2022.10.009.
- [28] Z. Dong, W. Liang, Y. Liang, W. Gao, and Y. Lu, “Blockchained supply chain management based on IoT tracking and machine learning,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2022, no. 1, p. 127, 2022, doi: 10.1186/s13638-022-02209-0.
- [29] L. Li, Y. Gong, Z. Wang, and S. Liu, “Big data and big disaster: a mechanism of supply chain risk management in the global logistics industry,” *International Journal of Operations & Production Management*, vol. 43, no. 2, pp. 274–307, 2023, doi: 10.1108/IJOPM-04-2022-0266.

- [30] A. Egwuonwu, C. Mordi, A. Egwuonwu, and O. Uadiale, "The influence of blockchains and internet of things on the global value chain," *Strategic Change*, vol. 31, no. 1, pp. 45–55, 2022, doi: 10.1002/jsc.2484.
- [31] S. Yadav, T.-M. Choi, S. Luthra, A. Kumar, and D. Garg, "Using Internet of Things (IoT) in agri-food supply chains: A research framework for social good with network clustering analysis," *IEEE Transactions on Engineering Management*, vol. 70, no. 3, pp. 1215–1224, 2022, doi: 10.1109/TEM.2022.3177188.
- [32] A. Rejeb et al., "Unleashing the power of the Internet of things and blockchain: A comprehensive analysis and future directions," *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 1–18, 2024, doi: 10.1016/j.iotcps.2023.06.003.
- [33] S. Khan, R. Singh, S. Khan, and A. H. Ngah, "Unearthing the barriers of Internet of Things adoption in the food supply chain: A developing country perspective," *Green Technologies and Sustainability*, vol. 1, no. 2, p. 100023, 2023, doi: 10.1016/j.grets.2023.100023.
- [34] R. M. L. Rebelo, S. C. F. Pereira, and M. M. Queiroz, "The interplay between the Internet of things and supply chain management: Challenges and opportunities based on a systematic literature review," *Benchmarking: An International Journal*, vol. 29, no. 2, pp. 683–711, 2022, doi: 10.1108/BIJ-02-2021-0085.
- [35] H. Tran-Dang, N. Krommenacker, P. Charpentier, and D.-S. Kim, "The Internet of Things for Logistics: Perspectives, application review, and challenges," *IETE Technical Review*, vol. 39, no. 1, pp. 93–121, 2022, doi: 10.1080/02564602.2020.1827308.

Big Data Analytics of Knowledge and Skill Sets for Web Development Using Latent Dirichlet Allocation and Clustering Analysis

Karina Djunaidi¹, Dine Tiara Kusuma^{2*}, Rahma Farah Ningrum³, Puji Catur Siswipraptini⁴, Dina Fitria Murad⁵
Faculty of Energy Telematics, Institut Teknologi PLN, Jakarta, Indonesia^{1, 2, 3, 4}
Information Systems Department-Binus Online Learning, Bina Nusantara University, Jakarta, Indonesia⁵

Abstract—Web development is a data-centric field and fundamental component of data science. The advent of big data analytics has significantly transformed the processes, knowledge domains, and competencies associated with Web development. Accordingly, educational programs must adjust to contemporary advancements by initially determining the abilities required for big data web developers to satisfy industry demands and adhere to current trends. This study aims to identify the knowledge areas and abilities essential for big data analytics and to create a taxonomy by correlating these competences with currently popular tools in web development. A mixed method consisting of semi-automatic and clustering methods is proposed for the semantic analysis of the text content of online job advertisements associated with the development of big data web applications. This methodology uses Latent Dirichlet Allocation (LDA), a probabilistic topic modeling tool, to uncover hidden semantic structures within a precisely specified textual corpus and average linkage hierarchical clustering as a clustering analysis technique for web developers. The results of this study are a web development competency map which is expected to help evaluate and improve the knowledge, qualifications and skills of IT professionals being hired. It helps to identify the roles and competencies of professionals in the company's personnel recruitment process; and meet industry skill requirements through web development education programs. The competency map consists of knowledge domains, skills and essential tools for web development such as basic knowledge, frameworks, design and user experience, database design, web development, cloud computing and other soft skills. Furthermore, the proposed model can be extended to several types of jobs in the IT sector.

Keywords—Big data analytics; hierarchical clustering; Latent Dirichlet Allocation; web development; knowledge; skill

I. INTRODUCTION

The revolution of Industry 4.0 carried out the concept of digitalization in all sectors producing big data. Big data consists of enormous volumes and a variety of data. It is impossible to manage and process the traditional management methods [1]. Big data analysis reveals hidden information, patterns, and correlations with new insights [2]. The valuable insights and implications derived from big data analytics are used in intelligent processes, such as guiding decision-making strategies in various organizations, including educational institutions, businesses, and governments. Big data are generated from many resources, such as websites, applications, emails, social media, and other multimedia platforms [3], [4], [5].

Websites as famous digital marketing in any organization, have caused increasing demand for data-oriented services, aligning knowledge and skill sets related to big data [2], [6]. Big data analytics is defined as the process of examining, processing, and analyzing large and complex data sets that cannot be handled by traditional methods. The goal of these analytics is to identify patterns, trends, and useful insights from structured and unstructured data [7] [8].

Recently, the technology lifecycle has shown a significant increase in big data-driven websites. Some modern products and services have been embedded in big data-oriented websites; thus, they have become an interesting topic for researchers[9], [10]. Digital transformation across business and industrial eras provides an exciting experience for software and service-based economics, for which modern websites can provide valuable information from big datasets [11], [12]. During this process, websites played a significant role in modernizing numerous sectors [13], [14], [15]. Web developers played an important role in developing ICT-based industries. Web developer is one of the information technologies (IT) occupations projected to grow by almost ten percent by 2033, it is much faster than other IT occupations [16]. The main task of web developers is to design and build a responsive website using popular programming languages such as HTML, cascading style sheet (CSS), and JavaScript. They are also responsible for testing, debugging, and integrating systems using application programming interface (API) services.

Big data causes new and challenging problems that need to be resolved using artificial intelligence. The Indonesian Ministry of Education, Culture, Research and Technology (Kemendikbudristek) stated that, only 15-20% of bachelor graduates have competencies match to their jobs. Skills and job profiles in the (Information Technology) IT sector is not clearly defined [17]. Therefore, some literature has discussed the big data of web developers and has become a hot topic among scientists in the last five years [18], [19]. Big data consists of approximately 5Vs; volume means a huge amount of data, variety means a type of data such as structured and unstructured data, velocity means high speed and real-time, veracity means reliable and accurate, and variability means volatility [1], [9]. These five characteristics of big data form the basis of the web development life cycle through methodologies and approaches in Big Data Analytics (BDA). This study aims to reveal hidden information and the value of implementing BDA in web

*Corresponding Author.

developers' occupations by identifying its knowledge domains and skill sets.

Given this context, BDA requires a diverse set of skills, programming languages, web development tools, and frameworks. The web development industry is a dynamic work environment that relies entirely on the resources of qualified people. The competence of BDA specialists strongly influences the quality of BDA-based products and services. BDA has grown in popularity, as has the requirement for qualification. A semi-automatic methodology was proposed to analyze collections of online job advertisements (ads). Our methodology is based on semantic analysis – hierarchical clustering (SA-HC) of BDA job ads using Latent Dirichlet Allocation (LDA) and average linkage hierarchical clustering. Latent Dirichlet Allocation (LDA). LDA is a generative statistical model used in a wide range of research in natural language processing and data analysis, such as topic modeling, sentiment analysis, and text analysis. This study revealed the core skills and knowledge required for BDA based on discovery topics, using LDA and hierarchical clustering analysis. The topics are mapped based on competency domains to reveal a structured taxonomy for BDA. Furthermore, the technologies required for BDA, such as programming languages, databases, and big data tools, are extracted.

The main contributions of this research are:

- A competency taxonomy for BDA developed by mapping the topics according to competency domain.
- The complex datasets provide a wide range of web developers topic area.
- A novel mixed method consists of latent Dirichlet allocation (LDA) and average linkage hierarchical clustering.
- BDA contributes significantly to decision-making processes because it has high granularity detailed information related to web developers' knowledge and skill sets.
- This research has been conducted by involving expert judgement in determining web development analysis.

II. RELATED WORKS

The methodology of this study provides a comprehensive explanation of big data analytics and semantics associated with them. It is also based on a content analysis of the textual content of BDA job ads using generative topic models to reveal the knowledge domains and skill sets required for BDA. Therefore, the background of the study is addressed under two subheadings: big data web developers and topic models / big data analytics.

A. Big Data Analytics

Big data analytics is the process of analyzing large volumes of documents to extract meaningful insights and values. High technology industries pioneered the method of deriving values from BDA. It includes a variety of data-intensive technologies that are capable of processing large volumes of data [1], [20], [21]. High-level management uses big data to impact grateful decision making, which is one of the parameters useful for BDA.

This makes a substantial contribution to decision making because large-scale data contains specific information. The BDA consisted of five stages:

- Data retrieval refers to a set of text, images, videos from the Internet, sensors, and e-commerce; for example, social media generates billions of related data every day.
- Data acquisition consists of collecting data from sources, preprocessing to clean datasets, and transforming the data for purposes such as classification or clustering.
- Data management file system was created for effective data storage and processing of large datasets. The industry deploys big data cloud models and systems, such as the Hadoop distributed file system (HDFS) and NoSQL.
- Data analytics is the process of extracting insights from massive datasets using artificial intelligence techniques such as machine learning and data mining. BDA reveals knowledge for decision making by identifying hidden patterns, links, and interconnections. User experiences such as customer service and decision support can be enhanced by BDA.
- Data visualization is a graphical representation commonly used in big data. Researchers can use general software such as R studio or MATLAB to create some visualizations. Other than that, industries usually use their own applications, such as GIS-based 3D visualization, to monitor traffic data.

B. Web Developers

Web developers is one of information technology (IT) occupation which projected to grow almost ten percent up to 2033, it is much faster than other IT occupations[16]. The main task of web developers is to design and build a responsive website using popular programming languages, such as HTML, CSS, and JavaScript. They are also responsible for testing, debugging, and integrating systems using API services. Reliable and fast interconnection between web and mobile development has become a trending issue. Web developers can access WSs through application programming interfaces (APIs) in social media platforms [18]. Furthermore, web development tools have been implemented to enhance the accessibility of artificial intelligence for researchers and end-users [19].

C. Latent Dirichlet Allocation (LDA)

LDA, a generative statistical model, has become a popular method for topic modeling in text mining [2]. Latent pertains to the identification of semantic content in corpus documents through the analysis of the underlying semantic structures. The generative approach in LDA ensures the allocation of terms in a document to random variables, followed by semantic clustering through a repeating probabilistic process grounded in Dirichlet distribution. LDA is an unsupervised learning methodology that does not require labeling or training datasets. It can be concluded that LDA can be efficiently applied to large corpus documents to identify semantic patterns. In the past five years, LDA has gained popularity in text mining studies spanning a variety of contexts, including e-commerce reviews, natural language processing, information extraction, sentiment analysis, and

social media trend analytics. Likewise, this approach has been used as a successful strategy in some studies that analyzed online job advertisements from businesses and industries. In the past, topic models were only created for textual data analysis but are currently being applied to a variety of data sources, including genetic data, photos, videos, and social networks. For these reasons, this study implemented LDA as a topic modeling method.

D. Average Linkage Clustering Analysis

Unsupervised learning such as hierarchical clustering, has become a popular method in data analysis. A superior cluster quality was provided by hierarchical clustering, thereby diminishing the sensitivity of clustering to various problem types. It comprises two methodologies: bottom-up and top-down. Agglomerative as a bottom-up approach initially treats each instance as an individual cluster, which is subsequently merged to form bigger clusters; this is known as Average-Linkage Hierarchical Clustering (ALHC) [22]. This process continues until all the clusters are combined into a single giant cluster containing all the instances. The hierarchical clustering method identifies common traits and job profiles in IT job posts, including competency, programming languages, web development tools, and frameworks [23].

III. RESEARCH METHOD

This study analyzed the content of web developer’s job advertisements. The proposed research methodology is described in Fig. 1 which consists of three main phases: data collection, text preprocessing, and LDA implementation. The following figure illustrates the overall process.

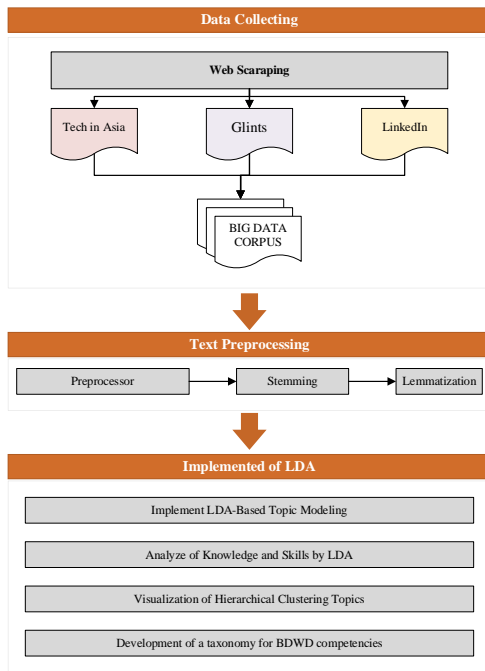


Fig. 1. Research methodology of big data analytics using LDA and clustering analysis.

A. Data Collection

The data used in this study were obtained from online job advertisements published by Glints [24], Tech in Asia [25], and LinkedIn [26]. A total of 2649 data were collected using the web scraping technique from January 2023 to July 2024. Job advertisements were searched using the keyword ‘web developer’, ‘web development’, and some expertise fields, including job title, job description, and required skills were collected. Table I presents the sample dataset.

TABLE I. SAMPLE OF DATA COLLECTION

Job Title	Job Description	Required Skills
Full stack Developer	Design, build, and optimize front-end and back-end code. Develop and maintain web applications. Integrate third-party APIs and services. Perform testing and debugging to ensure application quality. Collaborate with the design team to implement a responsive and intuitive user interface. Manage databases and perform query optimization. Provide continuous technical support and bug fixes. Keep up with the latest technological developments and implement best practices in development.	jQuery Debugging Laravel JavaScript CSS3 GIT SQL PHP HTML5
Web Developer	Develop new web application or customize existing application Learn new technology when required in the process of application development Problem solving and working with team on a project	Node.js REST API Laravel PHP

B. Text Preprocessing

Text preprocessing has become a crucial stage in information retrieval research. However, text data often comes in an unstructured form and is full of noise, especially when obtained from sources such as social media or websites. Therefore, text preprocessing is a crucial initial step in aligning data before being directed to further stages of analysis. The preprocessing stage plays a major role in removing text data from noise, which can damage the quality of the results [2], [27], [28], [29], [30]. Text preprocessing applied to the experimental data set consisted of several sequential stages. First, the text data were divided into words (tokens/parsing), known as tokenization, to obtain meaningful attributes [31], [32].

Tokenization divides text in the form of sentences or paragraphs into tokens/parts that are then represented by data vectors [33], [34], [35], [36]. Furthermore, web links, personal tags, and characters/affixes with no meaning were removed. The next stage was the stop word process. Stop words are used to reduce the number of words in a document, which affects the speed and performance of Natural Language Processing (NLP) [37], [38], [39].

The text preprocessing stages involved in converting textual data into keywords in WordStat ver.2024 are:

1) *Pre-processor*: This option allows custom text transformations to be analyzed before or instead of the execution of the other three standard processes: lemmatization, exclusion, and categorization. These transformations are achieved by executing specially designed external routines that

can be accessed in the form of Python scripts, external EXE files, or functions in dynamic link library.

2) *Stemming*: Stemming is a process used in text preprocessing to convert words into their base form [40], [41], [42]. For example, the word 'running' is changed to 'run. This helps in text analysis because it reduces the variation in words with similar meanings. Stemming ignores suffixes and prefixes to arrive at the base form of the word. Its use is common in applications such as information retrieval, sentiment analysis, and text mining. This can be useful for improving the accuracy of text analysis.

3) *Lemmatization*: Lemmatization is a process in text preprocessing that aims to change words into their basic form or 'lemma'. Unlike stemming, which only cuts off the endings of words, lemmatization considers context and changes words into their grammatically correct basic forms [43], [44]. After the preprocessing stage was completed, each text (job ad) in the dataset was defined as a word matrix. As a result of the preprocessing, the word space size for the entire dataset was reduced from 28232 to 22773. The dataset consisting of the job ads "TGL Web Developer and Digital Designer" is characterized by 22773 unique words, which also refers to the word matrix size for each ad. The number of matrices/vectors is 1868, which is also the number of job ads. The Document Term Matrix (DTM) created for this analysis consists of 1868 rows and 22773 columns. In other words, the DTM shows that 1868 job ads were represented by a word space consisting of 22773 terms. The DTM weighting process is performed by considering the word frequency.

C. Implementation of LDA-Based Topic Modeling

This step of the experimental analysis entails the application of a topic model to the dataset to reveal the domain knowledge and expertise necessary for BDA in a clear and comprehensible manner. The LDA model is a document generation model based on Bayesian theory, that excels in extracting themes and features from voluminous texts. This concept has found extensive application in fields such as text mining and information retrieval [45]. The inherent qualities of research that utilizes semantic analysis of job advertisements contribute to the effectiveness of LDA as a topic model [2]. LDA-based topic modeling assumes that the distribution of subjects in texts and the distribution of words within topics are mutually independent. Identical terms may manifest at varying degrees across distinct subjects. Likewise, specific subject matter may manifest to varying degrees in several written materials. The fundamental premise of the LDA model is derived from the Bayesian joint probabilistic model. The objective of this study is to use LDA-based topic modeling to reveal the underlying semantic structures (word clusters) in a textual corpus of job advertising.

Once the LDA model was implemented, the probability distribution for each topic is computed using Bayesian estimation methods in conjunction with the Dirichlet distribution. The WordStat ver.2024 tool was utilized to

implement an LDA model in this experimental investigation [46]. This tool is specifically designed to implement an LDA model. WordStat was implemented with varying iteration counts and was stabilized after 500 successive Gibbs sampling iterations.

The Bayesian inference model is the most important component of the LDA model [45], which is produced by the three-layer Bayesian probability of the "topic word" in the text. The diagram in Fig. 2 illustrates the topological structure of LDA. The fundamental algorithms of the data of a data-sampling algorithm and a feature-weight algorithm. The data samples in this study were collected using a web scraping technique.

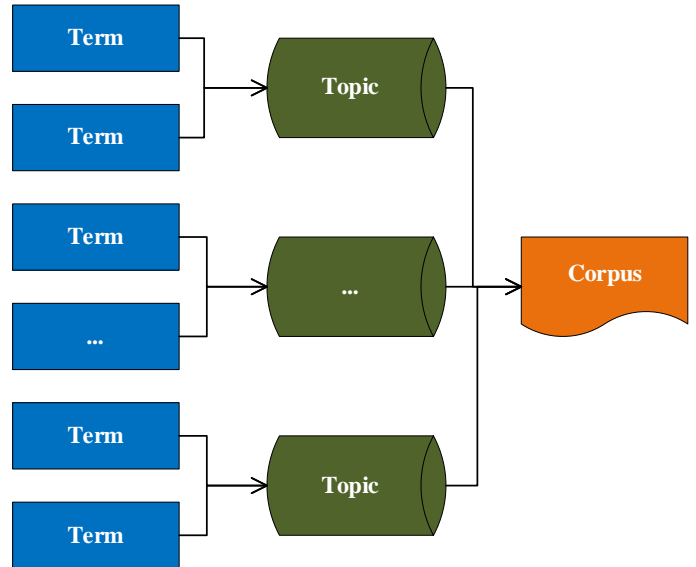


Fig. 2. Topological structure of LDA.

Within the LDA model, researchers explicitly allocated topic names to the identified themes, in accordance with the descriptive keywords. The naming of topics is based on the significance of all keywords and involves professional experts in the field of web developers through Forum Group Discussions (FGD). Therefore, the topic titles used may differ depending on the perspective of each researcher. LDA is an unsupervised generative probabilistic technique that is used to model a corpus.

Fig. 3 shows a diagram illustrating the LDA algorithm. The parameters used were as follows:

α and β are parameters of the previous distribution of θ . z is the designated theme for the n th word in the document count

ϑ is distribution of terms in number of topics theme

w is word in document

For the specified parameter θ , the mathematical equation to compute the probability distribution of the topic in Eq. (1) is as follows:

$$p(z|\theta) = \prod_n^n p(z|\theta) = \prod_{k=1}^K \theta_k^n \quad (1)$$

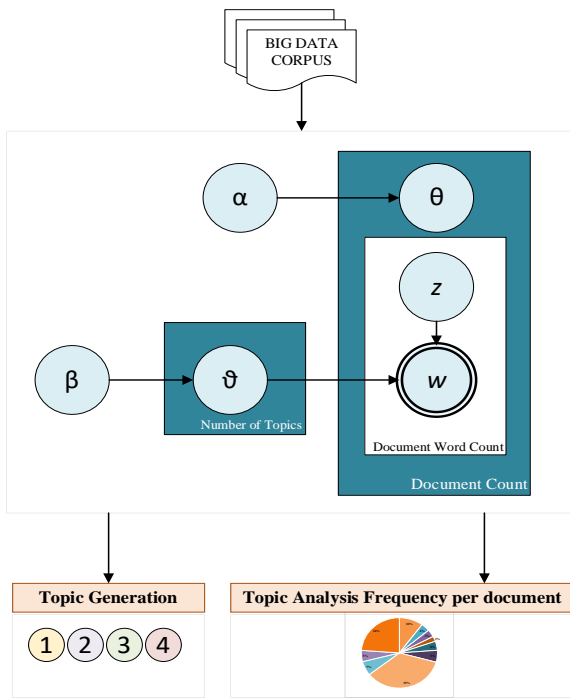


Fig. 3. Diagram illustrating the LDA algorithm.

Assuming the term-document matrix is defined and considering the number of documents comprising a corpus, these matrices often exhibit a significant size. Hence, it is customary in text mining to exclude sparse terms, which have a very low inclusion rate in documents. Typically, such an approach allows for a substantial reduction in the size of the matrix while preserving its essential relationships.

IV. RESULTS

BDA is presented to identify the core competencies of big data web development. First, the corpus document classifies the topics into skill sets. Then, the competency domain is mapped using these skill sets. Finally, most in-demand tools for BDA, programming languages, web development tools, and frameworks were analyzed to identify higher quality competencies. The results of the analysis are presented and discussed below.

A. Analyze Knowledge and Skills Using Latent Dirichlet Allocation

The corpus document formed by top three job advertisements comprised a wide spectrum of knowledge, skills, and job descriptions in the web development area. These spectra extended the coverage of the discovered topics of BDA. Three variables of job advertisements were used to combine the LDA-based topic modeling. Knowledge domains and skill sets of BDA were revealed and discovered from 26 trending topics with optimal granularity. As presented in Table II, topics were combined using descriptive LDA keywords and topic rates. The descending order and percentages are listed in Table II. Descending order means that the first term was the most occurrence and the last term was the least occurrence in a topic. The names of the discovered topics were automatically assigned using WordStat ver. 2024.

TABLE II. DISCOVERED TOPICS

TOPIC NAME	LATENT DIRICHLET ALLOCATION KEYWORDS	RATE %
Cascading style	cascading style sheets; responsive web design; css; front end; web development; client requirements;	10.04
Programming language	programming language; python; typescript; javascript; kotlin; golang	9.32
Graphic design	graphic design; corel draw; adobe photoshop; pattern; teamwork; management; communication; marketing social media	8.34
HTML CSS PHP	php; html; css; doctrine; laravel; mysql; jquery; bootstrap; wordpress; framework	7.04
Graphic design video	video; editing; graphic; design; illustrations; canva; adobe; video editing; photo editing; image editing; motion graphics; multimedia design	5.44
Digital marketing	digital marketing; marketing strategy; online marketing; product marketing; sales and marketing; creative writing; creative design; google ads; content marketing;	4.90
Code review	code review; code integration; development; system development	4.85
Javascript	javascript; react js; laravel node; angular js; vue js; tailwind css; node js; cassandra; development javascript;	4.76
Big data	big data; data engineering; scala; apache spark; data mining; olap cubes; data cubes	4.63
Design tools	After effects; adobe photoshop; adobe illustrator; canva design; google sketch up; auto cad; google sketch up; interior design;	4.35
Relocation provided	relocation provided; lead data engineer hadoop apps; staff data engineer; engineer data platform;	4.30
Performance tuning	performance tuning; testing; development; skills; continuous delivery; clean coding	4.06
Full stack	full stack developer; full stack engineer; full stack; full stack web developer; front end; back end;	3.55
Database design	database design; data architecture; sql; sql server; modeling; postgresql; nosql; stored procedures; database development; query optimization;	3.35
Business requirements	business requirements; problem solving; metrics driven; operational excellence; application testing; agile development; coding standards;	3.19
Object oriented	object; oriented; oop; object oriented; object oriented programming	2.66
Interpersonal skills	skills; analytical; interpersonal; solving; problem; communication; administration; systems interpersonal skills;	2.51
User experience	user experience; user research; architecture; large language models; requirements gathering; user interface design; debugging	2.32
Search engine	search engine; search engine optimization; digital marketing; google analytics; google ads; instagram	1.57
API	rest api; rest api laravel; git	1.50

TOPIC NAME	LATENT DIRICHLET ALLOCATION KEYWORDS	RATE %
Full stack developer	fullstack; programmer; developer; senior; engineer; backend; web programmer; angular developer; senior full stack developer;	1.46
Model view controller	model view controller; mvc; asp.net; sql; programming	1.44
Web service	amazon; service; web; aws; amazon web;	1.37
System UI	system; ui; administration; linux; ux; application; test; testing; mobile; integration	1.23
Online advertising	online advertising; paid advertising; google ads; instagram ads; digital marketing; market analysis	1.09
Information technology	information technology; service; communications; customer; security; management; product; service level agreements;	0.73

Table II shows that cascading style, programming languages, and graphic design were among the competencies with the highest demand in the BDA industry. Other knowledge and skills in the top ten were html css php, graphic design video, digital marketing, code review, JavaScript, big data, and design tools. The discovered topics also covered various emerging trends, such as database design, business requirements, object oriented, user experience, search engines and soft skill areas such as interpersonal skills which shed light on the priorities and demands in the ever-growing BDA industry.

B. Knowledge and Skill Mapping According to Competency Domains

This stage focuses on categorization and presents knowledge and skills in a structured manner. First, a mapping process was performed by associating knowledge and skills with the competency domains and workflows. Second, the knowledge and skills revealed by 26 topics were mapped into ten core competency maps developed for BDA.

Table III presents the distribution of knowledge and skills according to the competency map and their respective percentages. As presented in Table III, the first three competency areas are related to the function of web development, which consists of big data products, roles, and specialized web developers. The total rate of these competencies as the most important focus in web development area was 48.49%. The next five competency areas were related to the major discipline, comprising databases, web development frameworks, tasks, programming languages, and web development tools. The total rate for these competencies areas was 38.81%. The last two concern the interdisciplinary areas, consisting of educational background and soft skills, at a rate of 12.7%. These ten competency areas are discussed in detail below.

The first competency area, big data products in web developer area (8.96%). It contains five knowledge and skill items: digital marketing, social media, search engine optimization, data engineering, and data science. The second, role (33.92%) means a web developer must perform, such as web development, graphic design, software development, data engineering, etc. The third, specialized web developer, means specialization on web developer title (5.61%), contains some

items of front-end developers, full stack developers, Java developers, and others. The fourth, databases (2.1%) as a query language and data storage area, has the top five highest demands in BDA: SQL server, MySQL, MongoDB, Oracle, and powerBI. Fifth, web development frameworks (2.42%) used to create interactive and progressive user interfaces, consisted of six items: Apache spark, Laravel, angular js, etc. The sixth, task (10.52%), has the duties of web developer comprising user interface design, application testing, user experience, video editing, technical, content creation, and object-oriented. The seventh, programming languages (6.4%), has a lot of items, HTML CSS JavaScript, Python, typescript, scala, and VBScript. The eighth, web development tool (17.37%), contained a variety of frameworks, programming languages, and software. The ninth, educational background (4.41%), comprised four majors: software engineering, computer science, information technology, and electrical engineering. Finally, soft skills (8.29%) included problem solving, public speaking, creative design, positive attitude, communication skills, time management, analytical skills, and critical thinking.

TABLE III. COMPETENCY MAP

ID	Competency Areas	Knowledge and skills	Rate %	Total %
1	Big data product/output	Digital marketing Social media Search engine optimization Data engineering Data science	5.69 2.19 0.56 0.29 0.23	8.96
2	Role	Web development Graphic designer Software development Back end Full stack Front end End developer Application development Data engineer	6.79 6.67 5.56 4.31 2.87 2.77 2.29 1.52 1.13	33.92
3	Specialized web developer	Web developer Frontend developer Full stack developer Java developer Net developer Software engineer back end	2.45 0.96 1.13 0.39 0.38 0.31	5.61
4	Databases	SQL Server MySQL MongoDb Oracle Database Power BI	1.33 0.58 0.08 0.07 0.04	2.1
5	Web development framework	Model view controller Apache spark Php-laravel Javascript frameworks Angular js Java spring	0.37 0.19 0.89 0.37 0.15 0.45	2.42
6	Task	User interface design Application testing Video editing User experience Technical Cascading style sheets Data analytics Object oriented programming Content creation	2.68 1.25 1.18 1.12 1.07 0.94 0.86 0.72 0.68	10.52
7	Programming Languages	Html css javascript Asp net	3.98 0.99	6.4

ID	Competency Areas	Knowledge and skills	Rate %	Total %
		Typescript Development java Scala Visual basic Php rest api Python Query languages Vbscript Server side	0.36 0.32 0.18 0.18 0.16 0.08 0.07 0.05 0.03	
8	Web Development Tools	Adobe photoshop Programming language React js Rest api Amazon web Node js Spring framework Corel draw Web applications	9.83 1.48 1.36 0.88 0.85 0.84 0.80 0.70 0.64	17.37
9	Educational background	Software engineer Computer science Information technology Electrical engineering	2.05 2.08 0.23 0.04	4.41
10	Soft skills	Problem solving Public speaking Active listening Creative design Positive attitude Communication skills Time management Analytical skills Critical thinking	1.69 0.73 0.72 1.78 1.14 1.35 0.35 0.29 0.23	8.29

C. Identification of the High Demand Tools for BDA

Recently, collective environments of web development, a wide range of tools and technologies, such as programming languages, web development tools, and frameworks are used simulant. The corpus document was analyzed using a keyword indexing technique to reveal the tools and technologies required for BDA [29]. The findings of this analysis were divided into three main categories: programming languages, web development tools, and frameworks, which are discussed in detail in the following sections.

1) *Programming languages:* Programming languages are essential tools for application development and serve various applications. The job advertisement dataset was analyzed using keyword indexing to identify programming languages used in BDA. Table IV shows the top 12 programming languages required for BDA along with their percentages.

TABLE IV. PROGRAMMING LANGUAGES

Programming languages	Rate %
HTML CSS Javascript	59,2
Asp.Net	14,7
Typescript	5,31
Development Java	4,7
Scala	2,73
Visual Basic	2,73
PHP	2,43

Programming languages	Rate %
Golang	1,56
Python	1,21
Query Languages	1,06
VBscript	0,7
Server Side	0,46

According to the results in Table IV, HTML CSS JavaScript is the superior programming language in this field, followed by Asp. Net and Typescript. The total percentage of these three programming languages was 73.9%, a high percentage that shows their superiority. HTML CSS JavaScript shows the most superior and widely used programming as evidenced by the percentage produced is 59.18%, more than half of the existing value. In addition, the Server-Side programming language currently appears to have the least used trend in data science in recent years.

2) *Web development tools:* Table V shows Web development tools are often used in conjunction with programming languages to develop software applications more easily. These tools contain various types of utilities, such as frames, libraries, and applications. As seen in Table V, Adobe Photoshop, as a tool that can be used for UI / UX development in Web Development. React JS is a tool that has a JavaScript library used to build user interfaces, is in second place, followed by Rest API, a tool used to build web services so that applications can communicate with each other using the HTTP protocol. The fourth is Amazon Web, a cloud computing platform that provides various services for the development, hosting, and management of web applications and digital infrastructure. Almost the same rate value Node js which is a JavaScript-based runtime environment allows developers to run JavaScript code outside the browser, usually on a server. WordPress is a web development tool that is rarely used, as can be seen from the small percentage of tool use in Table V.

TABLE V. WEB DEVELOPMENT TOOLS

Web Development Tools	Rate %
Adobe Photoshop	48,39
React Js	6,68
Rest API	4,32
Amazon Web	4,17
Node Js	4,12
Spring Framework	3,92
Corel Draw	3,46
Web Applications	3,16
Search Engine	3,06
Graphic Illustrators	2,71
Canva Design	2,11

Web Development Tools	Rate %
Vue Js	2,11
Google Sketch Up	1,91
Net Framework	1,61
Mobile Application	1,41
Auto Cad	1,15
Data Cubes	1
Microsoft Azure	1
Java Virtual Machine	0,85
Query	0,8
Xml	0,75
Development Git	0,4
Cloud Computing	0,35
Apache Kafka	0,3
WordPress	0,25

3) *Framework*: According to Table VI, the most popular frameworks are PHP Laravel, Java Spring, Model View Controller, and Java Script, with 66.33%. Apache Spark, Angular Js, React Js, Next Js, Vue Js, and Express Js are in the middle position of their usage, with a total of 23.2%. Ruby on Rails has been the least utilized framework in recent years.

TABLE VI. FRAMEWORK

Frameworks	Rate %
Php Laravel	28,43
Java Spring	14,38
Model View Controller	11,76
Javascript Frameworks	11,76
Apache Spark	6,21
Angular Js	4,9
React Js	4,25
Next Js	3,92
Vue Js	3,92
Express Js	3,92
Redux Js	1,96
Angular Angularjs	1,96
Php Jquery	1,63
Ruby On Rails	0,98

4) *Visualization of hierarchical clustering topics*: Fig. 4 shows the relationship between various skills and topics related to web developers. The dendrogram shown in Fig. 4 is a graphical representation of hierarchical clustering. The dendrogram in Fig. 4 helps to visualize the relationship between skills or abilities related to the topic of the web developer in a

hierarchical manner. For example, Cascading Style Sheets (CSS), HTML, and JavaScript are grouped together because they are closely related to the development of web interfaces. Technically, JavaScript, CSS, HTML and Frameworks such as Angular and Spring are grouped more closely because these skills are often used together in web application development. In addition, the analytical skills and skills in the figure above have long branches, indicating greater differences from the other groups. In the Soft Skills grouping, Communication and Interpersonal skills were in a different group from technical skills, indicating that soft skills are important and conceptually different from technical web development skills.

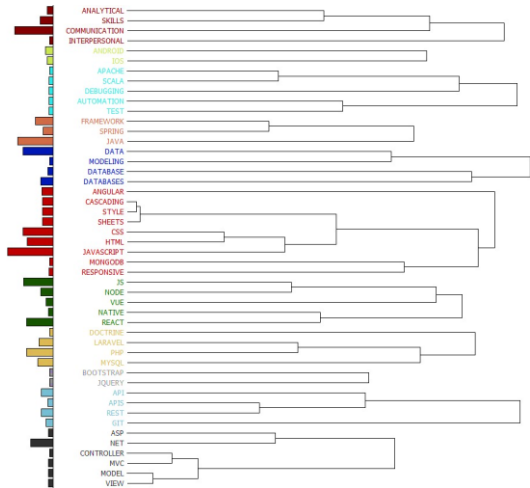


Fig. 4. Dendrogram agglomeration order of web developer and digital designers.

The visualization can be presented in a word cloud format in addition to being in a hierarchical diagram. Word Cloud is a method for visualizing text [47], [48], [49], [50]. This technique was used to identify and highlight the most frequently occurring words, thus providing insight into the dominant themes or topics in the text [51].



Fig. 5. Word cloud for web development.

Fig. 5 shows that the term 'Digital Marketing' is often paired with 'Web Development' as they complement each other in building and promoting an effective online presence. Some words that are often paired with 'web development' include various technical and non-technical aspects of web development. Here are some examples: HTML, CSS, JavaScript, Frameworks: Such as React, Angular, and Vue for the front-end, and Node.js and Django for the back end. Responsive design and Search Engine Optimization (SEO) are often paired with web development. Web Development is also often called Web Programming or Website Development or Web Application Development or Front-end Development or Back-end Development or Graphic Designer or can also be called Full-stack Development.

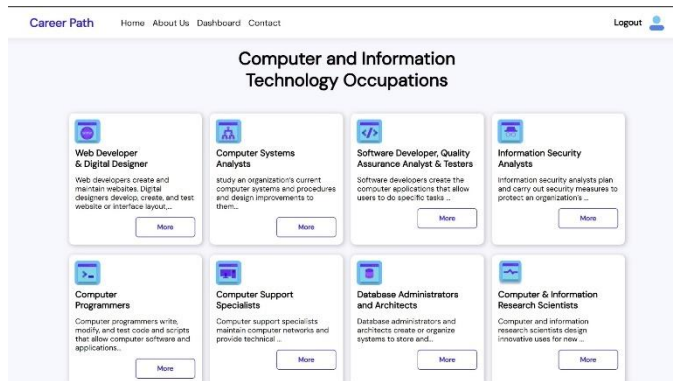


Fig. 6. Online dashboards of computer and information technology occupations.

As shown in Fig. 6, online dashboards of computer and information technology occupations has been deployed based on the ReactJS framework, and the Golang programming language.

V. DISCUSSION

This study has revealed the competencies for big data analytics over web development. First, the skill sets arranged by topic were identified from the data sets created using corpus documents of online job advertisement. Then, these skill sets were mapped into competency domains. Based on these results, the following ten competencies were identified:

- Big data product/output
- Role
- Specialized web developers
- Database
- Web development framework
- Task
- Programming languages
- Web development tools
- Educational programs
- Soft skills

The results show, in the big data web development field, HTML CSS Javascript, Asp.Net, and Typescript are the most

demanded programming languages; Adobe Photoshop, React Js, Rest API presented as the most demanded programming tools; PHP Laravel, Java Spring, Model View Controller are listed as the most demanded databases. The results of this study have important implications for web developers' programs, which are summarized below.

A. From Big Data to Data Science

Big data is related to data science which has characteristics such as volume, velocity, veracity, variability, and variety. Competence in big data requires a wide range of knowledge and skills. Such as in Table II, volume and variety of data are crucial factors in big data product/output with a contribution of 8.96%. This includes skills in digital marketing, social media, and search engine optimization and data engineering, which reflect the need for expertise in managing and analysing big data for various purposes, such as digital marketing and search engine optimization. The competence of each role of each person involved in web development work needs to adjust data with high speed and variety to support the development of more innovative products, this is related to the characteristics of big data velocity and variety with a role competency area of 33.92%. In addition, the ability of data to be relied on accurately is related to the database which includes skills in using SQL Server, MySQL, Mango DB and Oracle which contribute 2.1%. This ability is especially important in maintaining veracity which is the main basis for precise analysis and accurate data-based decisions. Big Data Product/Output and Web Development Framework, with a total contribution of 11.38%, indicate that data generated from various sources and having various types can be optimized for various analysis and processing needs. This reflects the existence of variability in data, which allows flexibility in managing and applying data for different analytical purposes. As illustrated in Fig. 5, hierarchical clustering is one of the data science analysis methods that groups similar abilities to illustrate the relationship between big data and data science.

B. The Extensive Range of Knowledge Areas and Competencies

The Web Development industry is one of the most in-demand and fastest growing professional fields worldwide. It has a highly dynamic and competitive work environment with an ever-increasing, changing, and evolving demand for knowledge, skills, and abilities. The global workforce will be impacted by the adoption of AI, automation, and Big Data Analytics (BDA) [21]. Our analysis reveals the knowledge and skill domains that are in high demand for Big Data Analytics (BDA). The analysis findings indicate that expertise in Big Data Analytics (BDA) requires a broad spectrum of highly varied and interrelated knowledge, skills, and ability domains. It involves collecting, storing, processing, and analysing large amounts of data to generate insights that can be used for decision making. Taking these findings into account, a conceptual competency map is proposed to organize these knowledge and skills. The map consists of the following ten competency domains: big data products/outputs, roles, specialized web developers, databases, web development frameworks, tasks, programming languages, web development tools, educational background, and soft skills (see Table III).

The competencies found indicate that BDA expertise has an interdisciplinary background that requires the integration of a broad set of technical and non-technical skills. Although competency priorities vary from position to position, employers in the BDA industry generally demand a set of technical and non-technical skills, defined as the job skill set. In this regard, our analysis results offer a more comprehensive perspective for BDA employers to identify the job skill set required for effective candidate assessment. The knowledge domains and skill set also indicate the need for a demand-driven educational background approach based on interdisciplinary collaboration to achieve a competency-based web development curriculum. Our findings are also in line with industry reports and academic research that emphasize the use of technical and non-technical skills together based on an interdisciplinary background containing data science, web development, software engineering, computer science, mathematics, business science, statistics, and communication science.

C. Reconciling Hard Skills and Soft Skills for Web Developers

Soft skills and hard skills have important roles in different types of jobs, although their roles can differ depending on the field and position. Hard skills are technical and specific skills that can be measured and are usually acquired through education or training. Technical skills in web development such as mastery of programming languages, tools, or frameworks. Professional qualifications such as certifications or licenses required for a specific job. Soft skills are interpersonal and character skills that are harder to measure but are essential for success in the workplace. Soft skills relating to interpersonal capabilities such as problem solving, analytical thinking, and communication. Soft skills depend on regular activities and organizational experiences of people [52], [53]. Problem solving means the ability to think critically and find solutions to problems that arise. Communication is the ability to convey ideas clearly and listen to others. This study's findings, general soft skills required for BDA specialists are highly recommended to whom it may concern. The findings related to soft skills include creative design, problem solving, and communication as the most favorite soft skills needed by industries (see Table III). This perspective has total rate of soft skills is approximately 8,29% in all topics. In many jobs, soft skills such as empathy and communication facilitate good cooperation in a team, while hard skills ensure that technical tasks are completed.

D. Insights into the Use of Tools and Technologies

Web development involves a variety of tools and technologies to create, manage, and maintain a website. This study proves several aspects of its use, including frameworks, databases, APIs, and responsive design. Web developers use tools and technologies consisting of programming languages, tools and frameworks in developing web applications. The selection of these tools is tailored to the needs and latest developments in the web development industry. The results of corpus data on BDA show that HTML CSS, ASP.Net, and Typescript are the most widely used programming languages in this field. Adobe Photoshop, React Js, and rest API are web development tools that are in high demand [54], [55]. Finally, this study proves that the most widely used web development

frameworks are PHP Laravel, Java Spring and Model View Control.

VI. CONCLUSION, LIMITATION, AND FUTURE WORKS

The finding of this study is big data analytics can clearly identify the industry needs of web development areas. A brief taxonomy of web development includes programming languages, databases, and web development tools that can improve the knowledge of students or job seekers in this field. LDA and hierarchical clustering algorithms prove that web developers can improve the innovation of businesses or organizations through digital marketing strategies. This study captures the updated industrial needs of the knowledge and skills of web developers because the datasets have been collected and managed in a wide and rigorous manner. The taxonomy of BDA competencies and skill sets was justified by three web developers' professionals through forum group discussions.

This study has several limitations. The first limitation comes from data collection using the web scraping technique; the text of job advertisement is biased because it is not specific enough or does not list relevant skills that are needed for the web developer's area. Second, the software used in the text preprocessing stage did not clean the data properly. There are some terms (stop words) still appear such as 'and', 'in', and 'to', so we must delete it manually. Third, the recurrence of phrases as synonyms forces us to determine the threshold for the most prevalent terms in job advertisements as 60.

In future work, some potential areas can be improved, such as comparing open-source tools, and Python to mine data. The latent semantic analysis (LSA) approach can be implemented to calculate the accuracy as a validated model.

ACKNOWLEDGMENT

This research was supported by the Ministry of Education, Cultural, Research, and Technology of Republic Indonesia and Institut Teknologi PLN based on Agreement Grant No. 0459/E5/PG.02.00/2024.

REFERENCES

- [1] P. V. Thayyib et al., "State-of-the-Art of Artificial Intelligence and Big Data Analytics Reviews in Five Different Domains: A Bibliometric Summary," *Sustainability* (Switzerland), vol. 15, no. 5, 2023, doi: 10.3390/su15054026.
- [2] F. Gurcan and N. E. Cagiltay, "Big Data Software Engineering: Analysis of Knowledge Domains and Skill Sets Using LDA-Based Topic Modeling," *IEEE Access*, vol. 7, pp. 82541–82552, 2019, doi: 10.1109/ACCESS.2019.2924075.
- [3] B. Kumar, S. Roy, A. Sinha, C. Iwendi, and E. Strážovská, "E-Commerce Website Usability Analysis Using the Association Rule Mining and Machine Learning Algorithm," *Mathematics*, vol. 11, no. 1, 2023, doi: 10.3390/math11010025.
- [4] P. E. Justin Zuopeng Zhang, Praveen Ranjan Srivastava, Dheeraj Sharma, "Big data analytics and machine learning: A retrospective overview and bibliometric analysis," *Expert Syst Appl*, vol. 184, 2023, doi: <https://doi.org/10.1016/j.eswa.2021.115561>.
- [5] H. Zhang, Z. Zang, H. Zhu, M. I. Uddin, and M. A. Amin, "Big data-assisted social media analytics for business model for business decision making system competitive analysis," *Inf Process Manag*, vol. 59, no. 1, 2022, doi: <https://www.sciencedirect.com/science/article/pii/S0306457321002430>.

- [6] T. Issa and P. Isaias, "Usability and Human-Computer Interaction (HCI)," in Sustainable Design, London: Springer London, 2022, pp. 23–40. doi: 10.1007/978-1-4471-7513-1_2.
- [7] A. Adel, "Future of industry 5.0 in society: human-centric solutions, challenges and prospective research areas," Journal of Cloud Computing, vol. 11, no. 1, 2022, doi: 10.1186/s13677-022-00314-5.
- [8] J. L. Hopkins, "An investigation into emerging industry 4.0 technologies as drivers of supply chain innovation in Australia," Comput Ind, vol. 125, no. 103323, 2021, doi: <https://doi.org/10.1016/j.compind.2020.103323>.
- [9] S. S. Alrumiah and M. Hadwan, "Implementing big data analytics in e-commerce: Vendor and customer view," IEEE Access, vol. 9, pp. 37281–37286, 2021, doi: 10.1109/ACCESS.2021.3063615.
- [10] L. Li and J. Zhang, "Research and Analysis of an Enterprise E-Commerce Marketing System Under the Big Data Environment," Journal of Organizational and End User Computing, vol. 33, no. 6, pp. 1–19, 2021, doi: 10.4018/joec.20211101.0a15.
- [11] A. Kamalaldin, D. Sjödin, D. Hullova, and V. Parida, "Configuring ecosystem strategies for digitally enabled process innovation: A framework for equipment suppliers in the process industries," Technovation, vol. 105, no. December 2019, 2021, doi: 10.1016/j.technovation.2021.102250.
- [12] C. Janiesch, B. Dinter, P. Mikalef, and O. Tona, "Business analytics and big data research in information systems," Journal of Business Analytics, vol. 5, no. 1, pp. 1–7, 2022, doi: 10.1080/2573234X.2022.2069426.
- [13] V. G. Goulart, L. B. Liboni, and L. O. Cezarino, "Balancing skills in the digital transformation era: The future of jobs and the role of higher education," Industry and Higher Education, vol. 36, no. 2, 2021, doi: <https://doi.org/10.1177/095042222110297>.
- [14] O. Cico, L. Jaccheri, A. Nguyen-Duc, and H. Zhang, "Exploring the intersection between software industry and Software Engineering education - A systematic mapping of Software Engineering Trends," Journal of Systems and Software, vol. 172, 2021, doi: 10.1016/j.jss.2020.110736.
- [15] J. Miranda et al., "The core components of education 4.0 in higher education: Three case studies in engineering education," Computers and Electrical Engineering, vol. 93, no. June, 2021, doi: 10.1016/j.compeleceng.2021.107278.
- [16] "U.S Bureau of Labor Statistics." Accessed: Jan. 13, 2022. [Online]. Available: <https://www.bls.gov/ooh/computer-and-information-technology/home.htm>
- [17] A. De Mauro, M. Greco, M. Grimaldi, and P. Ritala, "Human resources for Big Data professions: A systematic classification of job roles and required skill sets," Inf Process Manag, vol. 54, no. 5, pp. 807–817, Sep. 2018, doi: 10.1016/j.ipm.2017.05.004.
- [18] K. Mahmood, G. Rasool, F. Sabir, and A. Athar, "An Empirical Study of Web Services Topics in Web Developer Discussions on Stack Overflow," IEEE Access, vol. 11, no. February, pp. 9627–9655, 2023, doi: 10.1109/ACCESS.2023.3238813.
- [19] H. A. Goh, C. K. Ho, and F. S. Abas, "Front-end deep learning web apps development and deployment: a review," Applied Intelligence, vol. 53, no. 12, pp. 15923–15945, 2023, doi: 10.1007/s10489-022-04278-6.
- [20] N. Jayachandran, A. Abdrabou, N. Yamane, and A. Al-Dulaimi, "A Platform for Integrating Internet of Things, Machine Learning, and Big Data Practicum in Electrical Engineering Curricula," Computers, vol. 13, no. 8, p. 198, Aug. 2024, doi: 10.3390/computers13080198.
- [21] G. Li, C. Yuan, S. Kamarthi, M. Moghaddam, and X. Jin, "Data science skills and domain knowledge requirements in the manufacturing industry: A gap analysis," J Manuf Syst, vol. 60, pp. 692–706, Jul. 2021, doi: 10.1016/j.jmsy.2021.07.007.
- [22] M. Labbé, M. Landete, and M. Leal, "Dendrograms, minimum spanning trees and feature selection," Eur J Oper Res, vol. 308, no. 2, pp. 555–567, Jul. 2023, doi: 10.1016/j.ejor.2022.11.031.
- [23] P. C. Siswipraptini, H. L. H. S. Warnars, A. Ramadhan, and W. Budiharto, "Information Technology Job Profile using Average-Linkage Hierarchical Clustering Analysis," IEEE Access, vol. 11, no. September, pp. 94647–94663, 2023, doi: 10.1109/ACCESS.2023.3311203.
- [24] "Glints." Accessed: Jul. 15, 2024. [Online]. Available: <https://glints.com/id/en>
- [25] "Tech in Asia." Accessed: Jul. 15, 2024. [Online]. Available: <https://www.techinasia.com/>
- [26] "LinkedIn." Accessed: Jul. 15, 2024. [Online]. Available: <https://www.linkedin.com/>
- [27] S. García, S. Ramírez-Gallego, J. Luengo, J. M. Benítez, and F. Herrera, "Big data preprocessing: methods and prospects," Big Data Anal, vol. 1, no. 1, p. 9, Dec. 2016, doi: 10.1186/s41044-016-0014-0.
- [28] S. A. Alasadi and W. S. Bhaya, "Review of data preprocessing techniques in data mining," Journal of Engineering and Applied Sciences, vol. 12, no. 16, pp. 4102–4107, Sep. 2017, doi: 10.3923/jeasci.2017.4102.4107.
- [29] S. García, J. Luengo, and F. Herrera, "Tutorial on practical tips of the most influential data preprocessing algorithms in data mining," Knowl Based Syst, vol. 98, pp. 1–29, Apr. 2016, doi: 10.1016/j.knsys.2015.12.006.
- [30] N. M. Nawi, W. H. Atomi, and M. Z. Rehman, "The Effect of Data Preprocessing on Optimized Training of Artificial Neural Networks," Procedia Technology, vol. 11, pp. 32–39, 2013, doi: 10.1016/j.protcy.2013.12.159.
- [31] R. Friedman, "Tokenization in the Theory of Knowledge," Encyclopedia, vol. 3, no. 1, pp. 380–386, Mar. 2023, doi: 10.3390/encyclopedia3010024.
- [32] M. Kashina, I. D. Lenivtceva, and G. D. Kopanitsa, "Preprocessing of unstructured medical data: the impact of each preprocessing stage on classification," Procedia Comput Sci, vol. 178, pp. 284–290, 2020, doi: 10.1016/j.procs.2020.11.030.
- [33] E. Elakiya and N. Rajkumar, "Designing preprocessing framework (ERT) for text mining application," in 2017 International Conference on IoT and Application (ICIOT), IEEE, May 2017, pp. 1–8. doi: 10.1109/ICIOTA.2017.8073613.
- [34] S. Ahmad and R. Varma, "Information extraction from text messages using data mining techniques," Malaya Journal of Matematik, vol. 5, no. 1, pp. 26–29, Jan. 2018, doi: 10.26637/MJM0S01/05.
- [35] M. Fachrurrozi, N. Yusliani, and M. M. Agustin, "Identification of Ambiguous Sentence Pattern in Indonesian Using Shift-Reduce Parsing," 2014.
- [36] P. EBDEN and R. SPROAT, "The Kestrel TTS text normalization system," Nat Lang Eng, vol. 21, no. 3, pp. 333–353, May 2015, doi: 10.1017/S1351324914000175.
- [37] M. Khader, A. Awajan, and G. Al-Naymat, "The Effects of Natural Language Processing on Big Data Analysis: Sentiment Analysis Case Study," in 2018 International Arab Conference on Information Technology (ACIT), IEEE, Nov. 2018, pp. 1–7. doi: 10.1109/ACIT.2018.8672697.
- [38] D. J. Ladani and N. P. Desai, "Stopword Identification and Removal Techniques on TC and IR applications: A Survey," in 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), IEEE, Mar. 2020, pp. 466–472. doi: 10.1109/ICACCS48705.2020.9074166.
- [39] S. M. Basha and D. S. Rajput, "Evaluating the Impact of Feature Selection on Overall Performance of Sentiment Analysis," in Proceedings of the 2017 International Conference on Information Technology, New York, NY, USA: ACM, Dec. 2017, pp. 96–102. doi: 10.1145/3176653.3176665.
- [40] D. R. Rakhimova and A. O. Turganbaeva, "Normalization of Kazakh language words," Scientific and Technical Journal of Information Technologies, Mechanics and Optics, vol. 20, no. 4, pp. 545–551, Aug. 2020, doi: 10.17586/2226-1494-2020-20-4-545-551.
- [41] N. Yusliani, R. Primartha, and M. Diana, "Multiprocessing Stemming: A Case Study of Indonesian Stemming," Int J Comput Appl, vol. 182, no. 40, pp. 15–19, Feb. 2019, doi: 10.5120/ijca2019918476.
- [42] A. S. Rizki, A. Tjahyanto, and R. Trialih, "Comparison of stemming algorithms on Indonesian text processing," TELKOMNIKA (Telecommunication Computing Electronics and Control), vol. 17, no. 1, p. 95, Feb. 2019, doi: 10.12928/telkomnika.v17i1.10183.
- [43] R. Pramana, Debora, J. J. Subroto, A. A. S. Gunawan, and Anderies, "Systematic Literature Review of Stemming and Lemmatization Performance for Sentence Similarity," in 2022 IEEE 7th International Conference on Information Technology and Digital Applications (ICITDA), IEEE, Nov. 2022, pp. 1–6. doi: 10.1109/ICITDA55840.2022.9971451.

- [44] M. Javed and S. Kamal, "Normalization of Unstructured and Informal Text in Sentiment Analysis," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 10, 2018, doi: 10.14569/IJACSA.2018.091011.
- [45] L. Wang, "Design of Network Public Opinion Monitoring System based on LDA Model," in *2nd International Conference on Integrated Circuits and Communication Systems, ICICACS 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ICICACS60521.2024.10498592.
- [46] M. Pejic-Bach, T. Bertonecel, M. Meško, and Ž. Krstić, "Text mining of industry 4.0 job advertisements," *Int J Inf Manage*, vol. 50, pp. 416–431, Feb. 2020, doi: 10.1016/j.ijinfomgt.2019.07.014.
- [47] N. Chintalapudi, G. Battineni, M. Di Canio, G. G. Sagaro, and F. Amenta, "Text mining with sentiment analysis on seafarers' medical documents," *International Journal of Information Management Data Insights*, vol. 1, no. 1, p. 100005, Apr. 2021, doi: 10.1016/j.jjime.2020.100005.
- [48] J. E. Montandon, C. Politowski, L. L. Silva, M. T. Valente, F. Petrillo, and Y.-G. Guéhéneuc, "What skills do IT companies look for in new developers? A study with Stack Overflow jobs," *Inf Softw Technol*, vol. 129, p. 106429, Jan. 2021, doi: 10.1016/j.infsof.2020.106429.
- [49] L. T. Khrais, "Role of Artificial Intelligence in Shaping Consumer Demand in E-Commerce," *Future Internet*, vol. 12, no. 12, p. 226, Dec. 2020, doi: 10.3390/fi12120226.
- [50] C. Zucco, B. Calabrese, G. Agapito, P. H. Guzzi, and M. Cannataro, "Sentiment analysis for mining texts and social networks data: Methods and tools," *WIREs Data Mining and Knowledge Discovery*, vol. 10, no. 1, Jan. 2020, doi: 10.1002/widm.1333.
- [51] H. Ren, Y. Liu, G. Naren, and J. Lu, "The impact of multidirectional text typography on text readability in word clouds," *Displays*, vol. 83, p. 102724, Jul. 2024, doi: 10.1016/j.displa.2024.102724.
- [52] J. Lamri and T. Lubart, "Reconciling Hard Skills and Soft Skills in a Common Framework: The Generic Skills Component Approach," *J Intell*, vol. 11, no. 6, p. 107, Jun. 2023, doi: 10.3390/jintelligence11060107.
- [53] M. Hirudayaraj, R. Baker, F. Baker, and M. Eastman, "Soft Skills for Entry-Level Engineers: What Employers Want," *Educ Sci (Basel)*, vol. 11, no. 10, p. 641, Oct. 2021, doi: 10.3390/educsci11100641.
- [54] F. Gurcan and N. E. Cagiltay, "Big Data Software Engineering: Analysis of Knowledge Domains and Skill Sets Using LDA-Based Topic Modeling," *IEEE Access*, vol. 7, pp. 82541–82552, 2019, doi: 10.1109/ACCESS.2019.2924075.
- [55] K. Bajaj, K. Pattabiraman, and A. Mesbah, "Mining questions asked by web developers," in *Proceedings of the 11th Working Conference on Mining Software Repositories*, New York, NY, USA: ACM, May 2014, pp. 112–121. doi: 10.1145/2597073.2597083.

Optimizing Multi-Dimensional SCADA Report Generation Using LSO-GAN for Web-Based Applications

Fanxiu Fang¹, Guocheng Qi², Haijun Cao³, He Huang⁴, Lingyi Sun⁵, Jingli Yang⁶, Yan Sui⁷, Yun Liu⁸,
Dongqing You⁹, Wenyu Pei¹⁰

Pipe China Oil and Gas Control Center, Beijing, 10020, China^{1, 2, 4, 5, 6, 8, 9, 10}
Kunlun Digital Technology Co., Ltd. Beijing, 10020, China^{3, 7}
School of Software, Tsinghua University, Beijing, 100084, China⁴

Abstract—This paper addresses the challenges of custom-generating multi-dimensional data SCADA (Supervisory Control And Data Acquisition) reports using web technologies. To improve efficiency, reduce maintenance costs, and enhance scalability, the paper proposes a custom generation method based on the LSO-GAN (Light Spectrum Optimizer - Generative Adversarial Network) model. The study begins by analyzing the requirements for multi-dimensional SCADA reports and proposes a web-based design scheme. The LSO algorithm is employed to optimize the GAN model, enabling efficient generation of customizable SCADA reports. The proposed LSO-GAN model was validated using relevant SCADA data, with experimental results showing that the method outperformed other models in terms of accuracy and generation efficiency. Specifically, the LSO-GAN model achieved an RMSE of 14.98 and a MAPE of 0.93, surpassing traditional models such as Conv-LSTM and FC-LSTM. The custom report generation method based on LSO-GAN significantly improves the customization and generation of multi-dimensional data SCADA reports, demonstrating superior performance in both accuracy and operational efficiency.

Keywords—Web technologies; SCADA systems; report customisation; spectral optimisation algorithms; adversarial generative networks

I. INTRODUCTION

Report is a dynamic display of data information through tables, graphs and other diverse formats, it is a form of expression of data statistics [1]. As an important tool for analysing and displaying information and printing, reports can be used to quickly organize and analyze data and become an important basis for development decisions in various industries [2]. SCADA (Supervisory Control And Data Acquisition) system is a computer system used to monitor and control industrial processes [3]. SCADA systems based on Web technology in order to be accessed through a Web browser, they provide the ability to monitor and control remotely and are suitable for distributed control systems [4]. Multidimensional data reports are key components in SCADA systems, they allow users to analyse data from different perspectives and dimensions to better understand and optimise industrial processes [5]. Therefore, the study of customised methods for generating multidimensional data SCADA reports is beneficial to improve efficiency, reduce maintenance costs and increase scalability [6].

With the development of Internet and Web technologies and the diversification of users' needs for reports, the development of report generation methods has become the focus of attention of experts and scholars in the field, especially in the customisation of multi-dimensional data SCADA reports [7]. In the context of globalisation of information technology, the implementation technology of Web reporting tools has been constantly innovated and improved. At present, the Web reporting tool implementation technology is more, which is more widely used, the practicality of the better there are mainly the following three schemes [8]: (1) based on the COM components of the programme. Xie et al. [9] describe the VB environment based on the ADO pairs and COM components Excel data processing functions combined to achieve the report printing function. Chen et al. [10] developed an Excel-based put custom report dynamic library, using COM technology, the output of a new output report; (2) based on the ActiveX plug-in programme. Cuzzocrea et al. [11] combined with the Jasper Reports open source project to generate reports to meet the needs of dynamic generation of Web reports; (3) XML-based plug-in-free programme. Munz-Krner and Weiskopf [12] proposed an XML-based Web-oriented intelligent reporting system. Nasri and Weslati [13] studied the Web-based custom reporting tool design method. With the increasing data dimensionality in SCADA systems, the current report generation methods no longer meet the design requirements. For the current multi-dimensional data SCADA system requirements, this paper combines Web technology [14], intelligent optimisation algorithms [15] and neural network methods [16], and proposes a Web-based custom intelligent generation method for multi-dimensional data SCADA reports. The contributions of this paper include the following:

- (1) Describe the problem of multi-dimensional data SCADA report customisation and give relevant solutions to the problem;
- (2) Around the SCADA report customisation generation, combined with the LSO algorithm [17] and the GAN network [18], propose the SCADA report customisation generation algorithm based on the LSO-GAN;
- (3) Use the multi-dimensional data report relevant information to validate the report customisation algorithm. The results show that the method proposed in

this paper achieves the customisation of multi-dimensional data SCADA reports, and at the same time improves the intelligent generation efficiency by using LSO-GAN.

II. PROBLEM DESCRIPTION AND ANALYSIS

A. Requirements Analysis for Custom Reports on Multidimensional Data

The main engineering requirements for multidimensional data reporting enable a defined set of better report semantics to achieve custom report customisation (Fig. 1). Therefore, a good set of reporting tool software and methodology should fulfil the following functions: 1) report data management; 2) report design tools; 3) report file management; and 4) report integration and application [19], as shown in Fig. 2.



Fig. 1. Multi-dimensional data custom report style.

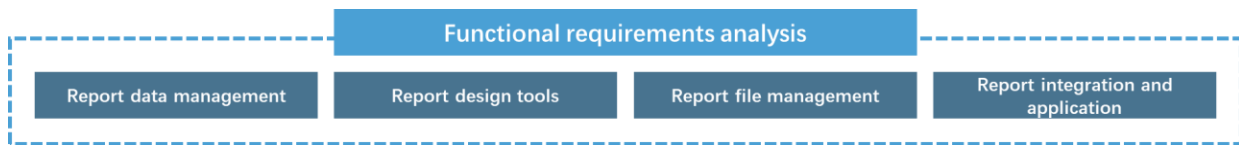


Fig. 2. Functional requirements analysis.

In addition to the analysis of functional requirements, it is also need to analyse the performance requirements of the tool, especially the custom reporting tool method of ease of use and security and reliability. 1) in terms of ease of use, report generation method needs to provide users with a friendly interface, mainly for: rapid formatting and content parsing, fast preview of the Web page, the data import and export of the report quickly; 2) security reliability, the main performance: identity verification, operation rights management, input information legality detection, deletion warning, real-time recording of important operations, as shown in Fig. 3.

B. SCADA Report Custom Generation Idea Design

1) Web-based custom reporting tool workflow: The overall workflow of the Web-based custom reporting tool is shown in Fig. 4.

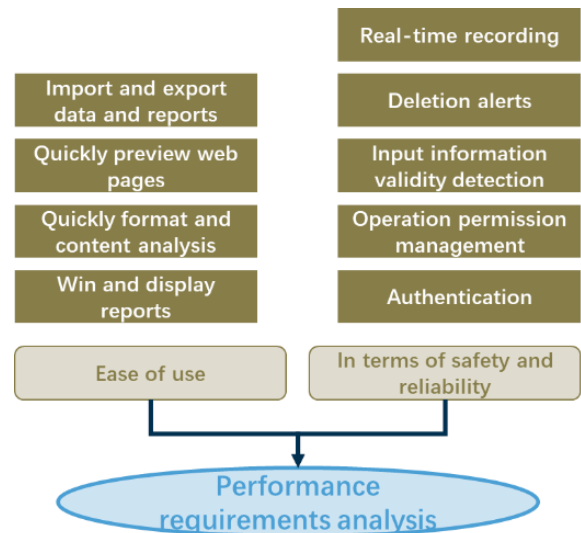


Fig. 3. Analysis of performance requirements.

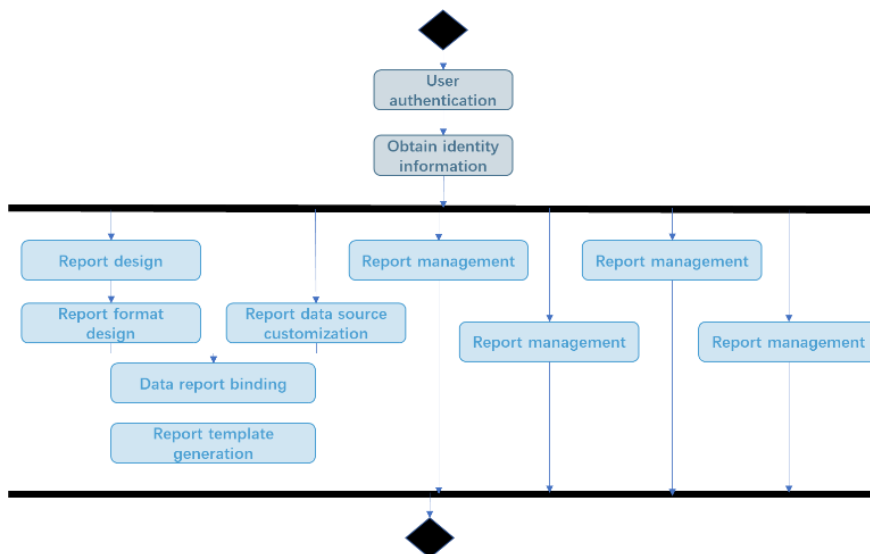


Fig. 4. Reporting tool flow.

As can be seen in Fig. 4, the user logs into the system for user authentication, obtains user identity and permission information, and enters the main interface of the SCADA reporting software. Depending on the user's privileges, one or more of the management functions, such as report customisation, report template management, report browsing and printing, report file export and system management, can be performed [20].

2) *Steps for customising multidimensional data reports:* Customising a multidimensional data report typically involves the following steps (Fig. 5): 1) identifying the purpose and user requirements of the report; 2) collecting and organising the required data sets; 3) selecting a tool that supports web development; 4) designing the visual layout of the report; 5) developing the report functionality; 6) testing and optimising; and 7) deploying and maintaining [21].

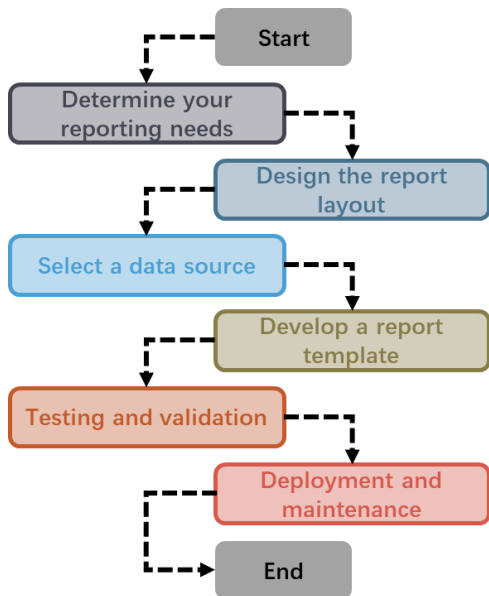


Fig. 5. Custom report workflow.

3) *Customised SCADA report generation and analysis:* In order to be in practical applications, many reports are basically the same in format and script, and the main difference lies in the different equipment of the data source. In order to further improve the efficiency of report generation, this paper uses deep learning technology and intelligent optimisation algorithms to introduce SCADA report custom generation algorithms, the specific analysis is shown in Fig. 6.

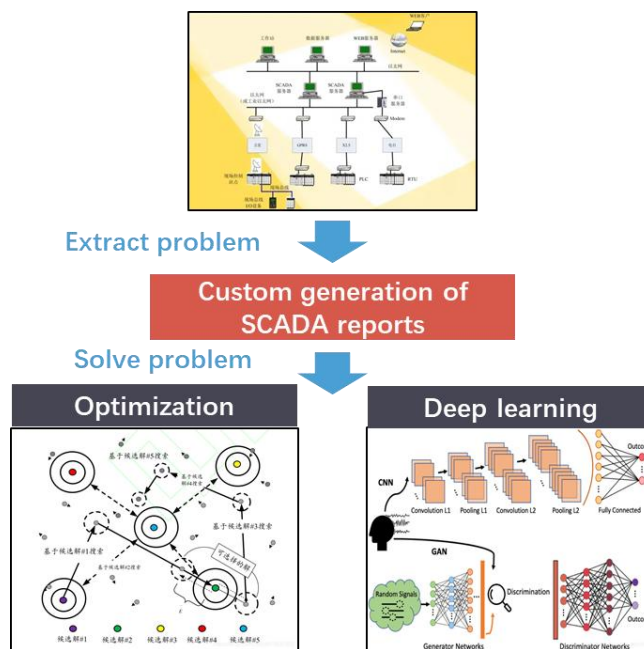


Fig. 6. Analysis of key technologies for customised generation of SCADA reports.

III. CUSTOMISED SCADA REPORT GENERATION

A. Adversarial Generative Networks

Adversarial Generative Networks (GANs) [21] were proposed by Goodfellow et al. and are shown in Fig. 7. The basic

GAN architecture consists of two fundamental components: a generator $G(z; \theta_g)$ and a discriminator $D(x; \theta_d)$ which work against each other. The generator captures the distribution P_g of data x from the noise variable $P_z(z)$ and generates

fake data that looks real and can deceive the discriminator; the discriminator distinguishes whether different categories are fake or not and acts as a classifier to model the probability of each category.

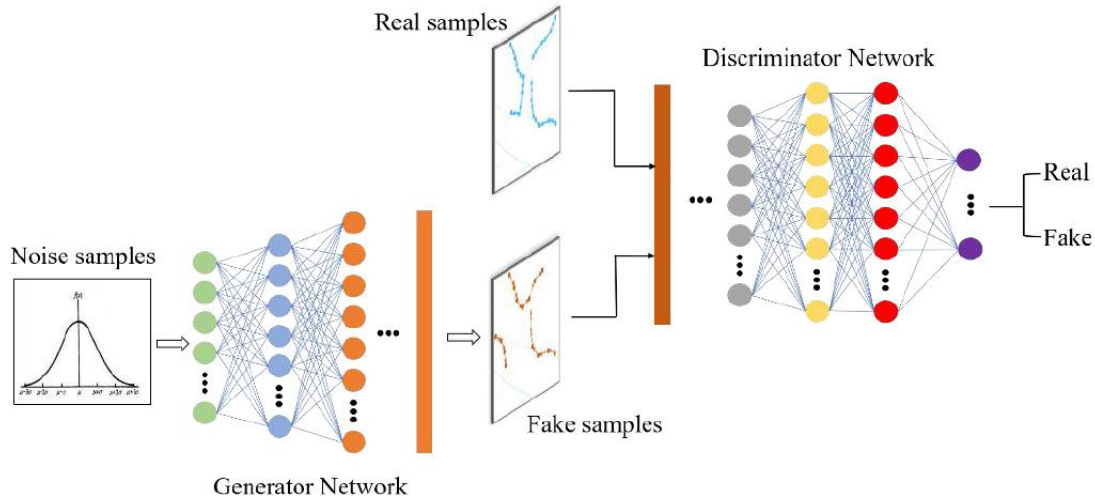


Fig. 7. Schematic diagram of GAN structure.

Ideally, the generator G can generate fake data with the data $G(z)$ and it is difficult for the discriminator to distinguish whether the data generated by G is real or not. Finally, the two

components reach dynamic equilibrium, i.e. $D(G(z)) = 0.5$:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p(z)} [\log (1 - D(G(z)))] \quad (1)$$

GANs use only backpropagation and do not require complex Markov chains to compare with other generative models such as Boltzmann machines [22]. The main advantage of GANs is that they can automatically learn the distribution of the data from the original set of samples, generating clearer and more realistic samples. Even complex distributions can be learnt by a GAN if it is trained well enough.

GAN training process: during the training process, the generator and the discriminator are updated alternately. First, the weights of the discriminator are fixed and the generator is trained to produce more plausible data; then, the weights of the generator are fixed and the discriminator is trained to better distinguish real data from generated data [23]. This process is repeated until an equilibrium point is reached, at which point the generator is able to produce samples that are virtually indistinguishable from the real data, as shown in Fig. 8.

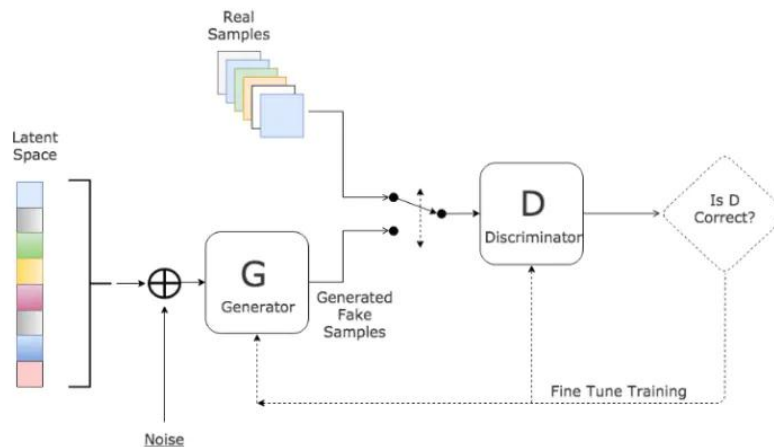


Fig. 8. GAN training process.

GANs are widely used in many fields such as image generation, image restoration, style migration, video generation, etc [24], as shown in Fig. 9. They are not only capable of generating high-quality images, but also play a role in tasks such

as image segmentation and video prediction. With the deepening of research, variants and improved versions of GANs have emerged to address the problems of the original GAN models in terms of training stability and pattern collapse.

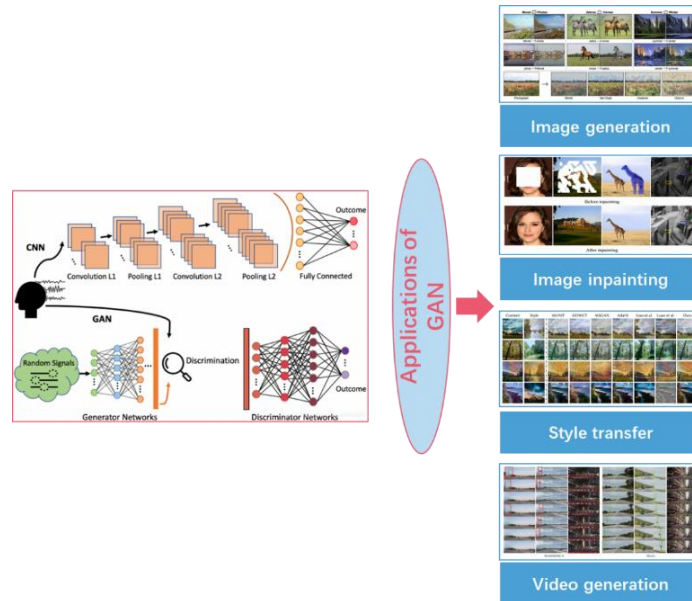


Fig. 9. GAN application.

B. LSO-GAN Network

1) *LSO algorithm*: Light Spectrum Optimizer (LSO) [25] is a meta-heuristic algorithm, an optimisation algorithm based on spectral analysis, which simulates the spectral distribution and peak search process in spectral analysis. The algorithm adaptively adjusts the resolution of the search space and the search speed in order to find the optimal solution quickly and accurately, which has the characteristics of fast convergence and high solution accuracy. The optimization process of the LSO algorithm includes the optimization strategies such as initialization, generation of new coloured rays, and dispersion of the coloured rays, which are shown in Fig. 10.

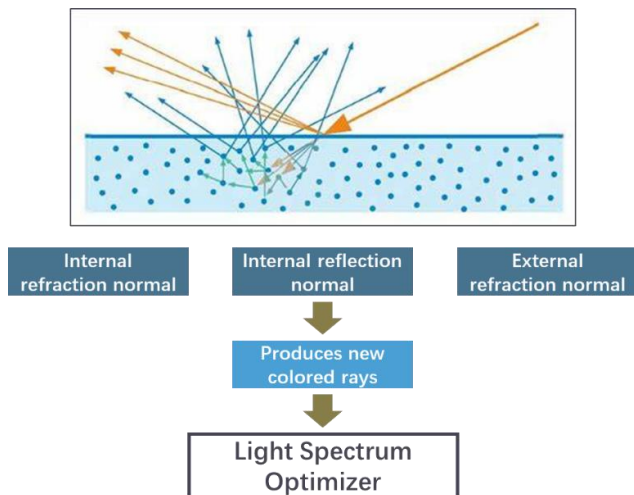


Fig. 10. Analysis of optimisation strategy of LSO algorithm.

a) *Initialisation*: The LSO algorithm uses a random initialisation strategy to model the white light population:

$$x^0 = lb + RV_1 \times (ub - lb) \quad (2)$$

where x^0 denotes the white light initialisation population, lb and ub denote the search lower and upper bounds respectively, and RV_1 denotes the random vector.

b) *The direction of the rainbow spectrum*: After initialisation, the internal refraction normal vector, internal reflection normal vector and external refraction normal vector are calculated as follows:

$$x_{nA} = \frac{x_t^r}{\text{norm}(x_t^r)} \quad (3)$$

$$x_{nB} = \frac{x_t^p}{\text{norm}(x_t^p)} \quad (4)$$

$$x_{nC} = \frac{x^*}{\text{norm}(x^*)} \quad (5)$$

Where x_{nA} , x_{nB} and x_{nC} denote the internal refraction normal vector, internal reflection normal vector and external refraction normal vector respectively, x_t^r denotes the current randomly selected ray, x_t^p denotes the current ray individual,

and x^* denotes the current optimal ray individual. *norm* denotes the normalisation method.

For incident light, it is calculated using the averaging method with the following formula:

$$X_{mean} = \frac{\sum_{i=1}^N x_i}{N} \quad (6)$$

$$x_{L0} = \frac{X_{mean}}{norm(X_{mean})} \quad (7)$$

where x_{L0} denotes the incident light, X_{mean} denotes the average position information of the light population, and N is the population size magnitude.

The mathematical model for the calculation of internal and external refracted and reflected rays is as follows:

$$x_{L1} = \frac{1}{k^r} (x_{L0} - x_{nA} (x_{nA} \cdot x_{L0})) - x_{nA} \left| 1 - \frac{1}{(k^r)^2} + \frac{1}{(k^r)^2} (x_{nA} \cdot x_{L0})^2 \right|^{\frac{1}{2}} \quad (8)$$

$$x_{L2} = x_{L1} - 2x_{nB} (x_{L1} \cdot x_{nB}) \quad (9)$$

$$x_{L3} = k^r (x_{L2} - x_{nC} (x_{nC} \cdot x_{L2})) + x_{nC} \left| 1 - (k^r)^2 + (k^r)^2 (x_{nC} \cdot x_{L2})^2 \right|^{\frac{1}{2}} \quad (10)$$

Where x_{L1} , x_{L2} and x_{L3} denote refracted, internally reflected and externally refracted rays respectively, and k^r denotes the refractive index, a random spectral colour can be defined:

$$k^r = k^{red} + RV_1 (k^{violet} - k^{red}) \quad (11)$$

where RV_1 denotes a random number.

c) *Generation of new coloured rays (Exploration search mechanism)*: After calculating the light direction, the random vector is used to calculate and select the candidate solution location information as modelled below:

$$x_{t+1} = \begin{cases} x_t + \varepsilon RV_1^n GI (x_{L1} - x_{L3}) \times (x_{r1} - x_{r2}) & p < rand \\ x_t + \varepsilon RV_2^n GI (x_{L2} - x_{L3}) \times (x_{r3} - x_{r4}) & p \geq rand \end{cases} \quad (12)$$

Where x_{r1} , x_{r2} , x_{r3} and x_{r4} denote randomly selected light individuals, RV_1^n and RV_2^n denote uniformly distributed vectors, ε denotes the scaling factor, and GI is the adaptive control factor based on the inverse incomplete function.

$$\varepsilon = a \times RV_3^n \quad (13)$$

$$GI = a \times r^{-1} \times P^{-1}(a, 1) \quad (14)$$

$$a = RV_2 \left(1 - \frac{t}{T \max} \right) \quad (15)$$

Where RV_3^n denotes normally distributed random numbers with 0 as the mean and 1 as the standard deviation, a denotes adaptive parameters, r denotes random numbers, P^{-1} is the inverse incomplete function, t is the current number of iterations, and $T \max$ is the maximum number of iterations. Different values of a have different values of adaptive control factors, which mainly control the balance of development and exploration behaviour operations, the specific curve changes are shown in Fig. 11.

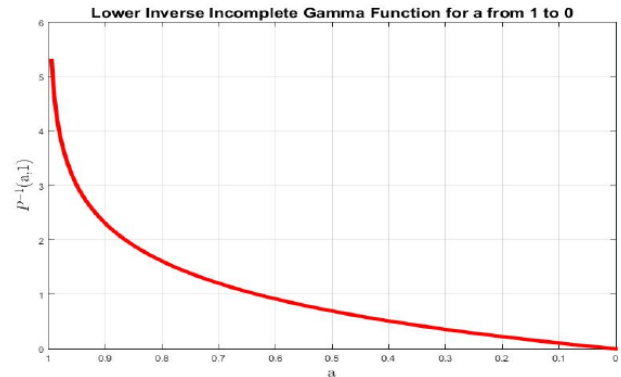


Fig. 11. Inverse incomplete function variation curve.

d) *Coloured light dispersion (Exploitation search mechanism)*: The optimisation model for the colour light dispersion phase is calculated as follows:

$$x_{t+1} = \begin{cases} x_t + RV_3 \times (x_{r1} - x_{r2}) + RV_4^n \times (R < \beta) \times (x^* - x_t) & R < P_e \\ 2 \cos(\pi \times r_1) (x^*) (x_t) & \text{Otherwise} \end{cases} \quad (16)$$

where RV_3 is a random number chosen between 0 and 1, RV_4^n is a random vector, x^* denotes the best ray individual, x_{r1} and x_{r2} denote randomly chosen ray individuals, r_1 denotes a random number, P_e denotes a predetermined probability, and R is a random number.

The final scattering phase is to generate new ray individuals based on random ray individuals and current individuals:

$$x_{t+1} = (x_{r1}^p + |RV_5| \times (x_{r2} - x_{r3})) \times U + (1 - U) \times x_t \quad (17)$$

Where RV_5 denotes normally distributed random numbers and U denotes that 0's and 1's are random vectors.

Switching Eq. (22) with Eq. (21) based on the difference in calculated fitness values is as follows:

$$x_{t+1} = \begin{cases} Eq.(21) & R < P_s | F' < R_1 \\ Eq.(22) & Otherwise \end{cases} \quad (18)$$

$$F' = \left| \frac{F - F_b}{F_b - F_w} \right| \quad (19)$$

Among them, F' denotes the normalised value of the fitness value of the current light individual (the relationship

between F' and R_1 is shown in Fig. 12), F , F_b and F_w denote the fitness values of the current individual, the optimal solution individual, and the worst solution individual, respectively, the predetermined probability P_s is used to promote the acceleration of the first dispersal stage and the second dispersal stage to the vicinity of the optimal solution, and R_1 and R are random numbers.

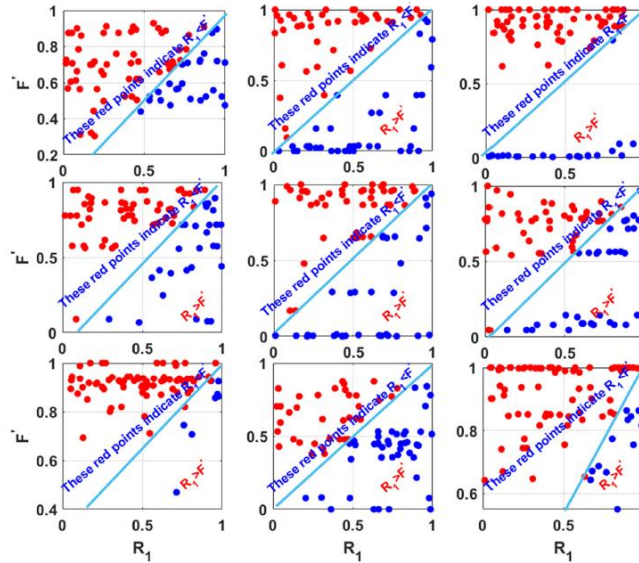


Fig. 12. Schematic diagram of the relationship between R1 and F'.

The pseudo-code of the LSO algorithm is shown in Table I.

TABLE I. PSEUDO-CODE OF LSO ALGORITHM

Algorithm 1: LSO algorithm pseudo-code

Inputs: ray population size N, maximum number of iterations Tmax;

- 1 Generating initialised random population rays;
- 2 t=0;
- 3 While t<Tmax
- 4 Evaluate light adaptation values;
- 5 t=t+1;
- 6 Update the optimal solution;
- 7 Calculate the ray normal vector;
- 8 Calculate refracted, internally reflected and externally refracted rays;
- 9 Update refractive index, scaling factor, adaptive control factor, adaptive parameters;
- 10 Generate random numbers p, q;
- 11 Update the light using the Exploration search mechanism;
- 12 Evaluate light adaptation values;
- 13 t=t+1; update the optimal solution;
- 14 Update the light using the Exploitation search mechanism;
- 15 End while

Output: optimal light and its adaptation value.

2) *LSO-GAN network*: In order to improve the efficiency of GAN network generation, this paper adopts the LSO algorithm to optimise the GAN hyper-parameters, and uses the RMSE as the fitness value to improve the optimisation of the GAN using the LSO optimisation strategy, and the specific optimisation structure is shown in Fig. 13.

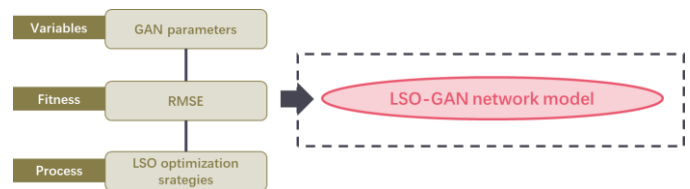


Fig. 13. Principle of LSO-GAN structure.

C. LSO-GAN in Multi-Dimensional Data SCADA Report Customisation

In order to improve the efficiency of report generation, this paper adopts LSO-GAN network to build multi-dimensional data SCADA report custom generation model, the specific application is shown in Fig. 14. Based on LSO-GAN network multi-dimensional data SCADA report custom generation method mainly includes five parts: multi-dimensional data SCADA report design requirements analysis, report layout design, data selection and processing, SCADA report generation, generation model performance analysis. As the key part of the multi-dimensional data SCADA report customisation

problem, SCADA report generation uses LSO-GAN network to construct multi-dimensional data SCADA report customisation generation model by training the acquired SCADA data, and

realises the intelligent generation of SCADA report customisation.

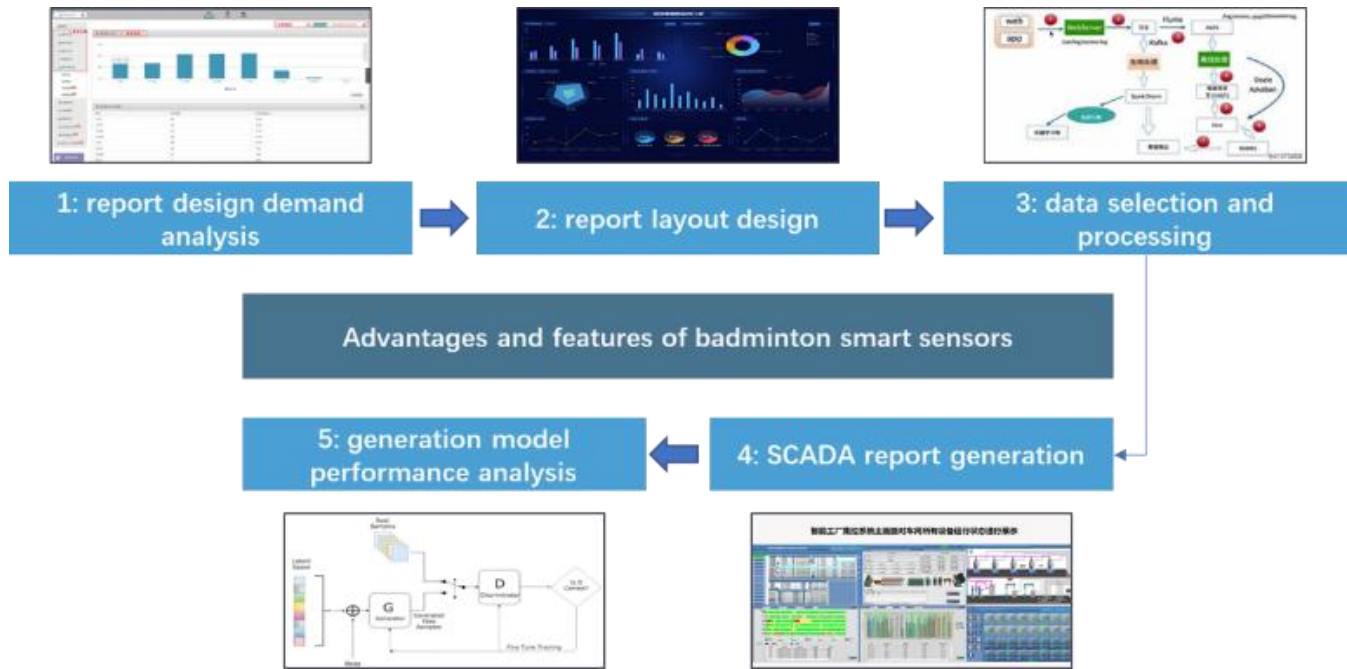


Fig. 14. LSO-GAN network application analysis.

IV. EXPERIMENTAL ANALYSIS AND DISCUSSION

A. Environmental Settings

In order to verify the high efficiency of the LSO-GAN network application proposed in this paper, this paper takes the

SCADA system report data as the analysed data, and adopts Conv-LSTM, FC-LSTM, DyConv-LSTM, CNN-LSTM as the comparison algorithms of the LSO-GAN network, and the specific parameter settings are shown in Table II.

TABLE II. COMPARISON NETWORK PARAMETER SETTINGS

Arithmetic	Parameterisation
ConvLSTM	A GAN structure is used with ConvLSTM for generator and discriminator, Tanh activation function is used for generator and the discriminator output consists of Sigmoid activation function using Adam optimiser with learning rate 0.0002, batch size 64 and iteration number 400.
FC-LSTM	A GAN structure is used with FC-LSTM for the generator and discriminator, using the Adam optimiser with a learning rate of 0.0002, a batch size of 4 and an iteration count of 400.
DyConv-LSTM	A GAN structure is used, with DyConv-LSTM for generator and discriminator, Adam optimiser, learning rate 0.0002, batch size 16, and number of iterations 400.
CNN-LSTM	A GAN structure is used with CNN-LSTM for the generator and discriminator, Adam optimiser with a learning rate of 0.0002, batch size of 64 and 1000 iterations.
LSO-GAN	The population size of the LSO algorithm is chosen to be 100 and the maximum number of iterations is set to 400.

B. Presentation and Discussion of Results

1) *Analysis of the effect of custom report generation:* In order to verify the feasibility of the report custom generation effect, this paper takes the enterprise SCADA system report design as a case study, and obtains the effect diagrams in Fig. 15 and Fig. 16.

Fig. 15 gives the effect of SCADA system report user function privilege configuration. Administrators can bind

functional rights for roles in the Role Management - Function Configuration module, and then bind roles for users in the Personnel Management - Edit Personnel module, at which time users have the rights to the configured functions. Figure 16 shows the effect of the multi-dimensional data report of SCADA system. Users in this page according to the needs of flexible configuration of the number of conditions, page configuration is complete, click the "query" button, the system to implement the logic of the number of query results in the form of a list of displays to meet user needs.

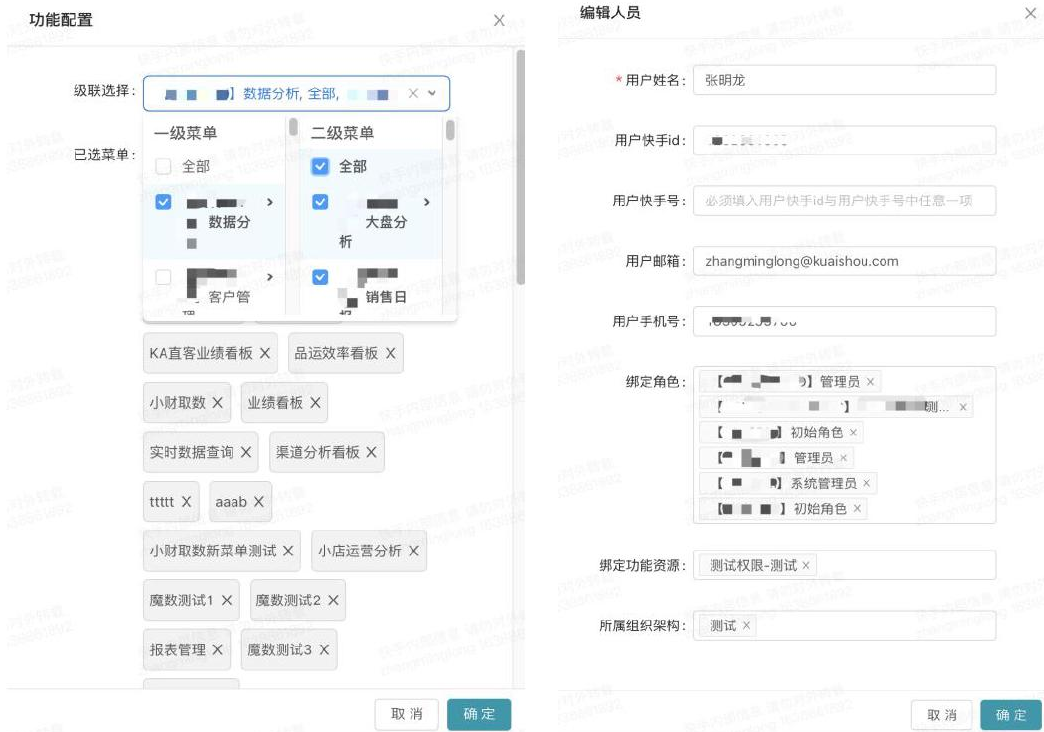


Fig. 15. Custom reporting user function permission configuration effect.

当前时间	营业执照	消耗	现金消耗	封面曝光数	封面点击数	封面点击率	素材曝光数	行为率	封面CPM	素材CPM	封面CPG	3s播放数	5s播放数
汇总	-	160,797,003.5 19	69,432,937.6 42	873,124,015	78,079,337	8.94%	10,179,758,72 3	2.63%	184.163	15.796	2.059	2,472,813,826	1,933,044,36 4
2021-12-05~2021-12-05		1,654,208.95	1.832	189,863	13,586	7.16%	221,373,772	2.78%	8712.645	7.472	121.758	57,634,672	43,471,386
2021-12-05~2021-12-05		1,528,909.753	74,792.463	27,351	3,448	12.61%	46,498,025	1.37%	55899.593	32.881	443.419	18,877,193	14,771,610
2021-12-05~2021-12-05		1,392,685.428	101,727.599	34,056	6,522	19.10%	43,259,191	1.49%	40893.981	32.194	213.537	12,876,944	10,623,677
2021-12-05~2021-12-05		1,257,503.193	1,093,877.046	21,506,373	747,729	3.48%	28,982,014	3.15%	58.471	43.389	1.682	1,323,518	1,009,014
2021-12-05~2021-12-05		1,146,485.65	1,057.586	740,905	56,022	7.56%	118,375,154	1.14%	1547.412	9.685	20.465	22,130,483	14,790,440
2021-12-05~2021-12-05		1,132,721.458	7,447.29	378,197	35,590	9.44%	170,725,405	1.66%	2995.957	6.635	31.738	42,212,029	32,337,961
2021-12-05~2021-12-05		1,117,412.602	960,950.927	2,352	43	1.83%	69,919,606	1.93%	475090.392	15.981	25986.34	15,926,117	10,108,795
2021-12-05~2021-12-05		1,099,084.89 6	0	261,861	19,878	7.59%	62,431,207	0.88%	4197.207	17.605	55.292	10,562,836	7,710,670
2021-12-05~2021-12-05		1,993,573.894	11,830.755	59,413	5,305	8.93%	74,547,412	2.42%	16400.306	14.67	208.14	52,272,198	48,878,439
2021-12-05~2021-12-05		928,488.368	12,527.29	23,781	2,688	11.3%	22,770,554	0.51%	39043.285	40.776	345.42	3,734,557	1,980,836

共 11582 条 10条/页 < 1 2 3 4 5 6 ... 1159 > 前往 2 页

Fig. 16. Multi-dimensional data report display effect.

2) *Algorithm performance analysis:* In order to verify the report generation effect of LSO-GAN network, this paper adopts Conv-LSTM, FC-LSTM, DyConv-LSTM, CNN-LSTM as the comparison algorithms, and evaluates the models in terms of RMSE, MAPE, training time, generation time, etc., and the specific results are shown in Table III, Fig. 17 to Fig. 20.

Table III gives the comparison of performance results of different SCADA multidimensional data report definition generation networks. From Table III, it can be seen that in terms of RMSE, the SCADA multidimensional data report definition generation method based on LSO-GAN network is better than other networks and has a value of 14.98, and in terms of MAPE, LSO-GAN network is better than Conv-LSTM, FC-LSTM, DyConv-LSTM, and CNN-LSTM, and has a value of 0.93.

TABLE III. COMPARISON OF THE PERFORMANCE OF DIFFERENT REPORT CUSTOM GENERATION NETWORKS

	Arithmetic	RMSE	MAPE
Conv-LSTM		15.98	0.89
FC-LSTM		16.79	0.86
DyConv-LSTM		28.67	2.25
CNN-LSTM		26.77	1.52
LSO-GAN		14.98	0.93

Fig. 17 gives the comparison of the accuracy of different networks for different time periods. From Fig. 17, it can be seen that the RMSE value of SCADA multidimensional data report definition generation method based on LSO-GAN network is less than other algorithms, which indicates that LSO-GAN network has better generation accuracy.

The training optimisation time and test generation time of different SCADA multidimensional data report custom generation methods are given in Fig. 18 and Fig. 19 respectively. From Fig. 18 and Fig. 19, it can be seen that the LSO-GAN network is ranked first in terms of both training optimisation time and test generation time, which are 38.4921s and 0.02772, respectively. This shows that the LSO-GAN network is the most efficient in generating reports.

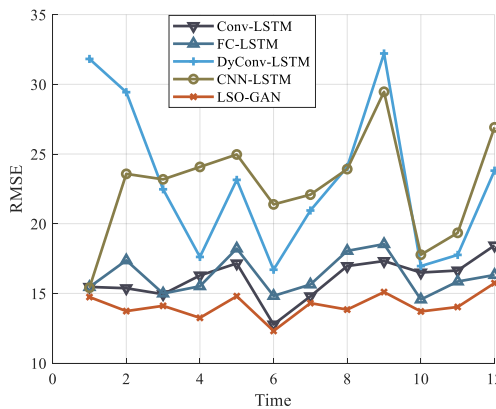


Fig. 17. Comparison of model accuracy performance of the compared algorithms.

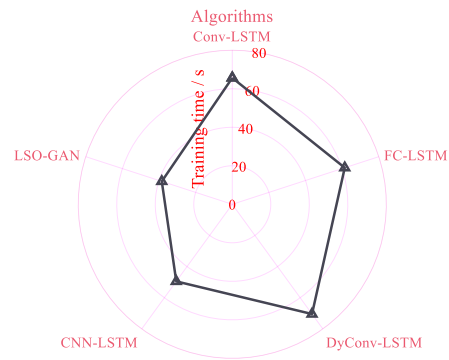


Fig. 18. Comparison of training time for different report custom generation networks.

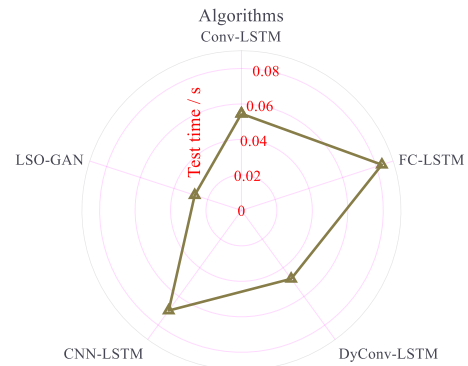
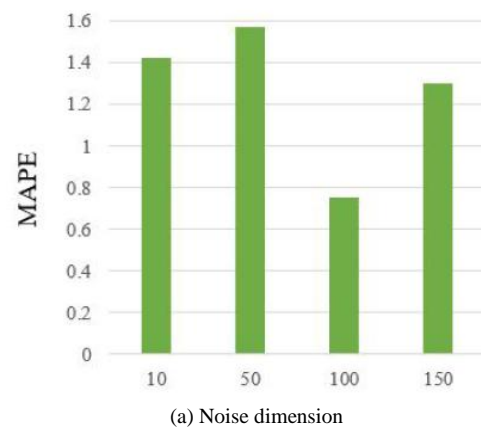


Fig. 19. Comparison of time spent on different report customisation generation networks.

Fig. 20 gives the effect of different parameters on the generative performance of the LSO-GAN algorithm. From Fig. 20, it can be seen that the error is high when the noise dimension is too low or too high; the model using 2-heads of self-attention has better performance than the model using 1-heads of self-attention layer, but with the increase of the number of heads, the error is increasing; the model is very sensitive to the hidden features, and the more hidden features in the dynamic convolution layer, the better the performance is, which indicates that the more the weights in the dynamic convolution, the better the ability to capture the statement Model.



(a) Noise dimension

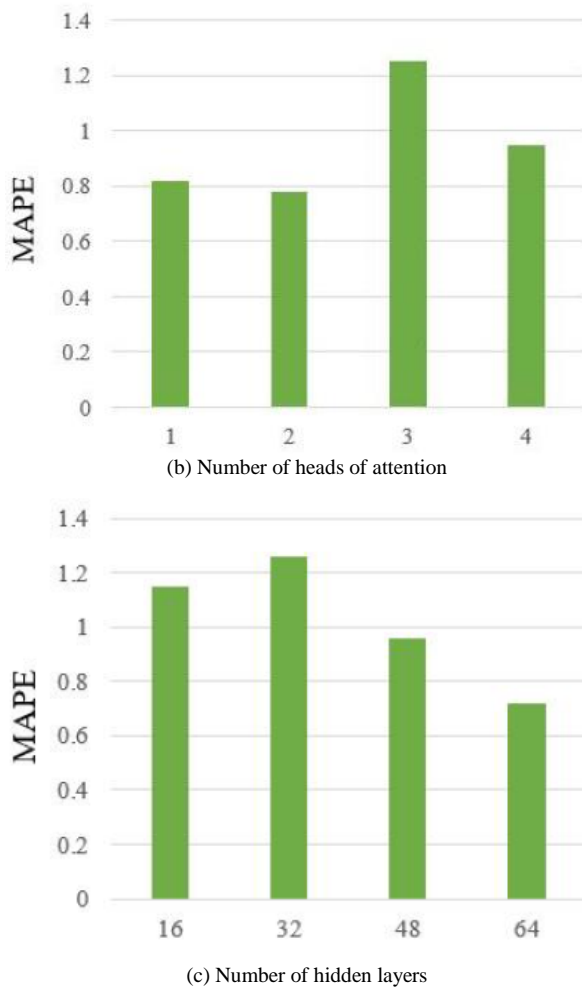


Fig. 20. Effect of parameters on report generation.

V. CONCLUSION AND OUTLOOK

The document discusses a method for customizing the generation of multi-dimensional SCADA (Supervisory Control and Data Acquisition) reports based on web technologies. By introducing the Light Spectrum Optimizer Generative Adversarial Network (LSO-GAN) model, the method aims to improve report generation efficiency, reduce maintenance costs, and enhance scalability. The study demonstrates that LSO-GAN outperforms traditional models like Conv-LSTM and FC-LSTM in terms of accuracy and generation efficiency, achieving an RMSE of 14.98 and a MAPE of 0.93 in experiments.

- The study analyzes the requirements for multi-dimensional SCADA report customization and designs a web-based generation workflow.
- The LSO-GAN model, which integrates the Light Spectrum Optimizer algorithm, improves the hyperparameter optimization of GANs, significantly enhancing the efficiency and intelligence of report generation.
- Experimental validation shows that the method excels in accuracy, training time, and generation time, especially for handling and generating multi-dimensional data.

In the future, we could work on expanding the model's application to other domains with different multi-dimensional datasets to validate its broader applicability; investigating ways to simplify the model while maintaining efficiency, thus reducing deployment and operational costs; exploring enhanced security measures, particularly in scenarios involving sensitive data, to improve the robustness of the report generation system.

REFERENCES

- [1] Parashar D .Unlocking multidimensional cancer therapeutics using geometric data science[J].Scientific Reports, 2023, 13(1).
- [2] Gupta A , Singh S , Rana H , Prashar V K , Yadav R. An OWA Based MCDM Framework for Analyzing Multidimensional Twitter Data: a Case Study on the Citizen- Government Engagement During COVID-19[J].International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2024, 32(03):355-383.
- [3] Deshpande S N , Jogdand R .A novel scheduling algorithm development and analysis for heterogeneous IoT protocol control system to achieve SCADA optimization: a next generation post covid solution[J].International Journal of Information Technology, 2023, 15:2123 - 2131.
- [4] Rusu F .Multidimensional Array Data Management[J].Found. Trends Databases, 2023, 12:69-220.
- [5] Abidine M Z E , Dutagaci H , Rousseau D .Ordinalysis: interpretability of multidimensional ordinal data[J].SoftwareX, 2023, 22:101343.
- [6] O'Brien K , Sood S , Shete R .Big Data Approach to Visualising, Analysing and Modelling Company Culture: a New Paradigm and Tool for Exploring Toxic Cultures and the Way We Work[J].International Journal of Management Science and Business Administration, 2022, 8.
- [7] An G , Cockrell C .Generating synthetic multidimensional molecular time series data for machine learning: considerations[J].Frontiers in Systems Biology, 2023.
- [8] Singh A K , Kumar J .A privacy-preserving multidimensional data aggregation scheme with secure query processing for smart grid[J].Supercomputing, 2023, 79(4):3750-3770.
- [9] Xie R , Bai S , Ma P .Optimal sampling designs for multidimensional streaming time series with application to power grid sensor data[J].The Annals of applied statistics, 2023, 17(4):3195-3215.
- [10] Chen H , Shen G Q , Feng Z , Liu Y. Optimization of energy-saving retrofit solutions for existing buildings: a multidimensional data fusion approach[J].Renewable and Sustainable Energy Reviews, 2024, 201.
- [11] Cuzzocrea A , Karras P , Vlachou A .Effective and efficient skyline query processing over attribute-order-preserving-free encrypted data in cloud-enabled databases[J].Future Generation Computer Systems, 2022, 126:237-251.
- [12] Munz-Krner T , Weiskopf D .Exploring visual quality of multidimensional time series projections[J].Visual Informatics, 2024, 8(2):27-42.
- [13] Nasri K , Weslati A .Targeting Household Deprivations for Multidimensional Poverty Alleviation: An Application to Tunisian Data[J].GLO Discussion Paper Series, 2022.
- [14] Yang Q .Designing remote sharing system of network education resources for software engineering specialty based on web technology[J]. International journal of computational systems engineering, 2024, 8(1/2):20-29.
- [15] Yuxuan Zhang,Nan Zhang. Optimal planning of local motion trajectories for intelligent sorting machines based on graph optimisation DWA algorithm[J]. Computer Measurement and Control,2024,32(09):315-321.
- [16] Hu X , Hu X , Li J Y K .Generative Adversarial Networks for Video Summarization Based on Key-frame Selection[J].Informacines Technologijos ir Valdymas, 2023, 52(1):185-198.
- [17] Liu Miaomiao,Zhang Yuying,Guo Jingfeng,Chen Jing. An adaptive lion group optimisation algorithm incorporating multi-strategy improvement[J]. Journal of Beijing University of Posts and Telecommunications,2024,47(01):85-93.

- [18] Huang S , Chen Y .Generative Adversarial Networks with Adaptive Semantic Normalization for text-to-image synthesis[J].Digital Signal Processing, 2022:120.
- [19] Poltavtseva M A , Andreeva T M .Methods of Multidimensional Aggregation of Time Series of Streaming Data for Cyber-Physical System Monitoring[J]. Automatic Control and Computer Sciences, 2022.
- [20] Zabaryo M , Barszcz T .Proposal of Multidimensional Data Driven Decomposition Method for Fault Identification of Large Turbomachinery[J]. Energies, 2022, 15.
- [21] Qi H , Li F , Tan S , Zhang X. Training Generative Adversarial Networks with Adaptive Composite Gradient[J].Data Intelligence, 2024(1).
- [22] Gonzalez-Abril L .Generative Adversarial Networks in Business and Social Science[J].Applied Sciences, 2024, 14.
- [23] Kong X , Bi J , Chen Q , Shen G, Chin T, Pau G. Traffic trajectory generation via conditional Generative Adversarial Networks for transportation Metaverse[J].Applied Soft Computing, 2024, 160.
- [24] Liu D R , Huang Y , Lee T S J .Hybrid Generative Adversarial Networks for News Recommendation[J].Journal of information science and engineering: JISE , 2023, 39(6):1437-1457.
- [25] Mohaned A, Reda M, Karam M S, Ripon K C. Light spectrum optimizer: a novel physics-inspired metaheuristic optimisation algorithm[J]. Mathematics, 2022, 10: 3466.

Optimizing Decentralized Exam Timetabling with a Discrete Whale Optimization Algorithm

Emily Sing Kiang Siew¹, San Nah Sze², Say Leng Goh³

i-CATS University College, Kuching, Sarawak 93350, Malaysia¹

Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak 94300, Malaysia²

Optimization and Visual Analytics Research Group, Faculty of Computing and Informatics, Universiti Malaysia Sabah, Labuan International Campus, Labuan 87000, Malaysia³

Abstract—In recent years, there has been increasing interest in intelligent optimization algorithms, such as the Whale Optimization Algorithm (WOA). Initially proposed for continuous domains, WOA mimics the hunting behavior of humpback whales and has been adapted for discrete domains through modifications. This paper presents a novel discrete Whale Optimization Algorithm approach, integrating the strengths of population-based and local-search algorithms for addressing the examination timetabling problem, a significant challenge many educational institutions face. This problem remains an active area of research and, to the authors' knowledge, has not been adequately addressed by the WOA algorithm. The method was evaluated using real-world data from the first semester of 2023/2024 for faculties at the Universiti of Sarawak, Malaysia. The problem incorporates standard and faculty-specified constraints commonly encountered in real-world scenarios, accommodating online and physical assessments. These constraints include resource utilization, exam spread, splitting exams for shared and non-shared rooms, and period preferences, effectively addressing the diverse requirements of faculties. The proposed method begins by generating an initial solution using a constructive heuristic. Then, several search methods were employed for comparison during the improvement phase, including three Variable Neighborhood Descent (VND) variations and two modified WOA algorithms employing five distinct neighborhoods. These methods have been rigorously tested and compared against proprietary heuristic-based software and manual methods. Among all approaches, the WOA integrated with the iterative threshold-based VND approach outperforms the others. Furthermore, a comparative analysis of the current decentralized approach, decentralized with re-optimization, and centralized approaches underscores the advantages of centralized scheduling in enhancing performance and adaptability.

Keywords—Examination timetabling; discrete whale optimization algorithm; variable neighborhood descent; capacitated; decentralized

I. INTRODUCTION

Educational timetabling involves assigning specific times to resources, events, and spaces while adhering to a predefined set of hard constraints and optimizing soft constraints. Resources typically encompass lecturers, teachers, students, administrative staff, or specialized equipment. Events may include lectures, classes, exams, or other academic activities.

Spaces refer to physical locations such as lecture halls, classrooms, or exam rooms.

Numerous formulations have been proposed for this problem, with the two most notable being the uncapacitated formulation introduced by [1] and the capacitated formulation featured as Exam 1 in the Second International Timetabling Competition, discussed by [2]. This study addresses a capacitated formulation of a real-world faculty exam timetabling problem (ETP) at the Universiti of Sarawak, Malaysia (UNIMAS). This problem stands out due to its unique combination of two approaches: online exam scheduling, which solely considers designated periods without considering physical room allocation, and physical exam scheduling, which involves assigning each exam to a specific period and room. Both exam scheduling strategies aim to prevent conflicts and optimize exam spacing, but the latter necessitates meeting room allocation constraints, such as dividing or sharing spaces.

Since ETP is an NP-complete decision problem [3], diverse approaches have been employed to address it. According to a recent study [4], there are six types of solution methods in the ETP. These are mathematical optimization, metaheuristics, heuristics, metaheuristics, hyper-heuristics, and hybrid approaches. The survey found that metaheuristics had been the approach most employed over the past 12 years.

Metaheuristics generally outperform exact search methods, as the latter often involves generating all possible solutions, which can be computationally intensive. Metaheuristic algorithms can be broadly divided into two categories: population-based algorithms, which emphasize exploration, and single-solution-based algorithms, which focus on exploitation. Effective metaheuristic design requires balancing two criteria: diversification, which involves exploring the search space broadly, and intensification, which focuses on refining and exploiting the most promising solutions [5].

An effective way to balance exploration and exploitation is by using a hybrid approach that integrates various techniques to enhance the performance of search algorithms. In this study, we introduce and design a novel hybrid method that combines the recently developed Whale Optimization Algorithm (WOA) with local search techniques to solve a real-world ETP. The following outline summarizes the main contributions of this work.

- Discrete WOA algorithm: A solution methodology that combines the WOA algorithm with local search methods is developed. This approach performs better than other VND variants in optimizing exam timetabling.
- Decentralized faculty exam timetabling: We propose a novel model that accommodates the preferences of multiple faculties with contradictory constraints, accounts for varied exam types, and ensures a more inclusive and flexible scheduling framework.
- Utilization of real-world data: The proposed discrete WOA approach is validated using real-world data, showcasing its robustness and practical relevance across various educational settings.

The paper is structured as follows: Section II presents a review of related works, followed by Section III, which outlines the problem description. Section IV discusses the original WOA, other applied methods and neighborhood structures. Section V details the algorithms of the proposed discrete WOA approaches. Section VI presents the experimental results, and Section VII compares the centralized and decentralized approaches. Lastly, Section VIII provides the conclusion of the study.

II. RELATED WORK

Real-world exam timetabling constraints are categorized into four main types [4]: exam-related, period-related, room-related, and invigilator-related. Real-world scenarios more commonly give rise to the capacitated formulation of ETPs, treating room capacities as adhered-to constraints. The constraints on room usage can vary significantly across problem formulations, ranging from limits on the number of exams allowed per room to considerations of individual room capacities and overall seating availability. Some studies extend this by considering the total seating capacity across all rooms within a time slot and the capacities of individual rooms [6–8]. In such cases, several exams may be assigned in the same room without restrictions on the number of exams if the total room capacity is sufficient to accommodate all the students requiring seating.

Dammak et al. [9] proposed a heuristic algorithm that modeled the exam-room assignment problem, allowing multiple exams in a single room. In contrast, other studies have also explored the possibility of scheduling multiple exams in a single room [10, 11]. Other room-related constraints studied include the distance between exam halls [12, 13], the allocation of exams across multiple rooms [12, 14] and assigning specific exams to designated rooms. We incorporate all these constraints—one exam per room, multiple exams per room, splitting exams across multiple rooms, and distance between rooms for split exams—into our approach on a faculty-specific basis.

Researchers have recently designed many intelligent algorithms, such as the Archimedes optimization algorithm [15], Fire Hawks algorithm [16], and WOA, to address various optimization challenges. Notably, the WOA, a swarm intelligence-based approach [17], models the hunting strategies

of humpback whales, mimicking their collective feeding behavior. Recent studies have enriched the growing literature by highlighting its successful practical applications and reporting enhanced results and performance [18]. Additionally, research suggests that WOA surpasses other optimization algorithms concerning global search capabilities and convergence speed [19]. The WOA offers several advantages, including simplicity of operation, minimal control parameters, and a robust capability to avoid local optima. These attributes have inspired researchers to employ WOA to address diverse practical challenges.

Although the WOA was originally developed for continuous problems, several studies have explored the use of the WOA for discrete optimization problems, including the knapsack problem [20, 21], feature selection [22–24], and workshop scheduling [25–27]. The primary strength of the WOA lies in its ability to maintain a balance between exploration and exploitation throughout the iterative process. While WOA has shown promise in various optimization problems across multiple domains, to the best of our knowledge, its application to real-world exam timetabling remains unexplored.

Hence, this study bridges this gap by adapting the WOA approach to meet the specific needs of our real-world ETP, offering a novel solution for discrete optimization in educational scheduling. Since WOA was originally designed for continuous optimization tasks, it relies on continuous updates to individual positions, making it unsuitable for discrete scheduling problems like exam timetabling. To overcome this limitation, we propose a modified discrete WOA, incorporating discrete updating strategies to tailor the algorithm to the discrete nature of timetabling. The real-world experimentation outlined in the following sections demonstrates its effectiveness in addressing practical exam timetabling while fulfilling institutional constraints.

III. PROBLEM DEFINITION

This paper presents a solution method for the ETP at UNIMAS. Specifically, we study the decentralized ETP within the Faculty of Economics and Business (FEB) and the Faculty of Computer Science and Information Technology (FCSIT). The data on students within the faculties has been collected and analyzed to evaluate the proposed solution algorithm.

The problem definitions are as follows:

- The exams will take place over two weeks.
- Each day is divided into two blocks.
- Exams have varying durations, such as 120, 150, and 180 minutes.
- If assigning an exam to a single room within a timeslot is not feasible, the exam must be split across multiple rooms.
- The exam day reserved for pre-assigned common courses should not be used for other exams.
- The exams include online exams conducted via an online platform and physical exams held in rooms.

- Online exams must be assigned to a predesignated slot.
- Shared or non-shared exam rooms will depend on each faculty's specific practices.
- There are two types of exam rooms: exam halls and faculty-owned exam rooms.
- Exam halls vary in availability based on the schedule set by the Centre, which range in size from medium to large.
- Faculty-owned exam rooms are consistently available for faculty exams and are typically small-sized.

The hard constraints are:

- H1: Each student may attend only one exam at any given time.
- H2: Each exam can only be scheduled once within the exam period.
- H3: The exam period must not exceed the designated days.
- H4: Rooms must be able to accommodate all students taking an exam during each timeslot.
- H5: Rooms can only be shared if the faculty permits; otherwise, no sharing is allowed.

The soft constraints include:

- S1: Minimize the number of rooms utilized.
- S2: Minimize proximity costs to ensure adequate time gaps between exams.
- S3: Minimize the splitting of exams across different areas or rooms.
- S4: Minimize violations of exams assigned to preferred timeslots.

A solution that violates soft constraints is not considered infeasible; this allows for defining specific objective values for each soft constraint. Consequently, the objective function f aims to minimize the total soft constraint violations, directing the optimization process toward reducing their overall impact. The end user typically determines the weights assigned to different types of soft conflicts. However, this study standardizes the weights by assigning fixed values: 1 for $S1$, $S2$, $S3$, and 2 for $S4$ to ensure reproducibility.

IV. METHODS

A. Constructive Heuristic Method

The proposed algorithm starts by generating initial feasible solutions, using a constructive heuristic method as the starting point. The process begins with assigning prioritized exams to their preferred time slots, followed by the allocation of online exams and, finally, the allocation of physical exams. We use a best-fit strategy for room allocation, choosing the smallest room that fits, minimizing room splits, and allocating several exams to the same room whenever feasible. The algorithm continues to allocate exams to rooms and periods while

ensuring compliance with hard constraints and adhering to the soft constraint of preferred slot assignments.

B. The Original WOA

The WOA is a recently developed swarm intelligence optimization algorithm commonly used to solve optimization and classical engineering problems. When whales locate prey, they swim in a spiral toward it while encircling and foraging using a bubble net. This process involves three hunting strategies: shrinking and attacking with a bubble-net attack, encircling prey, and randomly searching for prey. The first two strategies guide exploitation, while exploration is supported by the third within the WOA.

We present the mathematical model for each phase below, employing a uniform distribution to generate random numbers in the equations. In the following equations, t signifies the current iteration, x refers to the position vector, and $MaxIter$ indicates the maximum number of permitted iterations.

1) *Exploitation phase – encircling prey*: Humpback whales employ strategies described by the mathematical models in Eq. (1) and Eq. (2) to encircle and hunt their prey. As per Eq. (2), acting as search agents, whales adjust their positions relative to the prey—the current optimal solution, x^* . The coefficient vectors C and A , calculated using Eq. (3) and Eq. (4), adjust the search area to determine the whale's position relative to its prey. In both phases, the value of a decreases linearly from 2 to 0, while the vector r exhibits a uniform distribution within the interval $[0,1]$.

$$D = |C \cdot x^*(t) - x(t)| \quad (1)$$

$$x(t+1) = x^*(t) - A \cdot D \quad (2)$$

$$A = 2ar + a \quad (3)$$

$$C = 2r \quad (4)$$

2) *Exploitation phase – bubble-net attacking*: The shrinking, encircling behavior is governed by Eq. (5), while the position of a neighboring search agent is determined using a spiral equation as described in Eq. (6). D' denotes the distance from the i -th whale to the optimal solution, with b defining the shape of the logarithmic spiral and l being a random value within the range $[-1, 1]$.

$$a = 2 - t \cdot (2 / MaxIter) \quad (5)$$

$$x(t+1) = D' \cdot e^{bl} \cdot \cos(2\pi l) + x^*(t) \quad (6)$$

3) *Exploration phase – searching for prey*: For exploration, a random search agent is selected to guide the process, as mathematically represented by Eq. (7) and Eq. (8). Vector A contains random values exceeding one or falling below -1, while x_{rand} represents a randomly chosen whale from the population.

$$D = |C \cdot x_{rand} - x| \quad (7)$$

$$x(t+1) = |x_{rand} - A \cdot D| \quad (8)$$

Algorithm 1 delineates pseudocode for the original WOA, which starts by generating an initial population and evaluating it with a fitness function. During each iteration, a random value determines the update of a solution's position using either Eq. (2), Eq. (8), or Eq. (6) methods. The system returns the best solution x^* upon meeting the termination criteria.

Algorithm 1: Original WOA

```
Generate initial population  $x_i$  for  $i = 1, 2, \dots, n$ 
Compute each solution's fitness
Set the best search solution  $x^*$ 
 $t = 0$ 
While ( $t < MaxIter$ ) do
  For each solution do
    Update  $C, A, p, a$ , and  $l$ 
    If  $p < 0.5$  then
      If  $|A| < 1$  then
        Update the current solution's position by (2)
      Else
        Update the current solution's position by (8)
      End If
    Else
      Update the current solution's position by (5)
    End If
  End For
  Verify if any solution goes beyond the search space and amend it
  Compute each solution's fitness
   $t = t + 1$ 
  Update  $x^*$  if a better solution is found
End While
return  $x^*$ 
```

C. Variable Neighborhood Descent

Exploring a single neighborhood structure may result in finding a local optimum specific to that structure, but this is unlikely to be the global optimum. Conversely, identifying a solution that serves as a local optimum across multiple neighborhood structures enhances the likelihood of reaching the global optimum. This principle forms the foundation of the VND method. Specifically, a VND algorithm is employed to refine the solutions. VND is a deterministic variation of the Variable Neighborhood Descent framework initially proposed by [28]. It has been widely adopted as a local search method in numerous metaheuristics and implemented in diverse forms [29]. During its process, VND systematically explores different neighborhoods of a given solution to enhance its quality.

Algorithm 2 provides the pseudocode for the VND. The algorithm explores the neighborhood structures defined by the operators N_k , where $1 \leq k \leq k_{max}$, following a predefined order. $LocalSearch(x, N_k)$ indicates that a local search is performed

using the current neighborhood N_k , starting from solution x . The first-improvement strategy is applied to all neighborhood structures, as described in Subsection D. Specifically, when an improved solution is found within a specific neighborhood, the corresponding move is made, and the next neighborhood structure is explored. This procedure repeats until the maximum iteration limit is reached.

VND can employ various rules to transition between neighborhoods on its list and adopt diverse strategies to explore each. This flexibility gives rise to multiple VND variants, including Basic Sequential VND (BVND), Cyclic VND, Pipe VND (PVND), and Nested VND. These variants may utilize either the first-improvement or best-improvement search strategies. Algorithms 3 and 4 outline the neighborhood change procedure for both BVND and PVND, respectively. For the former, if a better candidate solution is found within a given neighborhood structure, the search resumes in the initial neighborhood structure based on the specified order. Otherwise, the search continues in the next neighborhood structure. For the latter, if the current solution improves within a particular neighborhood, exploration continues within that neighborhood.

Algorithm 2: Variable Neighborhood Descent

```
Procedure VND ( $x, N$ )
While ( $t < MaxIter$ ) do
   $stop = false$ 
   $k = 1$ 
   $x' = x$ 
  While ( $k \leq k_{max}$ )
     $x'' = LocalSearch(x, N_k)$ 
    neighborhood_change ( $x, x'', k$ )
  End While
End While
return  $x'$ 
```

Algorithm 3: Sequential Neighborhood Change for BVND

```
Procedure Sequential_neighborhood_change ( $x, x', k$ )
If  $f(x') < f(x)$  then
   $x = x'$ 
   $k = 1$ 
Else
   $k = k + 1$ 
End
```

Algorithm 4: Pipe Neighborhood Change for PVND

```
Procedure Pipe_neighborhood_change ( $x, x', k$ )
If  $f(x') < f(x)$  then
   $x = x'$ 
Else
   $k = k + 1$ 
End
```

D. Neighborhood Structure

The five types of neighborhood structures are implemented and described in the following:

- *Kick*: Assign exam e_1 to the period currently designated to exam e_2 , then reassign exam e_2 to a different period from the available options. The room for both exams may be available within the designated period.
- *Swap*: Exchange the periods of two exams, while their rooms may be swapped or assigned to different rooms.
- *Shift*: Move an exam to a different period and/or room(s).
- *Reallocate*: Move an exam assigned to a shared room to an unoccupied one.
- *Compact*: Relocate an exam to a shared room during the same period.

Swap, *Shift*, and *Kick* are moves related to both period and room assignments, whereas *Reallocate* and *Compact* focus specifically on room-related adjustments with contradictory objectives. *Reallocate* aims to address room-sharing preferences, especially in cases where certain faculties prohibit sharing. In contrast, *Compact*, which applies to most faculties, aims to reduce the number of rooms utilized, thereby encouraging room sharing. The neighborhood structure is set to $N = \{Swap, Shift, Kick, Reallocate\}$ for faculty exam timetabling where shared exam rooms are prohibited. Otherwise, the neighborhood structure is set to $N = \{Swap, Shift, Kick, Compact\}$.

V. PROPOSED APPROACH

This study proposes two algorithms: one based on the PVND algorithm and the other on WOA. Subsection A describes the first algorithm, Iterative Threshold-based Variable Neighborhood Descent (ITVND), while Subsection B presents the second algorithm, the modified discrete WOA.

A. Iterative Threshold-based Variable Neighborhood Descent

We propose a variant of the classic PVND, ITVND, with the pseudocode presented in Algorithm 5. The threshold-based pipe neighborhood change procedure is outlined in Algorithm 6. The ITVND algorithm incorporates a control parameter, the objective function threshold cT , and an input parameter, the iteration count L . We initially assign the objective value of the starting solution to cT . The iteration count increments in steps of L , and at every L -th iteration ($cT \bmod L = 0$), cT is updated to the current cost. The algorithm accepts all improving or sideways moves with an objective value below cT .

The algorithm explores the solution space using multiple neighborhood structures, where the neighborhood structure N_k is defined for $k = 1 \dots, k_{max}$. The core concept of ITVND is to maintain the objective value threshold across L iterations for various neighborhood structures. It permits accepting inferior solutions within the objective value threshold, introducing a more flexible acceptance condition, which slows the current cost reduction and prolongs the time needed to reach convergence.

Algorithm 5: Iterative Threshold-based Pipe Variable Neighborhood Descent

Procedure ITVND (x, N)

$t = 0$

While ($t < MaxIter$) do

$k = 1$

$x' = x$

 While ($k \leq k_{max}$)

$x'' = LocalSearch(x, N_k)$

 Threshold_pipe_neighborhood_change (x, x'', k, cT)

 End While

 If ($t \bmod L = 0$) then

$cT = f(x)$

 End If

$t = t + 1$

End While

return x'

Algorithm 6: Threshold Pipe Neighborhood Change

Procedure Threshold-based_pipe_neighborhood_change (x, x', k, cT)

If $f(x') < cT$ or $f(x') < f(x)$ then

$x = x'$

Else

$k = k + 1$

End

B. Discrete Whale Optimization Algorithm

As whales adjust their positions within a continuous domain using specific equations and operators, the original WOA becomes unsuitable for tasks like timetabling, which exhibit discrete characteristics. While the classical WOA algorithm relies on whale interactions to solve optimization problems, its simple neighborhood structure and limited disturbance tend to trap it in local optima. To address this problem, we replaced these equations and operators with a local search mechanism. The proposed modified discrete WOA approach considers two methods: (1) WOA-VD and (2) WOA-IVD. As shown in Algorithm 7, the process begins with a set of solutions generated using a constructive heuristic, then iteratively refined throughout the search process.

During the exploitation phase, the modified discrete WOA algorithm updates individual solutions using information from the best current solution. When probability p is less than 0.5, the algorithm utilizes local search with an improvement criterion, accepting a new solution only if it enhances the current solution's fitness, thereby supplanting Eq. (2). If p is equal to or greater than 0.5, PVND is chosen as a local search method to improve the solution in WOA-VD algorithm. In contrast, the ITVND local search method is used in the WOA-IVD algorithm to compare the effectiveness of the VND variation during the search process. Both methods replace Eq. (5) and take advantage of its ability to search for a larger area by changing neighborhoods in a planned way.

During exploration, local search with a threshold-based acceptance criterion replaces Eq. (8). A new candidate solution is accepted if its objective value falls below a dynamically updated cost bound. This approach helps explore the search space more effectively by mitigating the effects of premature

convergence and promoting broader exploration outside the immediate neighborhood of the current optimal solution.

Algorithm 7: Discrete WOA-VD

```

Generate initial population  $x_i$  for  $i = 1, 2, \dots, n$ 
Compute each solution's objective value
Initialize the best search solution  $x^*$ 
 $t = 0$ 
While ( $t < MaxIter$ ) do
    For each solution individual do
        Update  $a, A, C, l$  and  $p$ 
        If  $p < 0.5$  then
            If  $|A| < 1$  then
                Generate new candidate  $x'$ 
                If  $f(x') < f(x)$  then
                     $x = x'$ 
                End If
            Else
                Generate new candidate  $x'$ 
                If  $f(x') < cBound$  then
                     $x = x'$ 
                End If
                If  $(t \bmod L = 0)$  then
                     $cBound = f(x)$ 
                End If
            End If
        Else
            Apply PVND ( $x, N$ ) to improve  $x$ 
        End If
    End For
    Calculate each solution's objective value
     $t = t + 1$ 
    Update  $x^*$  if a better solution is found
End While
return  $x^*$ 

```

Our search method is based on local search and leverages the following characteristics:

- *Termination Criterion:* A fixed number of iterations as the termination criterion ensures that the algorithm's runtime remains consistent and independent of other parameters.
- *Search Space:* The search space is restricted to the feasible region, which only includes scheduling that fully satisfies all constraints, such as precedence and conflict avoidance.

VI. EXPERIMENTS

A. Experimental Settings

The experiment is conducted on a computer with Windows 11, equipped with an Intel® Core™ i7 processor, 16.0 GB of memory, and an integrated graphics card. We built the algorithms presented using IntelliJ IDEA, the Integrated Development Environment (IDE), and Java, specifically JDK

1.8, as the programming language. Table I presents the characteristics of the datasets collected from two faculties. These datasets specifically pertain to the first semester of the 2023/2024 academic year.

TABLE I. CHARACTERISTICS OF DATASETS FROM TWO FACULTIES

Description	Dataset	
	FCSIT	FEB
No. of enrolment	2,837	7,035
No. of students	1,026	2,011
No. of exams	31	55
Exam range per student	1 to 5	1 to 8
Conflict density	0.234	0.182
Exam Types	Online & Physical	Physical
Total exam period	12 days	12 days
Average Shared Hall Usage per Timeslot	1.8	3.1
Faculty-Owned Exam Room Count	10	6
Scheduling method	Manually arranged	Proprietary System

To evaluate the performance of our proposed algorithms, we compared them against other methods and the existing scheduling method, which generates the current solution. The five include three variations of VND: BVND, PVND, and ITVND, along with two variations of WOA: WOA-VD and WOA-IVD. For each of these algorithms, we conducted 30 independent runs per dataset.

B. Experimental Results

Table II below presents the descriptive statistics for the three employed algorithms compared to the existing scheduling method based on 30 runs. The bold formatting indicates the best values achieved by all the methods. We conducted all statistical analyses using SPSS version 29.

TABLE II. COMPARISON BETWEEN DIFFERENT METHODS

Dataset	Method	Slot	Best	Mean	Worst	Std Dev
FCSIT	Manual	12	55.6			
	BVND	8	50.4	53.7	57.3	1.97
	PVND	8	50.1	53.1	56.3	1.48
	ITVND	8	49.4	51.6	54.7	1.41
	WOA-VD	8	49.9	52.9	57.8	2.09
	WOA-IVD	8	48.7	51.2	53.7	1.28
FEB	Proprietary System	12	148.6			
	BVND	12	131.8	134.6	138.3	1.77
	PVND	12	132.2	134.7	137.0	1.24
	ITVND	12	130.3	132.4	134.6	1.22
	WOA-VD	12	131.0	132.9	135.3	1.13
	WOA-IVD	12	129.9	131.9	133.6	1.04

For the FCSIT dataset, Table II shows that WOA-IVD emerges as the best-performing method among the tested methods, with the lowest objective value (48.7) and the lowest standard deviation (1.28), demonstrating consistent performance. The second-best method is ITVND, which has a slightly higher objective value (49.4) but maintains low variability (1.41). In contrast, despite its competitive mean value, the WOA-VD method has the highest variability (2.09), indicating less consistent results than others.

In comparing the performance of various methods for the FEB dataset, all our proposed methods outperform the proprietary system [30], which employed two-stage heuristic methods across all instances, with a moderate gap. The best-performing method, WOA-IVD, achieved an average of 131.9 with a standard deviation 1.04. ITVND followed closely with an average value of 132.4 and a standard deviation 1.22. However, the results for BVND and PVND are somewhat inferior, with some instances where they perform poorly compared to the ITVND and WOA models.

Overall, the results demonstrate that adopting all the VND and WOA variation methods reduces exam session and objective value compared to the manual approach. The WOA-IVD method performs the best across both datasets, consistently achieving the lowest mean values and demonstrating superior efficiency compared to other methods. On the other hand, VND tends to perform the worst, showing higher values in comparison.

We conducted a one-way ANOVA analysis of variance, as presented in Table III for dataset FCSIT and Table IV for dataset FEB, to statistically demonstrate the differences among all employed methods, BVND, PVND, ITVND, WOA-VD, and WOA-IVD, presented in Table II. The results for both datasets indicate a statistically significant difference between the tested approaches, with a p-value below 0.001.

TABLE III. ANOVA FOR FCSIT DATASET OF THE ALGORITHMS

Source	DF	Sum of Squares	Mean Square	F	Signature
Between Groups	4	132.209	33.052	11.772	<.001
Within Groups	145	407.131	2.808		
Total	149	539.340			

TABLE IV. ANOVA FOR FEB DATASET OF THE ALGORITHMS

Source	DF	Sum of Squares	Mean Square	F	Signature
Between Groups	4	198.958	49.739	29.289	<.001
Within Groups	145	246.247	1.698		
Total	149	445.204			

Fig. 1 and 2 present the box plots for two datasets generated using Tableau software. The lower mean, median, and distribution values in both box plots show that WOA-IVD is consistently the best method. The fact that its box plot height is lower than other algorithms in both datasets further

demonstrates this. For the FCSIT dataset, as shown in the box plot in Fig. 1, WOA-VD performs the worst due to its higher spread and maximum values, suggesting poor performance in worst-case scenarios. For the FEB dataset, as shown in the box plot in Fig. 2, BVND has the largest spread and includes higher maximum values, indicating it performs less reliably and worse in the worst-case scenarios. The proposed WOA-IVD approach effectively balances the objectives of exploration and exploitation, as the tested approaches' performance ranks BVND, PVND, WOA-VD, ITVND, and WOA-IVD.

The results highlight the consistent effectiveness of our discrete WOA method, achieving an objective value reduction of approximately 12.41% for the FCSIT dataset and 12.58% for the FEB dataset. This consistency underscores the method's reliability and adaptability, making it a practical solution for diverse scenarios. However, the proposed discrete modified WOA may be constrained by the need to adapt the local search method in the WOA model, which is currently tailored to our ETP problem instance and might require modification to address different problem constraints or domains.

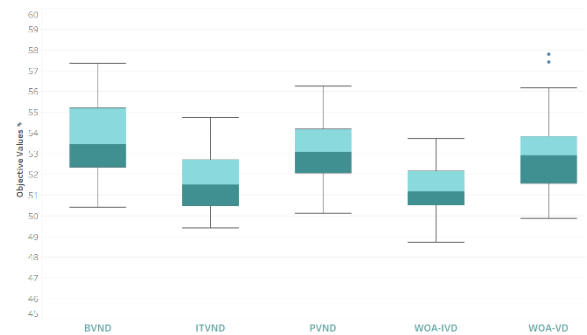


Fig. 1. Box plots of objective values for FCSIT dataset.

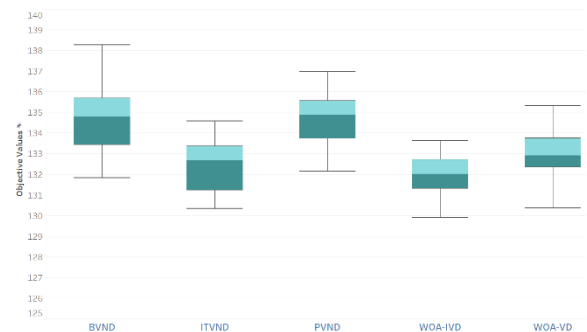


Fig. 2. Box plots of objective values for FEB dataset.

VII. CENTRALIZED AND DECENTRALIZED

A comparison of centralized and decentralized approaches to university exam timetabling during the pandemic was conducted by Modirghorassani and Hoseinpour [31], focusing on minimizing costs and ensuring social distancing, with the study underscoring the advantages of decentralization. Building on this and given that our current practice employs a decentralized approach at the faculty level, we intend to conduct a similar comparison using our objective function for cost evaluation to better align with our specific operational context. A comparative analysis is conducted to evaluate the effects of three scheduling methods:

1) *Decentralized approach*: the current practice whereby faculty schedule their timetables independently.

2) *Centralized approach*: All faculty exams are scheduled, with resources managed centrally. It employs a uniform 12-timeslot structure, ensuring consistency in completing exam sessions across datasets.

3) *Decentralized approach with re-optimization*: Similar to the first approach, only that re-optimization is performed after post-resource reallocation. If a faculty's exam session concludes earlier than others, the shared resources are reallocated for use by other faculties that take longer exam sessions.

The comparative analysis aims to determine the relative impacts of these approaches on scheduling efficiency and overall outcomes. The centralized and decentralized approaches are compared in Table V, with and without the re-optimization strategy for the decentralized approach. The comparison is made across four soft constraints (*S1–S4*) based on their objective values. The bold formatting indicates the best values achieved by the approaches.

TABLE V. PERFORMANCE OF SCHEDULING UNDER CENTRALIZED AND DECENTRALIZED APPROACHES

Value	Decentralized		Centralized
	Without Re-optimization	With Re-optimization	
S1	92.0	82.3	86.0
S2	37.2	43.9	19.5
S3	33.9	25.4	29.1
S4	16.0	20.4	9.8
Total	179.3	172.0	144.5

The decentralized approach with re-optimization has shown superior efficiency in resource allocation, as evidenced by its lowest values in both *S1* and *S3*, which pertain to room usage and splitting. However, it is less effective for other constraints, such as *S2* and *S4*, which are related to spread and preferred slot, where it performs worse than the decentralized approach without re-optimization. While re-optimization enhances performance in certain aspects, it may not universally improve outcomes across all constraints. Despite this, the centralized approach remains the most efficient overall.

For the decentralized approach, values without re-optimization are generally higher, indicating that re-optimization improves efficiency. In contrast, the centralized approach consistently produces lower values than both decentralized approaches, underscoring its overall efficiency. The total objective values further support this trend, with the centralized approach achieving the lowest total (144.5), followed by the decentralized approach with re-optimization (172.0) and without re-optimization (179.3). It shows a reduction of approximately 15.9% in the total objective value, calculated as $(172.0 - 144.5) / 172.0 \times 100$. This improvement highlights the effectiveness of the centralized approach in enhancing the solution's overall quality. However, the comparative analysis of centralized and decentralized

approaches is incomplete, as it encompasses only a limited subset of faculties rather than the entire scope, thereby leaving this aspect open for further exploration.

VIII. CONCLUSION

This study examines decentralized faculty exam timetabling to optimize resource allocation and satisfy institutional constraints while designing an approach that can be adopted across multiple faculties. The ETP under consideration accommodates two distinct exam modes and formulations within the same timetable. This structure is notably different from those commonly found in literature and, to our knowledge, has not been previously studied. We propose two approaches: ITVND, an improved version of VND, and a novel discrete WOA. Specifically, we embedded the different local search strategies in the WOA algorithm to ensure they work well in the discrete scheduling domain. We used a real-world dataset to validate the proposed algorithm's practicality, highlighting its applicability across various faculties while adhering to their specific constraints. Our search methods have been rigorously tested and compared internally and against proprietary software developed using heuristic and manual methods. These comparisons highlight that the discrete WOA outperforms other approaches, demonstrating superior performance, though it takes slightly longer. While the preliminary results provide proof of concept, further experimentation with additional examination timetabling datasets, such as benchmark sets, could provide valuable insights. We consider hybridizing the WOA algorithm with other metaheuristic algorithms for future studies.

ACKNOWLEDGMENT

This work was funded by *i*-CATS University College under the *i*-CATS Research and Innovation Grant Scheme 2024.

REFERENCES

- [1] Carter, Michael W., Gilbert Laporte, and Sau Yan Lee, "Examination timetabling: Algorithmic strategies and applications," *Journal of the operational research society* 47.3, 1996, pp.373-383. doi: 10.2307/3010580.
- [2] McCollum B, McMullan P, Parkes AJ, Burke EK, Qu R, "A New Model for Automated Examination Timetabling," *Annals of Operations Research* 194:291–315, 2012, doi: 10.1007/s10479-011-0997-x.
- [3] de Werra D, "The combinatorics of timetabling," *European Journal of Operational Research* 96(3):504–513, 1997, doi: 10.1016/S0377-2217(96)00111-7.
- [4] E.G. Talbi, "Metaheuristics: From Design to Implementation," Wiley Online Library, 2009.
- [5] Burke, E., Bykov, Y., Newall, W., Petrovic, S., "A time-predefined local search approach to exam timetabling problems," *IIE Trans.* 36 (6), pp. 509–528, 2004, doi: 10.1080/07408170490438410.
- [6] Caramia, M., Dell'Olmo, P., "Coupling stochastic and deterministic local search in exam timetabling," *Oper. Res.* 55 (2), pp. 351–366, 2007, doi: 10.1287/opre.1060.0354.
- [7] Cheong, C.Y., Tan, K.C., Veeravalli, B., "A multi-objective evolutionary algorithm for exam timetabling," *J. Sched.* 12, pp. 121–146, 2007, doi: 10.1007/s10951-008-0085-5.
- [8] Dammak, A., Elloumi, A., Kamoun, H., "Classroom assignment for exam timetabling," *Adv. Eng. Softw.* 37, pp. 659–666, 2006, doi: 10.1016/j.advengsoft.2006.02.001.
- [9] Akbulut, A., Yilmaz, G., "University exam scheduling system using graph coloring algorithm and rfid technology," *Int. J. Innov. Manag. Technol.* 4 (1), 66, 2013, doi: 10.7763/IJIMT.2013.V4.359.

- [10] Hassan, Mohammad Al-Haj, Osama Al-Haj Hassan, "Constraints aware and user friendly exam scheduling system," *Int. Arab J. Inf. Technol.* 13 (1A), pp. 156–162, 2016.
- [11] Kahar, M.N.M., Kendall, G., "The exam timetabling problem at Universiti Malaysia Pahang: Comparison of a constructive heuristic with an existing software solution," *Eur. J. Oper. Res.* 207 (2), pp. 557–565, 2010, doi: 10.1016/j.ejor.2010.04.011.
- [12] Cataldo, A., Ferrer, J. C., Miranda, J., Rey, P. A., & Sauré, A., "An integer programming approach to curriculum-based examination timetabling," *Ann Oper Res* 258, pp. 369–393, 2017, doi: 10.1007/s10479-016-2321-2.
- [13] Ozturk, Z. K., Gundogan, H. S., Mumykmaz, E., & Kececioğlu, T., "Exam scheduling under pandemic conditions: A mathematical model and decision support system," *Technological Forecasting and Social Change*, 208, 123687, 2024, doi: 10.1016/j.techfore.2024.123687.
- [14] S. Mirjalili, A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.* 95, pp. 51–67, 2016, doi: 10.1016/j.advengsoft.2016.01.008.
- [15] Hashim, F. A., Hussain, K., Houssein, E. H., Mabrouk, M. S., & Al-Atabany, W., "Archimedes optimization algorithm: a new metaheuristic algorithm for solving optimization problems," *Applied intelligence*, 51, 1531–1551, 2021, doi:10.1007/s10489-020-01893-z.
- [16] Ashraf, A., Anwaar, A., Haider Bangyal, W., Shakir, R., Ur Rehman, N., Qingjie, Z., "An Improved Fire Hawks Optimizer for Function Optimization," In: Tan, Y., Shi, Y., Luo, W. (eds) *Advances in Swarm Intelligence. ICSI 2023. Lecture Notes in Computer Science*, vol 13968. Springer, Cham, 2023, doi:10.1007/978-3-031-36622-2_6.
- [17] Mahmood, S., Bawany, N.Z., Tanweer, M.R., "A comprehensive survey of whale optimization algorithm: modifications and classification," *Indones. J. Electr. Eng. Comput. Sci.* 29 (2), 899, 2023, doi: 10.11591/ijeecs.v29.i2.pp899-910.
- [18] Nadimi-Shahraki, M.H., Zamani, H., Asghari Varzaneh, Z., Mirjalili, S., "A Systematic Review of the Whale Optimization Algorithm: Theoretical Foundation, Improvements, and Hybridizations," *Arch. Comput. Methods Eng.* 30, pp. 4113–4159, 2023, doi: 10.1007/s11831-023-09928-7.
- [19] Abdel-Basset, M., El-Shahat, D. & Sangaiah, "A modified nature inspired meta-heuristic whale optimization algorithm for solving 0–1 knapsack problem," *Int. J. Mach. Learn. & Cyber.* 10, pp. 495–514, 2019, doi: 10.1007/s13042-017-0731-3.
- [20] Abdel-Basset, M., El-Shahat, D. & Sangaiah, "A modified nature inspired meta-heuristic whale optimization algorithm for solving 0–1 knapsack problem," *Int. J. Mach. Learn. & Cyber.* 10, 495–514 (2019). <https://doi.org/10.1007/s13042-017-0731-3>.
- [21] Li, Y., He, Y., Liu, X. et al., "A novel discrete whale optimization algorithm for solving knapsack problems," *Appl Intell* 50, pp. 3350–3366, 2020, doi: 10.1007/s10489-020-01722-3.
- [22] Majdi M. Mafarja, Seyedali Mirjalili, "Hybrid Whale Optimization Algorithm with simulated annealing for feature selection," *Neurocomputing*, 260, pp. 302–312, 2017, doi: 10.1016/j.neucom.2017.04.053.
- [23] Tawhid, M.A., Ibrahim, A.M., "Feature selection based on rough set approach, wrapper approach, and binary whale optimization algorithm," *Int. J. Mach. Learn. & Cyber.* 11, pp. 573–602, 2020, doi: 10.1007/s13042-019-00996-5.
- [24] Mafarja, M., Thaher, T., Al-Betar, M.A., Too, J., Awadallah, M.A., Abu Doush, I. and Turabieh, H., "Classification framework for faulty-software using enhanced exploratory whale optimizer-based feature selection scheme and random forest ensemble learning," *Appl Intell* 53, pp. 18715–18757, 2023, doi:10.1007/s10489-022-04427-x.
- [25] T. Jiang, C. Zhang and Q. -M. Sun, "Green Job Shop Scheduling Problem With Discrete Whale Optimization Algorithm," in *IEEE Access*, vol. 7, pp. 43153–43166, 2019, doi: 10.1109/ACCESS.2019.2908200.
- [26] Liu, M.; Yao, X.; Li, Y., "Hybrid whale optimization algorithm enhanced with Lévy flight and differential evolution for job shop scheduling problems," *Appl. Soft Comput.*, 87, 105954, 2020, doi: 10.1016/j.asoc.2019.105954.
- [27] Zhao, F.; Xu, Z.; Bao, H.; Xu, T.; Zhu, N., "A cooperative whale optimization algorithm for energy-efficient scheduling of the distributed blocking flow-shop with sequence-dependent setup time," *Comput. Ind. Eng.*, 178, 109082, 2023, doi:10.1016/j.cie.2023.109082.
- [28] Mladenovi'c, N., Hansen, P., "Variable neighborhood search," *Comput. Oper. Res.* 24, pp. 1097–1100, 1997, doi: 10.1016/S0305-0548(97)00031-2.
- [29] Mjirda, A., Todosijevi'c, R., Hanafi, S., Hansen, P., Mladenovi'c, N., "Sequential variable neighborhood descent variants: an empirical study on the traveling salesman problem," *Int. Trans. Oper. Res.* 24, pp. 615–633, 2017, doi: 10.1111/itor.12282.
- [30] Sze, S. N., Phang, M. H., & Chiew, K. L., "Real-Life Faculty Examination Timetabling to Utilise Room Used," *Journal of Telecommunication, Electronic and Computer Engineering*, 9(3-11), pp. 51-54, 2017.
- [31] Modirghorassani, A., & Hoseinpour, P., "Decentralized exam timetabling: A solution for conducting exams during pandemics," *Socio-Economic Planning Sciences*, 92, p.101802, 2024, doi:10.1016/j.seps.2024.101802.

Fusion of Multimodal Information for Video Comment Text Sentiment Analysis Methods

Jing Han^{1*}, Jinghua Lv²

School of Creative Art and Fashion Design of Huzhou Vocational & Technical College, Huzhou 313099, Zhejiang, China¹
Department of Chinese Studies of Kyungsoong University, Pusan 48434, Pusan, Korea²

Abstract—Sentiment analysis of video comment text has important application value in modern social media and opinion management. By conducting sentiment analysis on video comments, we can better understand the emotional tendency of users, optimise content recommendation, and effectively manage public opinion, which is of great practical significance to the push of video content. Aiming at the current video comment text sentiment analysis methods problems such as understanding ambiguity, complex construction, and low accuracy. This paper proposes a sentiment analysis method based on the M-S multimodal sentiment model. Firstly, briefly describes the existing methods of video comment text sentiment analysis and their advantages and disadvantages; then it studies the key steps of multimodal sentiment analysis, and proposes a multimodal sentiment model based on the M-S multimodal sentiment model; finally, the efficiency of the experimental data from the Communist Youth League video comment text was verified through simulation experiments. The results show that the proposed model improves the accuracy and real-time performance of the prediction model, and solves the problem that the time complexity of the model is too large for practical application in the existing multimodal sentiment analysis task of the video comment text sentiment analysis method, and the interrelationships and mutual influences of the multimodal information are not considered.

Keywords—Video commentary text sentiment analysis; multimodal information fusion; M-S multimodal sentiment model; convolutional neural network

I. INTRODUCTION

Due to the prevalence of the Internet and the advancement of social media, video platforms like YouTube and Jitterbug have emerged as significant avenues for users to express their views and sentiments [1]. The massive video comment data generated on these platforms provide rich materials for sentiment analysis. The foundation for sentiment analysis of video comments is laid by the advancement of sentiment analysis technologies, particularly the breakthrough in text sentiment analysis [2]. Although sentiment analysis of video comment language has gained a lot of attention lately, there are still several issues with the current approaches. Many of the current video commentaries sentiment analysis methods mainly rely on textual modality, while ignoring the rich visual and auditory information contained in the video [3]. This unimodal approach to analysis has obvious limitations because emotional information in videos is not only expressed through text, but also conveyed through multiple channels such as facial expressions, voice intonation, and body language [4]. This process usually involves sentiment lexicon methods, which determine sentiment polarity by comparing with sentiment lexicon, and machine

learning methods, which improve the accuracy and adaptability of sentiment recognition by training machine learning models. Sentiment analysis has important applications in several fields, especially in opinion monitoring and marketing [5].

Currently Video Commentary Text Sentiment Analysis is a hot area of current research, and there are various methods and techniques, the dictionary-based method mainly relies on sentiment dictionary [6], which calculates the sentiment tendency by matching the sentiment words in the text. This method is simple and direct, but requires the support of a high-quality sentiment dictionary. Machine learning-based methods perform well on specific datasets by constructing feature vectors [7] and combining algorithms such as Support Vector Machines (SVMs) and Plain Bayes to perform sentiment classification, but this method requires the manual design of the features; deep learning methods automatically learn features through neural networks [8], which are widely used in text sentiment analysis, but the calculation is more complicated. Research indicates that multimodal sentiment analysis, which integrates visual and audio data, can markedly enhance the precision of sentiment recognition [9].

Aiming at the problems of sentiment polysemy, noise interference, dynamic change of sentiment vocabulary, cross-domain differences in sentiment, and insufficient deep semantic understanding in video comment text sentiment analysis [10], this paper proposes a M-S based multimodal sentiment model. The main main contributions of this paper are (1) analysing the main methods of video review text sentiment analysis and sorting out the related techniques; (2) designing an M-S based multimodal sentiment model; and (3) validating and analysing the evaluation model using related datasets.

II. VIDEO REVIEW OF TEXTUAL EMOTIONS ANALYSIS METHODOLOGY

At present, text emotion classification studies usually classify text emotion tendency into positive and negative. Positive emotion means that the text has a positive attitude, and negative emotion means that the text has a negative attitude. However, due to the characteristics of the Internet itself, such as openness and randomness, video comments usually contain a large number of non-standard linguistic phenomena, such as misspelled words, abbreviations, Internet slang, etc., so a large number of different sentiment analysis methods have emerged.

A. Dictionary-based Approach

The sentiment lexicon plays an important role in textual sentiment analysis by matching words with predefined

sentiment categories (e.g., positive, negative) [11], as shown in Fig. 1.

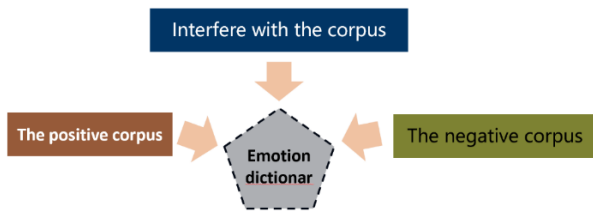


Fig. 1. Dictionary-based approach to text sentiment analysis

B. Machine Learning Methods

Support vector machines (SVMs), plain bayes, and logistic regression are among the frequently used algorithms in machine learning techniques, which are extensively employed in sentiment analysis [12], as shown in Fig. 2. These methods improve classification performance through feature selection and extraction techniques such as bag-of-words models and n-gram features. However, these methods may face computational inefficiency when dealing with large-scale data and require a large amount of labelled data for model training.

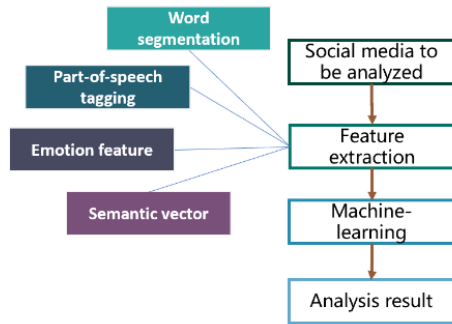


Fig. 2. Sentiment analysis of text based on machine learning approach analysis.

C. Deep Learning Methods

Deep learning methods have made significant progress in sentiment analysis, mainly through models such as Convolutional Neural Networks (CNNs), Long Short-Term Memory Networks (LSTMs) and BERT [13]. These models are able to automatically learn complex features in text without the

need to manually design features, which improves the accuracy of sentiment classification, as shown in Fig. 3.

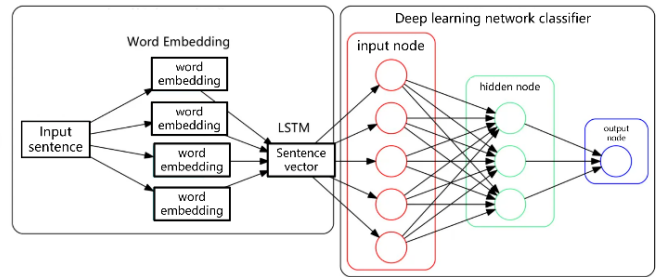


Fig. 3. Sentiment analysis of text based on deep learning approach.

D. Fine-Grained Sentiment Analysis

Fine-grained sentiment analysis focuses on identifying and analysing specific attributes in reviews and their emotional tendencies [14], which can provide more accurate sentiment analysis results. This approach classifies the polarity of comments by identifying object attributes and their corresponding sentiment words, as shown in Fig. 4.

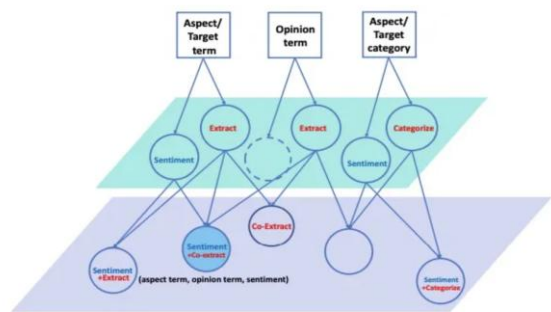


Fig. 4. Sentiment analysis of text based on fine-grained sentiment analysis.

E. Multimodal Sentiment Analysis

Multimodal sentiment analysis integrates many forms of information, including text, audio, and video, to facilitate sentiment recognition [15], thereby providing a more thorough comprehension of the user's emotional disposition. This method enhances the precision of sentiment analysis via modal fusion techniques, including feature-level fusion and decision-level fusion, as seen in Fig. 5.

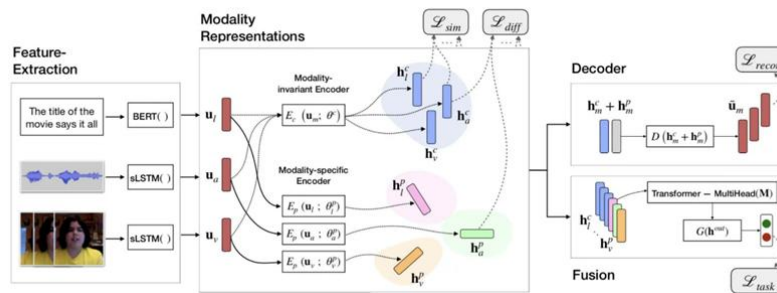


Fig. 5. Sentiment analysis of text based on multimodal information.

III. MULTIMODAL INFORMATION SENTIMENT ANALYSIS

The specific operation of the multimodal sentiment analysis task for video comment text is shown in Fig. 6. Firstly, the information of different modalities is fed into the feature extraction model, then the extracted features are fused into the feature fusion according to a certain fusion method, and finally the fused features are fed into the classifier to perform sentiment analysis on the multimodal data [16].

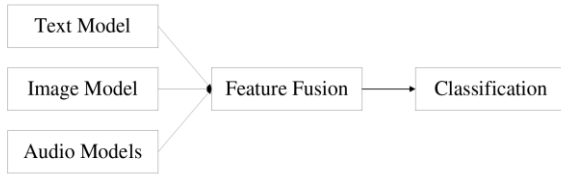


Fig. 6. Multimodal sentiment analysis architecture.

A. Feature Extraction in Multimodal Sentiment Analysis

Extraction of representative features from all modal data utilized in subsequent fusion tasks is referred to as feature extraction in multimodal sentiment analysis. Multimodal sentiment analysis tasks in machine learning, deep learning, artificial intelligence, natural language processing, image processing, and audio processing depend on feature extraction, and the accuracy and speed of the model operation can be directly impacted by the quality of feature extraction [17].

The objective of feature extraction is to derive more representative and interpretable features from the source data to enhance its description and differentiation. Feature extraction typically encompasses the subsequent steps: 1) Data preprocessing; 2) Feature selection; 3) Feature extraction; and 4) Dimensionality reduction of features.

B. Feature Fusion in Multimodal Sentiment Analysis

In the process of multimodal sentiment analysis, the feature extraction stage is not significantly different from the unimodal feature extraction method. However, the core difference between multimodal sentiment analysis and unimodal analysis lies in how to effectively fuse the information from different modalities to derive accurate sentiment polarity.

The methods of feature fusion mainly include feature-level fusion (Feature-level fusion) and decision-level fusion (Decision-level fusion), as shown in Fig. 7 and Fig. 8.

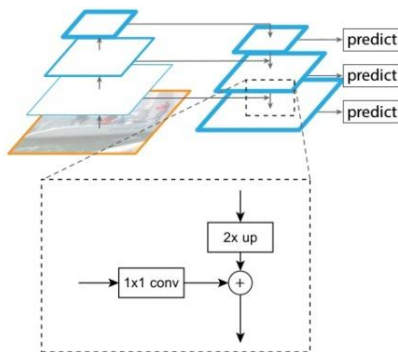


Fig. 7. Feature level fusion.

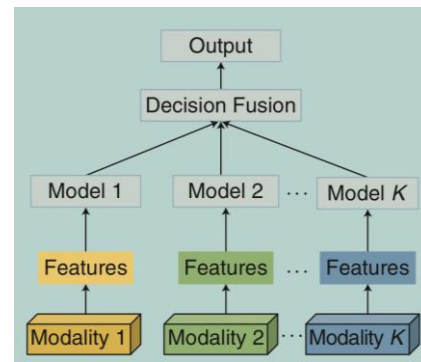


Fig. 8. Decision-level fusion.

Feature-level fusion (FLF) refers to the fusion of multiple features from different feature extraction methods or feature representations to generate a more representative and enriched feature vector. Feature-level fusion is one of the most commonly used methods for multimodal data fusion, and by fusing features from different modalities, a more comprehensive, accurate and robust feature representation can be extracted, thus improving the performance of the model.

Decision-level fusion (DLF) is a fusion strategy in which the features of each modality are first analysed individually, and the results of their respective analyses are subsequently integrated into a single decision vector to arrive at a final decision. The significant advantage of this fusion approach is that when data is missing for one modality, it is still possible to rely on information from other modalities for decision making.

IV. SENTIMENT ANALYSIS BASED ON THE M-S MULTIMODAL SENTIMENT MODEL RESEARCH

A. M-S Model Multimodal Fusion Sentiment Analysis Model

In this paper, a novel fusion model M-S model Multimodal Fusion Sentiment Analysis Model is proposed to perform the task of video comment text sentiment analysis as shown in Fig. 9. This model firstly uses a one-dimensional convolutional neural network to extract the text information, audio information, and image information, and the three modal features Feature-L, Feature-A, and Feature-V (text feature, audio feature, and image feature) obtained, and then the obtained features are sent to the Transformer-layer to be processed, and then the features containing cross-modal The features containing cross-modal attention information (Cross-modal attention Feature-X→Feature-Y), and the features of different modalities are sent to the main and sub dual channels for processing according to their importance, and finally, the processed features of the two channels are sent to the Fully Connected Layer (FC-layer) for processing, and are outputted to the Softmax classifier for classification.

This paper aims to propose a lightweight model with low time complexity that is easy to implement. Therefore, the most fundamental one-dimensional convolutional neural network is selected for feature extraction, while the Transformer model (Fig. 10) is utilized for feature processing to investigate the relationships and influences of various modalities. By fine-tuning critical parameters in the Transformer, it significantly contributes to both cross-modal attention and multi-modal

feature fusion. By modifying the critical parameters in the Transformer, it significantly contributes to both cross-modal attention and multi-modal feature fusion. The primary and

secondary dual-channel fusion approach introduced in this study enables the model to achieve high accuracy and stability.

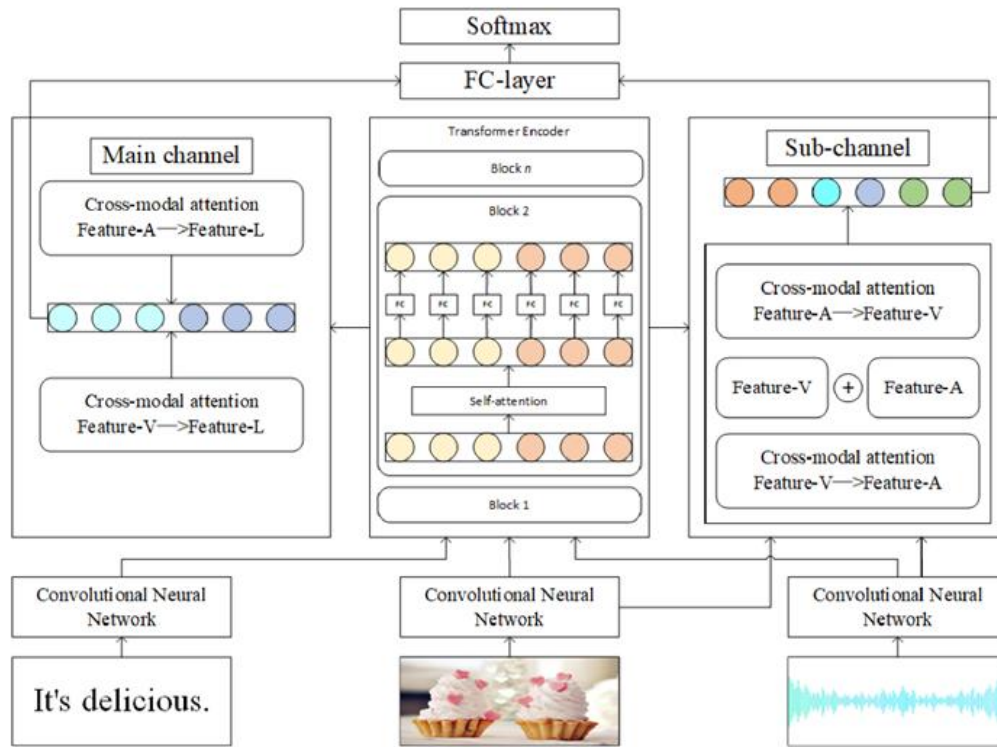


Fig. 9. M-S model.

Multimodal information feature extraction: In this paper, we use 1D Convolutional Neural Network (1D-CNN) for text information, image information, and audio information, which is a variant of convolutional neural network, and compared with traditional fully-connected neural network, 1D-CNN can better deal with localised Relationships. Considering that multimodal datasets are often video-based, resulting in three modes of data containing time series, 1D convolutional networks can effectively extract the features of each mode in the dataset, as shown in Fig. 11.

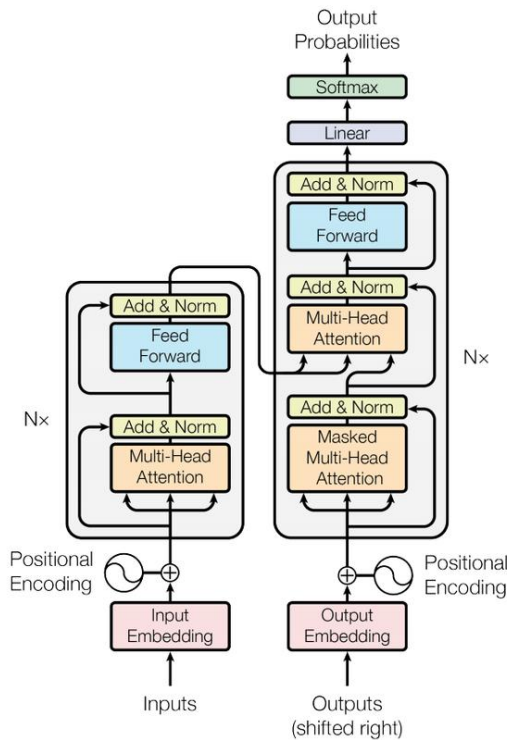


Fig. 10. Transformer model.

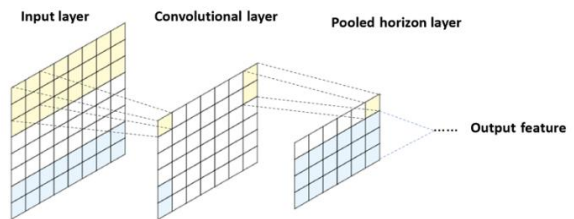


Fig. 11. One-dimensional convolutional neural network.

This paper employs a Transformer-based feature fusion model for multimodal sentiment analysis, emphasizing the significance of fused features in influencing outcomes. The objective is to preserve or enhance the maximum information prior to fusion, utilizing the Transformer encoding and decoding techniques for feature integration. The fusion method ensures

that the extracted features optimally preserve the characteristics prior to fusion, after which the three modal features are combined and encoded to provide a segment of the fused features post-fusion. Subsequently, the fully connected method is employed, incorporating a Dropout layer to mitigate overfitting. The FC-layer consists of two Dropout layers, two Linear layers, and one ReLU activation layer, after which the resultant features are input into a Softmax classifier to derive their sentiment class labels.

B. Multimodal Information Feature Extraction

A one-dimensional convolutional neural network is a specialized neural network designed for sequence data processing, comprising an input layer, a convolutional layer, and a pooling layer, as seen in Fig. 11. The input layer incorporates information from three distinct modalities, while the convolutional layer is trainable to derive an optimal set of convolutional kernels that minimize the loss function, hence facilitating automatic feature extraction.

$A = [a_1, a_2, L, a_s]^T$ is passed as model input to the input layer, where $A \in \mathbb{R}^{s \times d}$ is the time series, s is the length of the time series, and d is the number of eigenvalues. a_i denotes the feature vector at the time of i , with the dimension of d . The sequence data is mapped into convolutional layer by one dimensional convolution operation:

$$f_r(z) = \max(z, 0) \quad (1)$$

$$y_c^j = f_r(A \otimes W_c^j + b) \quad (2)$$

Where: \otimes denotes one-dimensional convolution operation; y_c^j denotes the j th feature mapping generated by the convolution kernel w_c^j , $j \in [1, n_c]$, n_c denotes the number of convolution kernels, and the convolution kernel $W_c^j \in \mathbb{R}^{m \times d}$ is a weight matrix, where m is the size of the convolution kernel, and m denotes the width of the local time window for extracting the time series features for the time series; b is the bias; and $f_r(z)$ is the activation function (ReLU activation function), which is used to non-linearise the data after the convolution operation.

The pooling procedure is employed to extract the most pertinent information on the sequence of features in the convolutional layer, hence creating the pooling layer. This article employs max-pooling to aggregate the features. Upon the final application of the pooling procedure, global max-pooling is employed to extract the most pertinent global temporal information, resulting in a sequence length of 1, and the features of the three modalities are extracted from the information y_p^j .

$$y_p^j(k) = \max(y_c^j(2k-1), y_c^j(2k)) \quad (3)$$

$$y_p^j_{,last} = \max(y_c^j) \quad (4)$$

V. EXPERIMENTS AND ANALYSIS OF RESULTS

A. Baseline Modelling

1) *MCTN model*: The method introduces an innovative strategy to learn robust transitions between joint representations modalities through a translation process [18]. Specifically, the translation process from source to target modality not only provides a new way to learn joint representations, but also requires only the modality of the source modality as input. In order to further strengthen the translation effect of modalities, the method utilises cyclic consistency loss to ensure that the joint representation retains the maximum information of all modalities, as shown in Fig. 12.

2) *LMF model*: While previous studies have explored the tensor expressivity of multimodal representations, these methods often suffer from a dramatic increase in dimensionality and computational complexity due to the transformation of inputs into tensors [19]. To address this issue, this method proposes a low-rank multimodal fusion strategy that utilises a low-rank tensor for multimodal fusion, thus significantly improving the efficiency, as shown in Fig. 13.

3) *TFN model*: The model redefines the problem of multimodal sentiment analysis as a problem of modelling intra- and inter-modal dynamics [20]. To this end, it introduces a novel model, the tensor fusion network, which is capable of learning both dynamics end-to-end. The method is optimised especially for the instability of spoken language in online videos and the accompanying gestures and speech, as shown in Fig. 14.

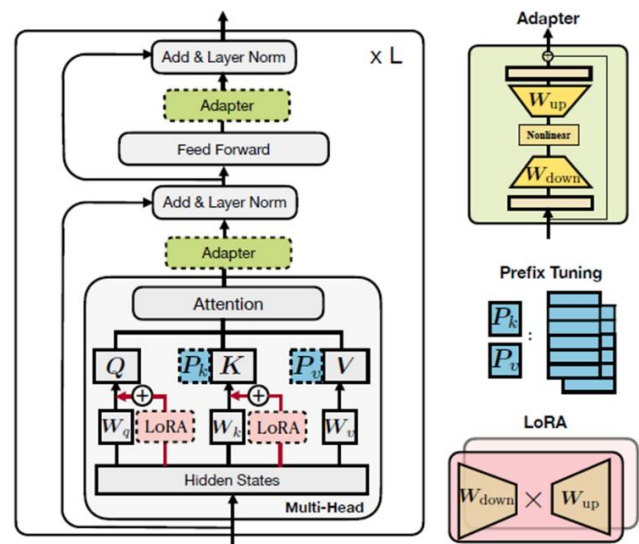


Fig. 12. MCTN model.

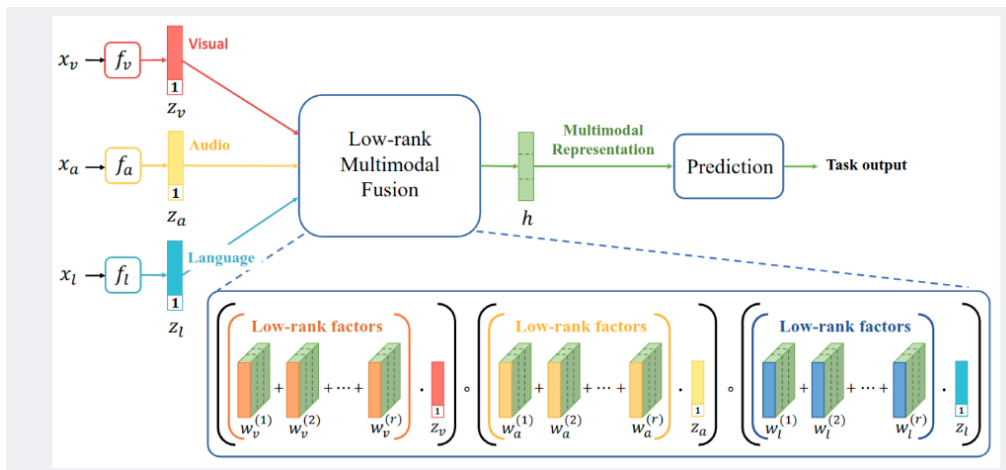


Fig. 13. LMF model.

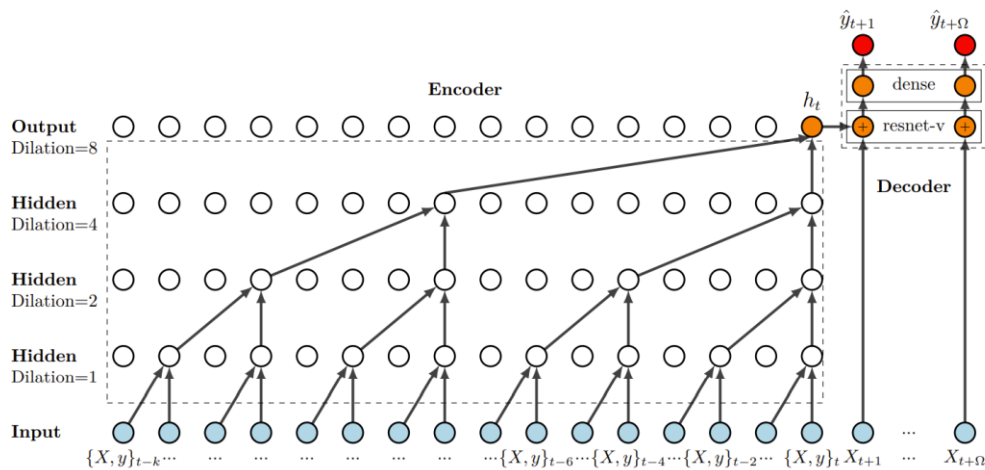


Fig. 14. TFN model.

4) *MFN model*: The proposed approach to address the multi-view sequence learning problem is the Memory Fusion Network (MFN) [21]. This neural architecture explicitly accounts for intra-view and inter-view interactions, modeling them sequentially across the temporal dimension. The MFN comprises an LSTM system designed to learn view-specific interactions in isolation while identifying cross-view interactions via a mechanism known as Delta-memory Attention Network (DMAN). This process culminates in a multi-view gated memory for temporal summarisation, as illustrated in Fig. 15.

5) *EF-LSTM model*: The EF-LSTM model does this by connecting multimodal input data and processing it using an

LSTM network (Hochreiter and Schmidhuber 1997). This design allows the model to process information from different modalities simultaneously, thus improving the performance of sentiment analysis.

6) *Mult model*: The Mult model effectively tackles the challenges of aligning information across different modalities and managing long-term dependencies within the same modality in an end-to-end framework, eliminating the necessity for explicit data alignment [22]. The model's foundation is its directed two-by-two cross-modal attention mechanism, which facilitates the examination of interactions across various time steps in a multimodal sequence, potentially allowing for alignment between modalities, as illustrated in Fig. 16.

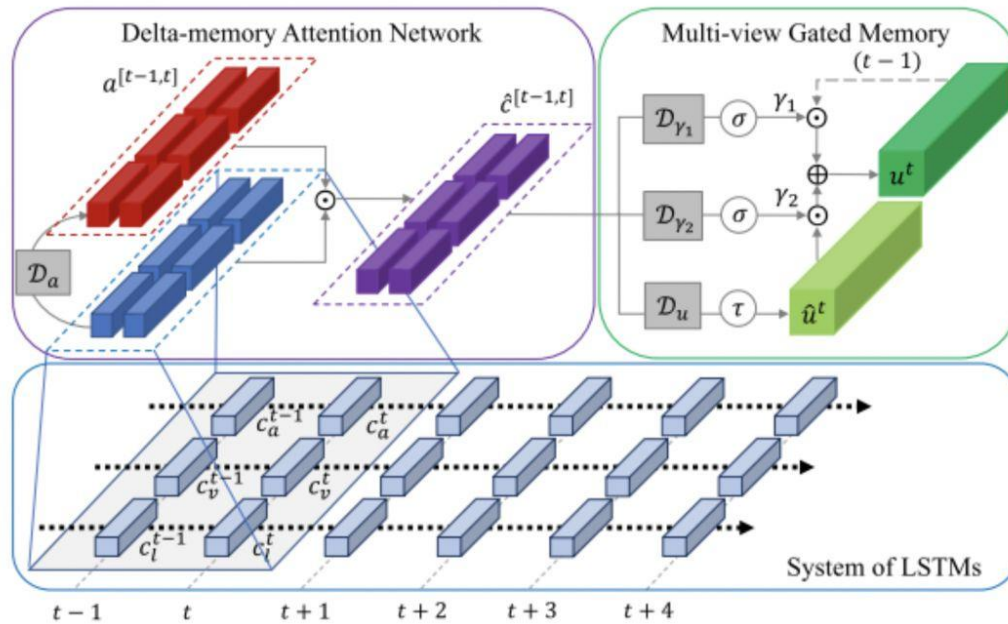


Fig. 15. Schematic diagram of the MFN model.

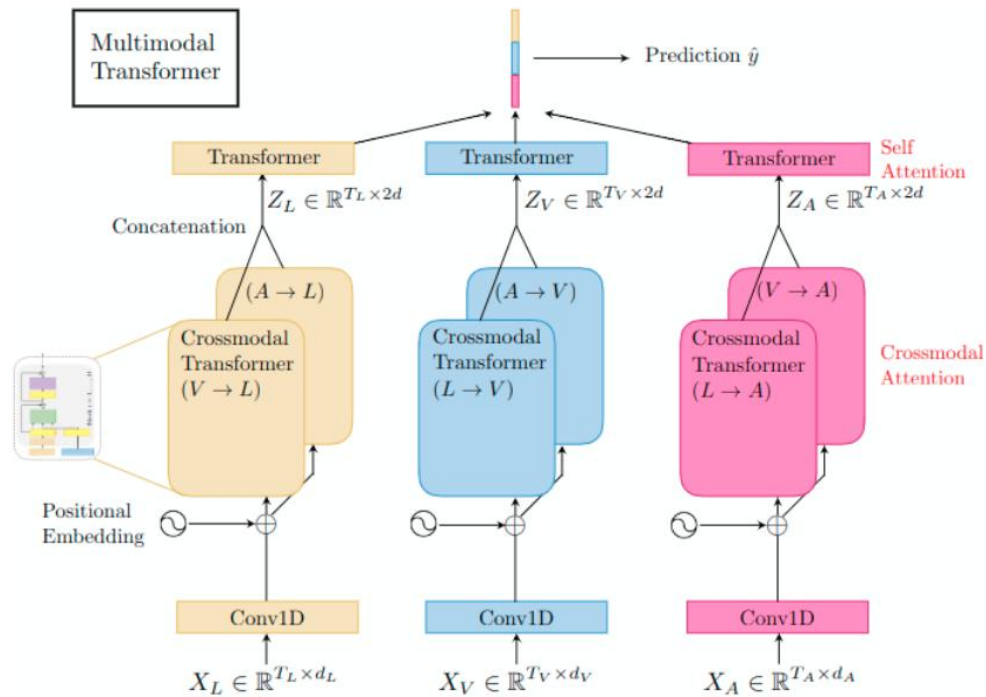


Fig. 16. Schematic diagram of the MulT model.

The dataset used in this paper is the CMU-MOSEI multimodal dataset, which is the first, largest and highly regarded multimodal emotion and sentiment recognition dataset.

B. Experimental Setup

In this paper, three experiments are set up, namely the comparison experiment, the principal mode selection experiment and the ablation experiment. Source of experimental

data were obtained from the text of the Communist Youth League video comments.

1) Comparison test: This experiment is intended to prove the applicability of the method used in the M-S model and the accuracy of the model, this paper uses the CMU-MOSEI dataset to train on each of the six baseline models 10 times respectively (the MulT model is the most effective of the baseline models,

so 50 experiments were carried out), and in order to prove the stability of the M-S model proposed in this paper, the was trained 50 times and the average results obtained.

2) *Main modality selection experiment*: Text features, image features and audio features are selected to be sent to the main channel for processing, and the features of the remaining modalities are sent to the subchannel for processing, and the results are obtained through the M-S model to determine the main modality to be sent to the main channel for processing.

3) *Ablation experiment*: Utilizing solely unimodal information (input from only one modality—text, image, or audio—while excluding data from the other modalities), the experimental findings indicate that information from all modalities in multimodal sentiment analysis positively influences the sentiment analysis task.

C. Experimental Analyses

Fig. 17 illustrates that the M-S model introduced in this paper yields superior outcomes across four metrics: sentiment 2 classification, sentiment five classification, sentiment 7 classification, and F1 value. This demonstrates that the proposed primary and secondary dual channels not only attain high accuracy but also exhibit versatile applicability in multimodal sentiment analysis tasks. Furthermore, in comparison to sentiment 2 classification, the model shows greater improvement in multicategory sentiment classifications (sentiment 5 and sentiment 7), suggesting that the proposed method is more adept at precise sentiment analysis. The Classification, Sentiment 7 Classification demonstrates greater improvement, suggesting that the method provided herein is more suitable for precise sentiment analysis.

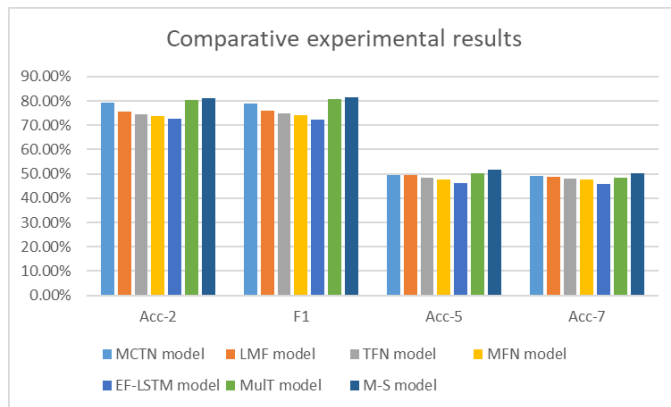


Fig. 17. Comparative experimental results.

Fig. 18 indicates that the results are markedly superior when text features are utilized as the primary modal compared to when image and audio features are employed. This demonstrates that the paper prioritizes text information as the main modality, directing the text features extracted by the one-dimensional convolutional neural network to the primary channel for processing, while the extracted image and audio features are allocated to the subchannel. This approach yields optimal results and substantiates the beneficial impact of the proposed main and subchannel methods on the multimodal sentiment analysis task. The vice-channel strategies presented in this paper positively influence the multimodal sentiment analysis challenge.

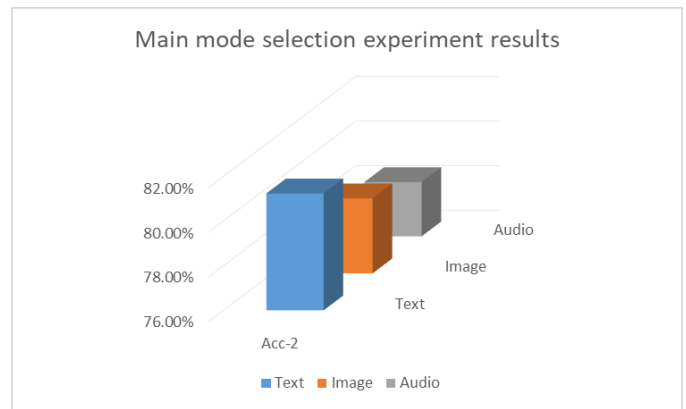


Fig. 18. Main mode selection experimental results.

In Fig. 19, the fusion method proposed in this paper in the CMU-MOSEI dataset when experiments are conducted using multimodal data, the effect is 6.94% more accurate with single text modal sentiment analysis, 21.11% more accurate with single image modal sentiment analysis, and even more accurate with 37.6% more accurate than the single audio modal, which demonstrates that each modal information in this method makes a positive effect on the present sentiment analysis task.

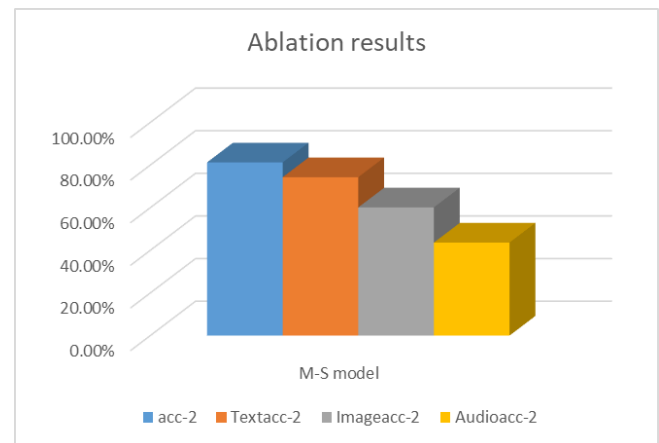


Fig. 19. Ablation results.

VI. CONCLUSION

The experimental results indicate that the M-S model introduced in this paper enhances the sentiment analysis of multimodal information in video review texts compared to the baseline model. The accuracy of the proposed model has been substantiated through various datasets and multiple experimental sets. Furthermore, the ablation experiments demonstrate that the textual, audio, and visual information utilized in this model contribute positively to the multimodal sentiment analysis task.

The M-S model proposed in this paper mainly solves the problems of the multimodal sentiment analysis task in which the time complexity of the model is too large for practical application, the interrelationships and mutual influences between multimodal information are not considered, and the weights of the multimodal features cannot be accurately defined by numerical values, but there are still a lot of challenges that need to be solved by researchers in this task. This paper

summarises some of the future work on the multimodal sentiment analysis task:

- Multimodal datasets are scarce, unsupervised or semi-supervised methods can be considered to train the model and get better results with fewer datasets;
- Use the topology of the hypergraph to establish the relationship between the multimodal data and obtain the feature tensor between the multimodalities;
- Majority of existing methodologies predominantly focus on textual, visual, and auditory data, while pose-oriented and ECG-based sentiment analysis remains exceedingly limited; future efforts should enhance collaboration with other disciplines to develop more comprehensive multimodal datasets;
- Most existing methodologies do sentiment analysis on aggregate data; future approaches may explore sentiment analysis of specific entities within the data or at the aspect level.

ACKNOWLEDGMENT

This work was supported by High-level Talent Project of Huzhou Vocational & Technical College (2024ZS03) and Jinghu Talent Training Project of Huzhou Vocational & Technical College.

REFERENCES

- [1] Sayrol E .Development of a platform offering video copyright protection and security against illegal distribution[C]//2005:76-83.DOI:10.1117/12.591742.
- [2] Zongyue W , Sujuan Q .A sentiment analysis method of Chinese specialised field short commentary[C]//IEEE International Conference on Computer & Communications.IEEE, 2017:2528-2531.DOI:10.1109/CompComm.2017.8322991.
- [3] Simm W , Ferrario M A , Piao S S ,et al. Classification of Short Text Comments by Sentiment and Actionability for VoiceYourView[J].IEEE, 2010.DOI. 10.1109/SocialCom.2010.87.
- [4] Simm W , Ferrario M A , Piao S S ,et al. Classification of Short Text Comments by Sentiment and Actionability for VoiceYourView[J].IEEE, 2010.DOI. 10.1109/SocialCom.2010.87.
- [5] Yu S .A Bullet Screen Sentiment Analysis Method That Integrates the Sentiment Lexicon with RoBERTa-CNN[J].Electronics, 2024, 13.DOI:10.3390/electronics13203984.
- [6] Shi P , Shi M .Emotion analysis method based on domain dictionary and machine learning[J]. 2021.
- [7] Hamouda A , Marei M , Rohaim M .Building Machine Learning Based Senti-word Lexicon for Sentiment Analysis[J]. Technology, 2011, 2(4):199-203.DOI:10.4304/jait.2.4.199-203.
- [8] Liang J , Chai Y , Yuan H ,et al. Deep learning for Chinese microblog sentiment analysis[J].
- [9] Poria S , Majumder N , Hazarika D ,et al. Multimodal Sentiment Analysis[C]//IEEE Educational Activities Department, PUB766, Piscataway, NJ, USA. IEEE Educational Activities Department, PUB766, Piscataway, NJ, USA, 2018.DOI:10.1007/978-3-319-95020-4_7.
- [10] Zhao L , Pan Z .Cross-Modal Semantic Fusion Video Emotion Analysis Based on Attention Mechanism[J].2023 8th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 2023:381-386.DOI:10.1109/ICCCBDA56900.2023.10154781.
- [11] Feng X , Ju F , Hou H ,et al. Sentence Level Fine-grained Emotion Computation Based on Dependency Syntax Improvement Dictionary 1[J]. 2022.
- [12] Shetty P , Kini S , Fernandes R .A Comprehensive Analysis of 'Machine Learning and Deep Learning' Methods for Sentiment Analysis in Twitter[J].SN Computer Science, 2024, 5(7):1-13.DOI:10.1007/s42979-024-03216-2.
- [13] Oumaima B , Amine B , Mostafa B .Deep Learning or Traditional Methods for Sentiment Analysis: a Review[C]//The Proceedings of the International Conference on Smart City Applications.Springer, Cham, 2024.DOI:10.1007/978-3-031-53824-7_3.
- [14] Gonda R , Park J .Fine-Grained Sentiment Analysis of Covid-19 Quarantine Hotels through Text Mining[J]. Conference on Industrial Engineering and Operations Management, 2023.DOI:10.46254/an13.20230207.
- [15] Yujie W , Yuzhong C , Chen J D .A knowledge-augmented heterogeneous graph convolutional network for aspect-level multimodal sentiment analysis[J] . Computer speech & language, 2024, 85(Apr.):101587.1-101587.19.DOI:10.1016/j.csl.2023.101587.
- [16] Zhang H , Wang Y , Yin G ,et al. Learning Language-guided Adaptive Hyper-modality Representation for Multimodal Sentiment Analysis[J]. 2023.
- [17] Cheng H , Yang Z , Zhang X ,et al.Multimodal Sentiment Analysis Based on Attentional Temporal Convolutional Network and Multi-Layer Feature Fusion[J].IEEE transactions on affective computing, 2023(4):14.DOI:10.1109/TAFFC.2023.3265653.
- [18] Pham H, Liang P P, Manzini T, et al. Found in translation: learning robust joint representations by cyclic translations between modalities[J]. arXiv:1812.07809(2019).
- [19] Liu Z, Shen Y, Lakshminarasimhan V B, et al. Efficient low-rank multimodal fusion with modality-specific factors[J]. arXiv:1806.00064 (2019).
- [20] Zadeh A, Chen M, Poria S, et al. Tensor fusion network for multimodal sentiment analysis[J]. arXiv:1707.0725 (2017).
- [21] Zadeh A, Liang P P, Mazumder N, et al. Memory fusion network for multi-view sequential learning[J]. arXiv: 1802.00927 (2018).
- [22] Tsai Y H, Bai S J, Liang P P, et al. Multimodal transformer for unaligned multimodal language sequences[C]//Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. florence: 2019. 6558-6569.

Enhancing Stock Market Forecasting Through a Service-Driven Approach: Microservice System

Asaad Algarni

Department of Computer Sciences-Faculty of Computing and Information Technology,
Northern Border University, Rafha 91911, Saudi Arabia

Abstract—Predicting stock market is a difficult task that involves not just a knowledge of financial measures but also the ability to assess market patterns, investor sentiment, and macroeconomic factors that can affect the movement of stock prices. Traditional stock recommendation systems are built as monolithic applications, with all components closely coupled within a single codebase. While these systems are functional, yet they are difficult integrating several services and aggregating data from diverse sources due to their lack of scalability and extensibility. A service-driven approach is needed to manage the growing complexity, diversity, and speed of financial data processing. However, microservice architecture has become a useful solution across multiple sectors, particularly in stock systems. In this paper, we design and build a stock market forecasting system based on the microservice architecture that uses advanced analytical approaches such as machine learning, sentiment analysis, and technical analysis to anticipate stock prices and guide informed investing choices. The results demonstrate that the proposed system successfully integrates multiple financial analysis services while maintaining scalability and adaptability due to its microservice architecture. The system successfully retrieved financial metrics and calculated key technical indicators like RSI and MACD. Sentiment analysis detected positive sentiment in Saudi Aramco's Q3 2021 report, and the LSTM model achieved strong prediction results with an MAE of 0.26 and an MSE of 0.18.

Keywords—Stock market; microservice architecture; deep learning; technical indicators; sentiment analysis

I. INTRODUCTION

Analyzing the stock market to predict the price movement of shares is of interest to investors and traders. There is an extensive variety of interrelated factors affecting the stock price, including microeconomic considerations and geopolitical developments [1, 2]. A study found that psychological factors such as investor emotion and behavioral biases had a considerable impact on share prices [3]. Traditional methods for predicting stock prices depend on statistical models and technical analysis [4]. However, they are unable to completely comprehend the complex interdependencies and behavioral patterns that are inherent in investor psychology and market dynamics [5]. Consequently, deep learning has become recognized as a highly efficient tool for deriving meaningful insights from extensive, complex datasets [6].

The financial markets are very volatile and process intensive data [8]. Hence, they require sophisticated systems that can handle huge datasets, run complex analyses, and give real-time insights that can be used. Most traditional stock

recommendation systems, on the other hand, are built as monolithic applications, with all components closely coupled within a single codebase. While these systems are functional, yet they are difficult in scalability, maintainability, and extensibility [7].

Also, it is critical for investors and financial analysts to be able to forecast stock movement with precision, as they must make informed decisions in a constantly changing environment. Forecasting systems that rely on monolithic architectural structures are difficult to adapt to ever-changing markets. These systems focus on limited data inputs, like historical price data, while overlooking critical factors such as real-time sentiment analysis and technical indicators, which play a significant role in market behavior. To the best of our knowledge, current approaches struggle to combine diverse analytical methods—like machine learning, sentiment analysis, and fundamental analysis—into a unified, effective system. This may lead to inconsistent performance and a lack of flexibility when applied to dynamic and complex market conditions.

Microservice architecture is now widely adopted across various sectors. Microservice divides a large system into small, self-contained services that provide benefits such as improved scalability, greater flexibility, and error tolerance [9]. The development of stock systems based on a microservice architecture enables the integration of various analytical functions such as artificial intelligence, technical analysis, sentimental analysis, portfolio management, and risk assessment. Using this modular strategy, traders and investors may precisely adjust market circumstances to make better-informed decisions.

Thus, this study proposes the design and implementation of a stock recommendation system based on microservices that use specialized services to deliver informed stock recommendations. The system's goal is to enhance the precision and relevance of stock recommendations by operating machine learning algorithms, fundamental and technical analysis, and sentiment analysis of financial reports. This research extends the existing literature on financial systems by examining how microservices can improve the efficiency, scalability, and usefulness of generating stock recommendations.

The paper is organized as follows: a literature review is conducted in Section II. Section III demonstrates the methodology, including the proposed system. Sections IV and

V present the study's results and discussion. Section VI presents the research conclusion and future work.

II. BACKGROUND AND RELATED WORK

A. *Microservice Architecture*

Microservice architecture is becoming an effective solution across multiple sectors, particularly in stock systems because it excels at managing complex, large-scale applications that demand scalability, maintainability, and flexibility. Microservices empower teams to develop, deploy, and scale each component independently, enhancing efficiency and innovation by breaking down a system into smaller and independent components. Each service is designed to handle a certain business function and communicates with other services through lightweight protocols, typically using APIs [10].

Microservices systems can significantly improve the efficiency and responsiveness of financial trading systems. The shift from a monolithic architecture to a micro-services-based approach enables developers to create independent services that satisfy certain requirements, improving scalability and maintainability. For instance, a microservice reference architecture based on domain engineering can address the high complexity and maintenance costs associated with large-scale financial trading systems by providing a general solution for similar scenarios, allowing new microservices to coexist with legacy systems and supporting rapid business development through DevOps and other mechanisms [11].

Microservices in power trading platforms also highlight their ability to enhance load performance and scalability, which are critical for the real-time and complex nature of stock trading [12]. Furthermore, microservices can be integrated with IoT technologies to create efficient inventory management systems and benefit stock systems by ensuring real-time data processing and business logic execution [13]. Implementing microservices architecture within ERP financial systems showcases their ability to address complex business demands with exceptional availability, security, and scalability, which are crucial for stock systems managing substantial amounts of data and transactions [14].

Adopting microservice architecture in stock recommendation systems may offer numerous benefits, including improved scalability, maintainability, and responsiveness. Thus, making it a valuable approach for modern financial trading platforms and related applications. The effective management of systems based on micro-services is vital for ensuring service quality across various microservice components, which is particularly important in stock systems where resource demands and quality of service requirements are high [15]. Integrating microservices with social-media messaging bots for automated inventory management further illustrates their versatility and potential for enhancing customer and retailer connectivity in stock systems [16]. Additionally, microservices can be applied to assess the impact of news on the stock market, which makes the system respond to news events, analyze them, and predict their effects on market trends. Thus, it provides valuable insights for traders and investors. Additionally, microservices in asset tracking systems, such as those based on indoor positioning systems,

showcase their potential for real-time data processing and resource-efficient operations, which can be applied to stock systems for tracking and managing assets efficiently [17].

The potential implications of a booming stock market prediction system through deep learning algorithms services built on microservice architecture are essential. Accurate forecasting could inform investment decisions and contribute to risk management strategies, portfolio optimization, and the development of automated trading systems. Furthermore, the valuable insights derived from this research have the potential to revolutionize financial forecasting and expand economic modeling applications.

B. *Machine Learning Algorithms*

Integrating machine learning and pattern recognition improved the accuracy of stock trading systems [18]. Esteemed machine learning methods have been utilized to predict fluctuations in stock prices. Some of these powerful algorithms are Long Short-Term Memory (LSTM), Support Vector Machine (SVM), Random Forest, and Decision Tree, each algorithm offers unique advantages. LSTM, a recurrent neural network, demonstrates exceptional efficacy in encapsulating time-based dependencies and intricate patterns within stock market data, as shown by its exceptional forecasting accuracy compared to conventional models like ARIMA [19] [20]. Despite the complexity of machine learning models, it is suggested that straightforward strategies are effective for predicting price soaring. A study conducted by [21] evaluates the effectiveness of using deep learning and technical indicators to predict short-term stock price movements. The authors developed a four-layer Long Short-Term Memory (LSTM) model incorporating various technical indicators, achieving an impressive 83.6% accuracy in forecasting stock trends. Authors in study [22] offer a comprehensive comparative analysis of nine machine learning models and two deep learning methods. The RNN and LSTM with continuous data emphasized superior performance among others. It also demonstrated a notable improvement in accuracy across all models when technical indicators are converted to binary data inputs.

Numerous studies have generated precise estimations by detecting non-linear correlations in stock market data [23, 24]. Random Forest, an ensemble learning method, is particularly adept at managing large datasets and mitigating overfitting by aggregating multiple decision trees. Although Decision Trees are easy to interpret and implement, they are at risk of overfitting, especially with small datasets. Nonetheless, they can remain beneficial when integrated with other models or used with an ensemble technique like Random Forest [25]. Comparative studies have shown that advanced models like LSTM and SVM surpass decision trees in performance. Nonetheless, decision trees remain valuable for data exploration and feature selection [26]. However, the selection of a model typically relies on the requirements of the prediction task, including the dataset's size and design, the necessary balance between accuracy and interpretability, and the available computational resources [23]. Each model offers distinct advantages. In some cases, the most precise forecasts arise from a combination of these models, which integrate the

optimal elements of each to mitigate the shortcomings [23, 25 - 27].

Disparate data sources, such as traditional time-series data and web platforms such as Google and Wikipedia, have been shown to greatly improve prediction accuracy [28]. The authors of study [45] emphasize the value of using NLP and finance to forecast stock prices. In their study, they emphasized the use of historical price data, text data from the news or social media, and their combination to improve accuracy. They discuss algorithms like RNN, GRU, and LSTM for price analysis, sentiment analysis for assessing investor emotions, and methods such as graph networks and event-driven approaches for identifying company correlations. A study investigates sentiment analysis of online investor opinions to support investment decisions and risk assessment. Using data from Sina Finance, it combines machine learning methods, like SVM and GARCH models, with sentiment analysis. Results show strong correlations between forum sentiment and stock price volatility, with machine learning outperforming semantic approaches and sentiment having a greater impact on value stocks than growth stocks [46].

The stock market is a complicated financial system with shifting prices, and investors seek to maximize gains while minimizing risks. Recent advances in neural networks and hybrid models have exceeded older approaches in terms of prediction accuracy. A study introduced the SMVF-ANP technique, which analyzes market transmission processes and financial aspects using multi-layer perceptron (MLP), dynamic artificial neural networks, and GARCH models. Results reveal that SMVF-ANP outperforms typical strategies and models, with a prediction accuracy of 97.2%, an efficiency rate of 96.9%, and higher returns [47]. Tong et al. studied the effect of social media mood on irrational herding behavior in the Chinese stock market. They discovered that emotion had a considerable impact on illogical conduct [48]. Furthermore, authors in study [49] extracted insights from unstructured financial text in quarterly company reports. They utilized FinBERT, a pre-trained language model based on the BERT framework. Their results demonstrate that FinBERT outperforms state-of-the-art methods, achieving an accuracy of 84.77% on quarterly reports.

To meet the varied demands of investors, advanced stock recommender systems have been expertly developed, which utilize techniques such as K-Nearest Neighbor, Singular Value Decomposition, and Association Rule Mining, all of which take into account investors' unique preferences and risk profiles to maximize portfolio returns [29]. Multi-agent recommender systems utilizing hybrid filtering techniques and involving collaborative and content-based filtering have been designed to adaptively recommend profitable stocks based on investor preferences and macroeconomic factors [30]. Incorporating social media text and company correlations into stock movement prediction models has enhanced the ability to capture multimodal signals, providing a robust tool for investment decision-making [31]. A systematic review of recent developments in machine learning methods, such as deep learning and ensemble methods, highlights the importance of these technologies in improving stock market movement forecasts and reducing investment risks [32]. In

addition, a deep reinforcement learning approach that combines artificial neural networks (ANN), long short-term memory (LSTM), natural language processing (NLP), and deep Q networks (DQN) was proposed to forecast the next day's stock price. The proposed model outperforms standard algorithms in terms of accuracy, demonstrating its efficacy in automating stock market investment decisions [33]. Thus, the remarkable innovations in stock recommendation systems highlight their extraordinary sophistication, empowering them to deliver precise, timely, and personalized investment guidance that perfectly aligns with the dynamic needs of investors and traders.

C. Technical Analysis

Technical indicators are vital in forecasting stock market movements that offers valuable insights into market trends and behaviors. The metrics are typically divided into two primary classifications: trend indicators and volume indicators. Trend indicators play a pivotal role in revealing not only the direction but also the strength of a market trend, which is essential for crafting savvy investment choices. Noteworthy instances of trend indicators encompass Moving Averages (MA), Moving Average Convergence Divergence (MACD), and Exponential Moving Average (EMA) [34]. These indicators are utilized to smooth out price data, facilitating the detection of trends over a designated timeframe. For example, the MACD, a momentum indicator, illustrates the correlation between two moving averages of the price of an asset trend. At the same time, the EMA assigns greater importance to the most recent prices, increasing its sensitivity to new information [35].

On the other hand, volume indicators are crucial as they reveal the true power behind price movements by examining trade volume. They assist investors in understanding the degree of interest surrounding a specific stock or market. Common volume indicators include Relative Volume (RVOL), Volume Weighted Average Price (VWAP), and Chaikin Money Flow (CMF) [34]. RVOL compares the current volume to the average volume over a specific period, indicating whether a stock is actively traded. The VWAP determines the mean price at which a security was traded during the day, considering volume and price, in contrast, the CMF assesses buying and selling pressure over a specific period.

The precision of stock predictions could be increased by combining these metrics with sophisticated machine learning algorithms. Combining technical indicators with ensemble learning methods like Random Forest and Gradient Boosting has improved predictive precision, attaining a success rate of 91.45% in forecasting the opening price of stock [36]. Similarly, using LSTM models with technical indicators as voters have demonstrated high accuracy in stock market predictions, with RMSE values indicating strong performance [34]. Jaideep and Matloob developed a machine learning classifier using five technical and 23 fundamental indicators. Their findings highlight the importance of analyst ratings in predicting stock prices [37]. Thus, these methods emphasize the need of selecting the proper set of technical indicators to improve the efficacy of prediction models. Furthermore, utilizing correlation-based feature selection models to identify crucial technical indicators has undeniably revolutionized the forecasting capabilities of deep learning algorithms,

particularly Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) [38].

Hybrid machine learning models that integrate various classifiers and optimization strategies have successfully forecasted stock price fluctuations, especially in volatile market conditions affected by external elements such as political turmoil and pandemics [39]. However, the strategic use of trend and volume indicators, along with advanced machine learning techniques, provides a robust framework for predicting stock market movement; thus, they offer rich visions for investors and financial analysts [4, 34-36, 38, 40, 41-43].

However, existing systems focus on historical price data and fail to integrate other valuable inputs, such as sentiment analysis, machine learning algorithms, and technical indicators. Despite their effectiveness, these systems' advanced technologies are often used in isolation. In response to these limitations, we propose a microservice-based architecture that brings together modularity, scalability, and advanced analytics.

III. METHODOLOGY

A. The Proposed System

Making well-informed investment choices is challenging due to the total volume of data accessible to investors. These data sources range from historical stock prices and technical indicators to company fundamentals and sentiment extracted from news articles and financial reports. The ability to process, analyze, and interpret these data effectively is crucial for investors. Hence, we propose a microservice-driven stock prediction system to address this need. Designing the system to provide investors with comprehensive stock analysis and actionable recommendations can optimize their portfolios and minimize risks as well.

The proposed system is built on microservices architecture, and each service was designed to handle a distinct part of the stock analysis process. These microservices operate independently while effortlessly communicating with one another through RESTful APIs. This architecture ensures the system is scalable, efficient, and adaptable to future enhancements. The proposed system was tested using real-world stock data to assess its effectiveness in response to each service and providing initial and informed stock recommendations. The system's performance was evaluated based on its ability to retrieve stock data, generate reliable technical and fundamental insights for traders, and summarize a semantic analysis for a given financial report. (The source code for the system can be accessed through the fine-grained personal access token link provided in the footer at the bottom of the page or by clicking on this link).

The system combines essential financial metrics, sophisticated machine learning algorithms, and sentiment analysis to provide a strong data-informed decision. Each microservice is responsible for a particular function or task in the stock recommendation process, enabling efficient real-time processing of large data sets. Fig. 1 presents the architecture of the system. The major components of the system are detailed as follows:

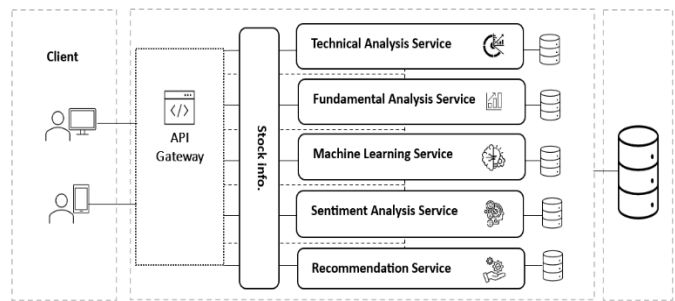


Fig. 1. The architecture of the proposed system.

1) *Stock data and fundamental analysis services*: The first step in the system involves collecting financial data from publicly available sources. This microservice retrieves vital financial metrics for individual stocks, including historical prices and trading volumes. The fundamental analysis microservice can process the data to compute key financial metrics like Market Capitalization, Dividend Yield, Earnings Per Share (EPS), and Revenue. These metrics help evaluate the financial health and performance of the selected company. We used the Yahoo Finance API in our system.

2) *Technical analysis service*: The system's technical analysis microservice calculates a range of widely used technical indicators that assist in identifying trends, momentum, and market conditions. To detect price trends and possible reversals, we run indicators such as moving averages (MA), the Relative Strength Index (RSI), and the Moving Average Convergence Divergence (MACD). Also, this component employs volume-based indicators like On-Balance Volume (OBV) and volatility indicators like the Average True Range (ATR). Thus, these technical indicators are important for traders and investors to determine a stock's price shifts and the overall direction of the trend.

3) *Sentiment analysis service*: Sentiment analysis has become an important tool in predicting stock market trends, given that news and financial reports can greatly affect investor actions and stock valuations. The sentiment analysis process was developed using a Flask-based API that integrates the FinBERT model, a variant of BERT specifically fine-tuned for sentiment analysis in financial contexts. The FinBERT, which is available via Hugging Face's model repository, is known for its domain specific capabilities in financial sentiment classification [48]. It is initialized in the application through the transformers pipeline with a sentiment analysis task. The model outputs predictions across three sentiment categories: positive, negative, and neutral. It processes the text data to determine whether the general sentiment is positive, negative, or neutral, providing a complementary perspective to the technical and fundamental analyses.

4) *Machine learning predictions service*: In its early version, the system uses an LSTM (Long Short-Term Memory) model to estimate future stock values, which is a recurrent neural network designed to manage time-series data. The LSTM model predicts future stock swings using previous

https://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@github.com/aalgarni2/Stock_Market_Forecasting_Microservice_System.git

stock prices and other services. This approach identifies long-term linkages and patterns in stock price data. It aids in providing short-term projections, which are critical for making buy, sell, or hold choices. The model is designed to react dynamically to new data, gradually improving the accuracy of its predictions as time progresses. The system scales and normalizes historical stock data, technical indicators, and semantic analysis findings before passing them through an LSTM model for training. After completion, the model produces forecasted stock values, which are then compared to the current market price to inform the recommendation service. During model execution, we divided the dataset into training (80%) and testing (20%) sets. We set up the model with 50 epochs, a batch size of 32, and an Adam optimizer with a learning rate of 0.001. The model's performance was assessed using Mean Absolute Error (MAE) and Mean Squared Error (MSE). Other machine learning approaches such as Support Vector Machine, Random Forest, and Decision Tree will be implemented in future work. Thus, this modular approach makes it easier to update and expand the system with new capabilities as they become available.

5) *Recommendation service:* The recommendation microservice will integrate the outputs of the fundamental analysis, technical indicators, and machine learning predictions to deliver a final recommendation. The algorithm would be designed to aggregate results from four microservices, fundamental analysis, technical indicators, machine learning predictions, and sentiment analysis, to generate a final stock recommendation. Each service's output is weighted, combined, and compared against set thresholds. Then, the service determines whether the recommendation is to 'Buy,' 'Hold,' or 'Sell'.

6) *User interaction via streamlit interface:* The system's user interface is designed to let investors enter stock ticker symbols, select a date period for analysis, and upload a financial report for sentiment evaluation. The interface retrieves and shows results based on the services that have been selected. The modular architecture also can enable future integration of additional analytical services or advancements with minimal disturbance.

IV. RESULT

A. Fundamental Analysis and Technical Analysis Performance

The component of fetching the historical stock data was tested with several stocks by entering the ticker, start date, and end date. All runs demonstrated that stock data can be retrieved successfully from the Yahoo Finance API. For the fundamental analysis service, it successfully retrieved financial metrics. Additionally, the proposed system effectively calculated and returned multiple technical indicators such as RSI, and MACD. These indicators effectively identified trends, overbought and oversold conditions, and market volatility. The ability to visualize these indicators over a selected date range helped investors gain deeper insights into price movements and market dynamics. Indicators like MACD and RSI provided

actionable insights into buying or selling opportunities. Thus, the system's technical indicators are aligned with known market trends, providing an additional layer of validation to the stock's performance.

During testing of the three components, we entered the stock ticker "2222.SR" (Saudi Arabian Oil Company); the system retrieved stock data for the specified date. Also, when we run the fundamental analysis, the system provides the available and accessible data. For unavailable metrics, it shows 'N/A'. The accuracy and reliability of this data depend on the real-time updates provided by Yahoo Finance, which proved to be timely and comprehensive during testing. Fig. 2 illustrates the system's interface after fetching the stock data.

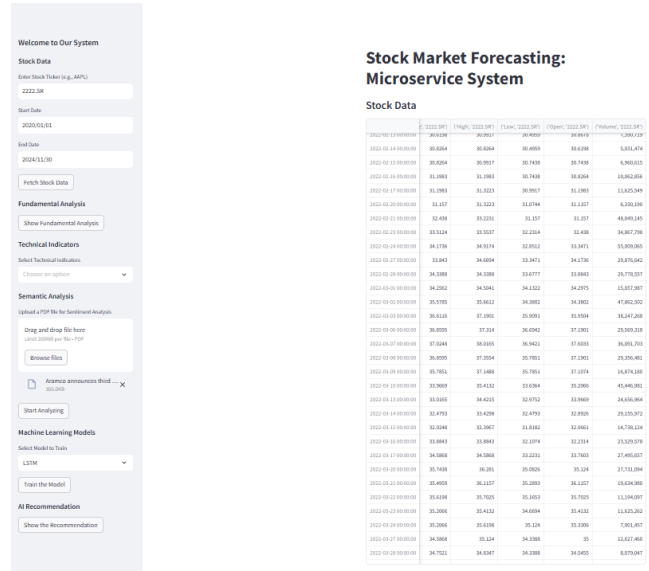


Fig. 2. The proposed system main interface.

B. Sentiment Analysis Services Performance

The sentiment analysis services perform sentiment analysis on any uploaded PDF document. Sentiment analysis is increasingly used in financial prediction. Incorporating news sentiment or financial report analysis improves accuracy in forecasting stock trends [44]. For our proposed system, we implemented the FinBERT model for the sentiment analysis microservice. We tested this service by uploading several PDF files.

The sentiment analysis service uses FinBERT to analyze text from PDF documents. The FinBERT tokenizer divides the text into pieces using tokens. The FinBERT algorithm evaluates each segment to identify the sentiment labels (positive, neutral, or negative), which are then accepted to obtain a sentiment measurement. In an analysis of Saudi Aramco's third-quarter report for 2021, the sentiment analysis service detected positive sentiment. Fig. 3 shows the sentiment analysis of the report. The sentiment analysis was instrumental when applied to earnings reports. Positive sentiment can be associated with subsequent price increases. In contrast, negative sentiment may correlate with price declines. However, while sentiment extracted from financial documents can predict short-term stock price movements, external market factors and investor behavior may influence price fluctuations.



Fig. 3. Sentiment analysis report.

C. Machine Learning Service and Recommendation System

The machine-learning service is designed to employ various machine learning algorithms for stock price prediction. They are LSTM, Support Vector Machine (SVM), Random Forest, and Decision Tree. In this early development, we only implemented LSTM. Other algorithms are left for future work. During testing the service, the LSTM model achieved a Mean Absolute Error (MAE) of 0.26 and a Mean Squared Error (MSE) of 0.18 for predicting the stock price of 2222.SR (Saudi Aramco). The results demonstrate the potential of LSTM for forecasting in financial markets. Fig. 4 show the result of MAE and MSE after training the model.



Fig. 4. The result of MAE and MSE of the model.

The recommendation engine service will integrate the results of fundamental analysis, technical indicators, and semantic analysis and pass them on to the machine learning algorithm service. The recommendation engine's strength is its ability to integrate several types of information which helps in producing a comprehensive insight of the stock's performance. This guarantees that the recommendation is well-rounded and takes into account both quantitative and qualitative considerations. Also, it limits the risk of making decisions that are based on only one factor. The full implementation of this service is left for future work.

V. DISCUSSION AND LIMITATION

The results of testing the system imply that the proposed system can be an effective tool for providing data-driven stock recommendations based on multiple sources. The system offers a comprehensive stock analysis approach by integrating fundamental analysis, technical indicators, sentiment analysis, and machine learning predictions. The modular design of the services facilitates integration with other components of the stock recommendation system. Also, the modular approach allowed for a flexible selection of technical indicators [50]. Extending support for more complex strategies or combining

indicators could enhance the analytical capability of the system further. FinBERT's pre-training on financial data provided domain-specific sentiment analysis, making the results relevant for stock market insights. The service's modular design supports easy scaling, such as adding more sentiment categories or integrating additional pre-trained models. Future work could focus on improving sentiment granularity and analyzing sentiment trends across sections of a document to uncover deeper insights. The LSTM model's architecture can capture temporal dependencies; hence, it is suitable for time-series data like stock prices. However, the ability to choose among models provided flexibility for different user requirements and computational resources. Future improvements include implementing additional time-series-specific models and enabling hyperparameter optimization for models.

Furthermore, the proposed system demonstrates that its internal structure is scalable and efficient. Each microservice operated independently. The utilization of RESTful APIs allows services to communicate without interruption as well as ensures that changes to one service do not impact the overall system. Hence, the system's modularity increases its flexibility and usability by making it simple to add or improve current services.

There are specific constraints that need to be defined. First, there is a need for stress testing to assess performance under high-load conditions. Second, sentiment analysis is heavily dependent on the quality of the input data. Hence, the outcomes of the results can be affected if there are specific constraints that need to be defined. First, there is a need for stress testing to assess performance under high-load conditions. Second, sentiment analysis is heavily dependent on the quality of the input data. So, the outcomes of the results may be affected negatively if there is any biased data.

However, the system's modular architecture allows for future enhancements, such as incorporating more advanced sentiment analysis models, employing advanced machine learning approaches to better account for market volatility, and leveraging recommendation services with explainable AI. This system has the potential to significantly aid investors in making well-informed decisions while also being adaptable to future advancements in financial technology.

VI. CONCLUSION AND FUTURE WORK

The proposed system adopts a comprehensive methodology for stock market analysis, integrating traditional financial metrics with contemporary machine-learning approaches. The system assists investors in making informed decisions by offering services like fundamental analysis, technical indicators, sentiment analysis, and machine learning algorithms. However, this paper presents a microservice-based architecture that embeds state-of-the-art technologies, such as LSTM for time-series prediction and FinBERT for sentiment analysis in stock price forecasting. By comparison, the LSTM model generates smaller error metrics than traditional baselines and thus has demonstrated the capability of learning and representing the complex, nonlinear stock price pattern. The value added to the system for further processing was done by the incorporation of FinBERT, through the analysis of the

financial reports based on sentiment analysis. The paper's parameter optimization will be expanded to capture more market scenarios and embed additional analysis tools for the betterment and adaptability of the system.

The recommendation service developed in this version is still on a very high level of abstraction and, hence, is left for further research and development. Also, we aim, in future work, to introduce Explainable AI (XAI) into the stock trading systems to make the predictions transparent and explain why they are predicted this way. It will be interesting to see whether XAI can be integrated into microservice architecture. The other related research direction could try to investigate how the XAI modules connect with the existing services and, more importantly, provide explanations that are correct, understandable, and trusted.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the Project Number "NBU-FFR-2025-231-02".

REFERENCES

- [1] Zhang, Y. The impact of stock price fluctuations on the financial market. *Highlights in Business Economics and Management* 2024, 39, 638–643. <https://doi.org/10.54097/rcy13226>.
- [2] Sundar, S.; Dhyani, B.; Chhajer, P. Factors Affecting Stock Market Movements: An Investors Perspective. *European Economic Letters (EEL)* 2023. <https://doi.org/10.52783/eel.v13i1.172>.
- [3] East, R.; Wright, M. Taming the Animal Spirits: Predicting Psychologically Based Stock Price Movements. *Research Square (Research Square)* 2022. <https://doi.org/10.21203/rs.3.rs-2090235/v1>.
- [4] Jose, N. J.; P, N. V. Integrating Technical Indicators and Ensemble Learning for Predicting the Opening Stock Price. *International Journal of Information Technology Research and Applications* 2024, 3 (2), 1–15. <https://doi.org/10.59461/ijitra.v3i2.96>.
- [5] Ho, T.-T.; Huang, Y. Stock Price Movement Prediction Using Sentiment Analysis and CandleStick Chart Representation. *Sensors* 2021, 21 (23), 7957. <https://doi.org/10.3390/s21237957>.
- [6] Badhan, A. K.; Bhattacharjee, A.; Roy, R. Deep Learning Techniques in Big Data Analytics. In *Studies in big data*; 2024; pp 171–193. https://doi.org/10.1007/978-981-97-0448-4_9.
- [7] Nassima, A. M.; Hanae, S.; Karim, B. Dynamic Decomposition of Monolith Applications Into Microservices Architectures. *2024 Mediterranean Smart Cities Conference (MSCC)*, 2024, 1–4. <https://doi.org/10.1109/mscc62288.2024.10697026>.
- [8] Dimov, A.; Emanuilov, S.; Bontchev, B.; Dankov, Y.; Papapostolu, T. Architectural Approaches to Overcome Challenges in The Development of Data-Intensive Systems. *AHFE International* 2022. <https://doi.org/10.54941/ahfe1002521>.
- [9] Blinowski, G.; Ojdowska, A.; Przybyłek, A. Monolithic vs. Microservice Architecture: A Performance and Scalability Evaluation. *IEEE Access* 2022, 10, 20357–20374. <https://doi.org/10.1109/access.2022.3152803>.
- [10] Oyeniran, N. O. C.; Adewusi, N. A. O.; Adeleke, N. A. G.; Akwawa, N. L. A.; Azubuko, N. C. F. Microservices architecture in cloud-native applications: Design patterns and scalability. *Computer Science & IT Research Journal* 2024, 5 (9), 2107–2124. <https://doi.org/10.51594/csitj.v5i9.1554>.
- [11] Sheng, X.; Hu, S.; Lu, Y. The Micro-service Architecture Design Research of Financial Trading System based on Domain Engineering. *Proceedings of the 2018 International Symposium on Social Science and Management Innovation (SSMI 2018)* 2019. <https://doi.org/10.2991/ssmi-18.2019.32>.
- [12] Yang, N.; Liu, Y.; Lv, W.; Gao, C.; Zheng, S.; Zhang, Q. Application of microservices in power trading platforms. In *CRC Press eBooks*; 2022; pp 176–183. <https://doi.org/10.1201/9781003330165-25>.
- [13] Sithiyopasakul, P.; Piyatananugoon, C.; Chaowalittawin, V.; Krungseanmuang, W.; Sathaporn, P.; Kanjanasurat, I.; Purahong, B.; Archevapanich, T.; Lasakul, A. Inventory Management System based on IoT and Microservices Architecture Design. *2022 International Electrical Engineering Congress (iEECON)* 2023. <https://doi.org/10.1109/ieecon56657.2023.10126548>.
- [14] Yan, W.; Shuai, F. Application of Microservice Architecture in Commodity ERP Financial System. *International Journal of Computer Theory and Engineering* 2022, 14 (4), 168–173. <https://doi.org/10.7763/ijcte.2022.v14.1324>.
- [15] Xu, M.; Dustdar, S.; Villari, M.; Buyya, R. Special issue on efficient management of microservice-based systems and applications. *Software Practice and Experience* 2023, 54 (4), 543–545. <https://doi.org/10.1002/spe.3298>.
- [16] M, D.; Raswanth, S. R.; Chaitanya, V. S. S. K.; Kalavalapalli, V. S. S.; Kumar, P.; Srivastava, G. Secure Automated Inventory Management system using Abstracted System Design Microservice Architecture. *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)* 2023. <https://doi.org/10.1109/icccnt56998.2023.10306924>.
- [17] Sasmita, D.; Kusuma, G. P. Microservices for Asset Tracking Based on Indoor Positioning System. *International Journal of Computing and Digital Systems* 2024, 15 (1), 861–873. <https://doi.org/10.12785/ijcds/160162>.
- [18] Lokesh, D.; Aravind, V.; Vani, V.; Karthik, N. Stock Recommendation System for Better Investment Plan. *2024 International Conference on Signal Processing, Computation, Electronics, Power and Telecommunication (IconSCEPT)*, 2024, 1–6. <https://doi.org/10.1109/iconcept61884.2024.10627901>.
- [19] Ruke, A.; Gaikwad, S.; Yadav, G.; Buchade, A.; Nimbarkar, S.; Sonawane, A. Predictive Analysis of Stock Market Trends: A Machine Learning Approach. *2024 4th International Conference on Data Engineering and Communication Systems (ICDECS)*, 2024, 1–6. <https://doi.org/10.1109/icdecs59733.2023.10503557>.
- [20] Liu, L. A Comparative Examination of Stock Market Prediction: Evaluating Traditional Time Series Analysis Against Deep Learning Approaches. *Advances in Economics Management and Political Sciences* 2023, 55 (1), 196–204. <https://doi.org/10.54254/2754-1169/55/20231008>.
- [21] Lee, M.-C.; Chang, J.-W.; Hung, J. C.; Chen, B.-L. Exploring the Effectiveness of Deep Neural Networks with Technical Analysis Applied to Stock Market Prediction 2021. <https://doi.org/10.2298/CSIS200301002L>.
- [22] Nabipour, M.; Nayyeri, P.; Jabani, H.; S, S.; Mosavi, A. Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data: A Comparative Analysis. *IEEE Access* 2020, 8, 150199–150212. <https://doi.org/10.1109/ACCESS.2020.3015966>.
- [23] Zhang, X. Stock price prediction of PayPal by Linear Regression, SVM, Random Forest and LSTM. *Applied and Computational Engineering* 2024, 52 (1), 201–207. <https://doi.org/10.54254/2755-2721/52/20241568>.
- [24] Nabipour, M.; Nayyeri, P.; Jabani, H.; S, S.; Mosavi, A. Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis. *IEEE Access* 2020, 8, 150199–150212. <https://doi.org/10.1109/access.2020.3015966>.
- [25] Halder, S.; Bagchi, D.; Samanta, A.; Mondal, D.; Samanta, S.; Hati, S.; Kundu, A. Strategic Selection of Machine Learning Models for Short-term Trading Optimization. *International Journal for Multidisciplinary Research* 2024, 6 (3). <https://doi.org/10.36948/ijfmr.2024.v06i03.20529>.
- [26] Kotapati, L.; Sajjala, R. B.; Gudi, S.; Hareesh, B. V. N.; Tokala, S.; Enduri, M. K. Forecasting Stock Markets Trends using Machine Learning Algorithms. *2023 IEEE 15th International Conference on Computational Intelligence and Communication Networks (CICN)*, 2023, 278–282. <https://doi.org/10.1109/cicn59264.2023.10402238>.

- [27] Sonkavde, G.; Dharrao, D. S.; Bongale, A. M.; Deokate, S. T.; Doreswamy, D.; Bhat, S. K. Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications. *International Journal of Financial Studies* 2023, 11 (3), 94. <https://doi.org/10.3390/ijfs11030094>.
- [28] Weng, B.; Ahmed, M. A.; Megahed, F. M. Stock market one-day ahead movement prediction using disparate data sources. *Expert Systems With Applications* 2017, 79, 153–163. <https://doi.org/10.1016/j.eswa.2017.02.041>.
- [29] Gonzales, R. M. D.; Hargreaves, C. A. How can we use artificial intelligence for stock recommendation and risk management? A proposed decision support system. *International Journal of Information Management Data Insights* 2022, 2 (2), 100130. <https://doi.org/10.1016/j.jjimei.2022.100130>.
- [30] Taghavi, M.; Bakhtiyari, K.; Scavino, E. Agent-based computational investing recommender system. *RecSys '13: Proceedings of the 7th ACM Conference on Recommender Systems*, 2013. <https://doi.org/10.1145/2507157.2508072>.
- [31] Sawhney, R.; Agarwal, S.; Wadhwa, A.; Shah, R. R. Deep Attentive Learning for Stock Movement Prediction From Social Media Text and Company Correlations. *Conference on Empirical Methods in Natural Language Processing*, 2020. <https://doi.org/10.18653/v1/2020.emnlp-main.676>.
- [32] Bustos, O.; Pomares-Quimbaya, A. Stock market movement forecast: A Systematic review. *Expert Systems With Applications* 2020, 156, 113464. <https://doi.org/10.1016/j.eswa.2020.113464>.
- [33] Awad, A. L.; Elkaffas, S. M.; Fakh, M. W. Stock Market Prediction Using Deep Reinforcement Learning. *Applied System Innovation* 2023, 6 (6), 106. <https://doi.org/10.3390/asi6060106>.
- [34] Vincent, N.; Saputra, B. J.; Izzatunnisa, S. Y.; Lucky, H.; Iswanto, I. A. Stock Market Prediction System Using LSTM with Technical Indicators as Voters. *2023 4th International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, 2023, 3, 229–234. <https://doi.org/10.1109/aidas60501.2023.10284633>.
- [35] Ramakrishnan, S.; Devi, K. R.; E, Y.; T, V.; C, T. A. Ensemble Algorithm to Speculate Stock Trend by Analyzing Technical Indicators on Historical Data. *2022 International Conference on Inventive Computation Technologies (ICICT)* 2023, 970–975. <https://doi.org/10.1109/iciict57646.2023.10134017>.
- [36] Jose, N. J.; P, N. V. Integrating Technical Indicators and Ensemble Learning for Predicting the Opening Stock Price. *International Journal of Information Technology Research and Applications* 2024, 3 (2), 1–15. <https://doi.org/10.59461/ijitra.v3i2.96>.
- [37] Singh, J.; Khushi, M. Feature Learning for Stock Price Prediction Shows a Significant Role of Analyst Rating. *Applied System Innovation* 2021, 4 (1), 17. <https://doi.org/10.3390/asi4010017>.
- [38] Ifleh, A.; Kabbouri, M. E. Stock price indices prediction combining deep learning algorithms and selected technical indicators based on correlation. *Arab Gulf Journal of Scientific Research* 2023. <https://doi.org/10.1108/agjsr-02-2023-0070>.
- [39] Zouaghia, Z.; Aouina, Z. K.; Said, L. B. Stock Movement Prediction Based On Technical Indicators Applying Hybrid Machine Learning Models. *2022 International Symposium on Networks, Computers and Communications (ISNCC)* 2023, 1823, 1–4. <https://doi.org/10.1109/isncc58260.2023.10323971>.
- [40] Chang, V.; Xu, Q. A.; Chidozie, A.; Wang, H. Predicting Economic Trends and Stock Market Prices with Deep Learning and Advanced Machine Learning Techniques. *Electronics* 2024, 13 (17), 3396. <https://doi.org/10.3390/electronics13173396>.
- [41] Borkar, S. N.; Jadhav, A.; Dhablia, A.; NandkishorAher, R.; Aher, N. D.; Aware, A. A. Selection of Technical Indicators for Stock Market Prediction: Correlation Based Approach. *2022 6th International Conference on Computing, Communication, Control and Automation (ICCUBEA)* 2023, 1–6. <https://doi.org/10.1109/iccubea58933.2023.10391999>.
- [42] Rouf, N.; Malik, M. B.; Arif, T.; Sharma, S.; Singh, S.; Aich, S.; Kim, H.-C. Stock Market Prediction Using Machine Learning Techniques: A Decade Survey on Methodologies, Recent Developments, and Future Directions. *Electronics* 2021, 10 (21), 2717. <https://doi.org/10.3390/electronics10212717>.
- [43] Fei, Y.; Zhou, Y. Intelligent Prediction Model of Shanghai Composite Index Based on Technical Indicators and Big Data Analysis. Highlights in Business Economics and Management 2023, 17, 370–389. <https://doi.org/10.54097/hbem.v17i.11486>.
- [44] Antweiler, W.; Frank, M. Z. Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards. *The Journal of Finance* 2004, 59 (3), 1259–1294. <https://doi.org/10.1111/j.1540-6261.2004.00662.x>.
- [45] Choi, J.; Yoo, S.; Zhou, X.; Kim, Y. Hybrid Information Mixing Module for Stock Movement Prediction. *IEEE Access* 2023, 11, 28781–28790. <https://doi.org/10.1109/ACCESS.2023.3258695>.
- [46] Wu, D.D.; Zheng, L.; Olson, D.L. A Decision Support Approach for Online Stock Forum Sentiment [2] Analysis. *IEEE Trans. Syst. Man Cybern. Syst.* 2014, 44, 1077–1087. <https://doi.org/10.1109/TSMC.2013.2295353>.
- [47] Guan, Z.; Zhao, Y. Optimizing Stock Market Volatility Predictions Based on the SMVF-ANP Approach. *Int. Rev. Econ. Financ.* 2024, 95, 103502. <https://doi.org/10.1016/j.iref.2024.103502>.
- [48] Li, T.; Chen, H.; Liu, W.; Yu, G.; Yu, Y. Understanding the Role of Social Media Sentiment in Identifying Irrational Herding Behavior in the Stock Market. *Int. Rev. Econ. Financ.* 2023, 87, 163–179. <https://doi.org/10.1016/j.iref.2023.04.016>.
- [49] Galphade, M.; Nikam, V.B.; Yedurkar, D.; Singh, P.; Stephan, T. Semantic Analysis Using Deep Learning for Predicting Stock Trends. *Procedia Comput. Sci.* 2024, 235, 820–829. <https://doi.org/10.1016/j.procs.2024.04.078>.
- [50] Yang, J.; Wang, Y.; Li, X. Prediction of Stock Price Direction Using the LASSO-LSTM Model Combines Technical Indicators and Financial Sentiment Analysis. *PeerJ Comput. Sci.* 2022, 8, e1148. <https://doi.org/10.7717/peerj-cs.1148>.

AUTHOR'S PROFILE

ASAAD ALGARNI received his Ph.D. in Software Engineering from North Dakota State University, Fargo, ND, USA. Currently, he serves as an Assistant Professor in the Department of Computer Sciences within the Faculty of Computing and Information Technology at Northern Border University, Saudi Arabia. His research interests encompass software engineering, artificial intelligence, and computer vision applications.

Improved Whale Optimization Algorithm with LSTM for Stock Index Prediction

Yu Sun¹, Sofianita Mutalib^{2*}, Liwei Tian³

School of Management, Guangdong University of Science and Technology, Dongguan, Guangdong Province, China¹

School of Computing Sciences-College of Computing-Informatics and Mathematics,

Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia^{1,2}

School of Computing, Guangdong University of Science and Technology, Dongguan, Guangdong Province, China³

Abstract—After the COVID-19 pandemic, the global economy began to recover. However, stock market fluctuations continue to affect economic stability, making accurate predictions essential. This study proposes an Improved Whale Optimization Algorithm (IWOA) to optimize the parameters of the Long Short-Term Memory (LSTM) model, thereby enhancing stock index predictions. The IWOA improves upon the traditional Whale Optimization Algorithm (WOA) by integrating logistic chaotic mapping to increase population diversity and prevent premature convergence. Additionally, it incorporates a dynamic adjustment mechanism to balance global exploration and local exploitation, thus boosting optimization performance. Experiments conducted on five representative global stock indices demonstrate that the IWOA-LSTM model achieves higher accuracy and reliability compared to WOA-LSTM, LSTM, and RNN models. This highlights its value in predicting complex time-series data and supporting financial decision-making during economic recovery.

Keywords—Long short-term memory network; chaotic mapping; dynamic adjustment mechanism; improved whale optimization algorithm; financial time series forecasting

I. INTRODUCTION

Stock market indices are published by stock exchanges or financial institutions and are important financial indicators that reflect market fluctuations. These indices directly affect investor sentiment and decision-making and serve as key references for investors. Given the significant impact of stock market movements on the global economy, predicting these movements has been a top priority for researchers and investors. Their goal is to develop effective investment strategies and reduce risk. Despite the passage of time, the complex patterns and return metrics that emerge in the stock market within the framework of unsupervised automated prediction remain difficult to predict accurately. This underscores the critical need for a progressive predictive approach that combines human expertise with advanced technological capabilities to improve the accuracy and reliability of stock and economic forecasts [1].

In recent decades, the adoption of machine learning techniques and metaheuristic algorithms in financial time series forecasting has attracted significant attention. Conventional neural networks, such as Recurrent Neural Networks (RNNs), have been shown to be effective in capturing complex fluctuations in the stock market. Chen et al. introduced a deep learning prediction model based on RNN, which integrates social media news content (sentiment and topic features) with

technical indicators to strengthen the predictive accuracy of stock market volatility [2]. Haromainy et al. utilized a genetic algorithm-optimized RNN model to predict stock trends for stock prices, demonstrating its ability to capture nonlinear features in stock market data [3]. Additionally, Zhao et al. incorporated fuzzy logic with RNN to improve stock market volatility prediction [4].

Despite the advantages of RNN models, prediction accuracy remains limited due to the high nonlinearity and chaotic nature of the stock market. As deep learning continues to grow, LSTM network, known for its superior time series processing capabilities, has become a prominent research focus. Abdul Quadir et al. utilized the LSTM algorithm to analyze normalized time series data, addressing the vanishing gradient issue observed in simpler RNN and determining the relationship between historical and future values [5]. In the medical field, Academician Zhong Nanshan's team utilized an LSTM-based recurrent neural network to study and predict the peaks and sizes of COVID-19 [6]. In the domain of energy consumption forecasting, LSTM-based methods have demonstrated outstanding predictive performance [7]. Notably, in the research area of financial time series forecasting, LSTM has shown remarkable predictive performance. Li et al. further validated the potential of LSTM in capturing complex stock price patterns and improving prediction accuracy [8]. Singh et al. combined Convolutional Neural Networks (CNN) with LSTM to propose a hybrid model for Indian stock portfolio management, which outperformed traditional models [9].

However, LSTM networks also face challenges, such as susceptibility to overfitting, sensitivity to hyperparameter selection, and the risk of falling into local optima, which can limit prediction accuracy [10]. Hyperparameter selection in LSTM models typically relies on manual experience, which significantly impacts model performance. To address these issues, metaheuristic optimization algorithms have been extensively utilized for hyperparameter optimization [11–13]. For example, Zhang et al. applied Particle Swarm Optimization (PSO) to optimize LSTM hyperparameters for predicting short-term fluctuations in the highest prices of U.S. stocks, demonstrating the superiority of the optimized LSTM model [14]. However, the PSO method is prone to slow convergence and inefficiency in high-dimensional spaces, particularly in multimodal optimization problems or complex search spaces, leading to local optima and reduced operational efficiency [15]. In response to these challenges, several studies have explored

*Corresponding Author.

the integration of the Whale Optimization Algorithm (WOA) with deep learning models. Xin et al. combined the WOA with the LSTM model to predict the stock market, achieving significant improvements in forecasting accuracy [16]. Hasan enhanced the convergence mechanism of WOA and applied it to the optimization of time-series prediction models, yielding excellent results in temperature and humidity forecasting [17].

Inspired by the aforementioned studies, this paper proposes a novel approach that integrates an Improved Whale Optimization Algorithm (IWOA) with LSTM networks. IWOA enhances population diversity and reduces the risk of premature convergence by incorporating chaotic mapping. Additionally, it features an adaptive factor mechanism that dynamically balances exploration and exploitation during the optimization process, leading to improved efficiency and accuracy. By leveraging IWOA's robust optimization capabilities, the hyperparameters of the LSTM model are fine-tuned more effectively, enabling the model to better capture the complexity of nonlinear and chaotic patterns in financial time series.

II. METHODS

A. LSTM

In the early stages of time series forecasting, recurrent neural networks (RNNs) gained widespread use due to their capability to process sequential data. Unlike traditional feedforward neural networks (e.g., backpropagation neural networks, BPNN), which propagate signals in a single direction, RNNs introduce weighted connections between hidden layer neurons. This architecture enables the output of hidden layer neurons at each time step to depend on information from the previous time step, allowing the network to effectively capture temporal dependencies. By incorporating both feedforward and internal feedback connections, RNNs exhibit dynamic temporal behavior that influences their internal states. However, in practice, the hidden state of an RNN at each time step is determined by both the hidden layer values from the previous time step and the input values at the current time step, which restricts its ability to retain long-term memory [18].

To overcome the limitations of traditional RNNs, Graves extended the Long Short-Term Memory (LSTM) neural network, which effectively addresses these challenges [19]. LSTM replaces the hidden layer nodes of standard RNNs with specialized memory units, allowing the network to better retain and manage temporal information, particularly for modeling long-term dependencies. The core component of LSTM is the cell state, which functions as a channel for transmitting information across the network. LSTM introduces input, forget, and output gates to enhance the functionality of global memory cells. These gates regulate the retention, updating, or discarding of information at each time step, enabling the network to efficiently learn long-term dependencies. Fig. 1 illustrates the architecture of the LSTM model.

The structure of LSTM is highly efficient in managing long-term relationships within time series data, especially when events are delayed. In LSTM, three gates regulate the cell state, each employing a Sigmoid activation function and

pointwise multiplication. The Sigmoid output, ranging from 0 to 1, determines how much information passes through: a value of 0 indicates complete "blocking," while a value of 1 signifies full "pass-through." This gating mechanism enables LSTM to efficiently retain and propagate long-term dependency information in time series data. The workflow of the LSTM network is described by the following equations.

$$i_t = \sigma(W_i * [h_{t-1}, X_t] + b_i) \quad (1)$$

$$f_t = \sigma(W_f * [h_{t-1}, X_t] + b_f) \quad (2)$$

$$o_t = \sigma(W_o * [h_{t-1}, X_t] + b_o) \quad (3)$$

$$\hat{C}_t = \tanh(W_c * [h_{t-1}, X_t] + b_c) \quad (4)$$

$$C_t = f_t * C_{t-1} + i_t * \hat{C}_t \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6)$$

In Eq. (1) to Eq. (6), W refers to the weight vectors, while b denotes the bias terms.

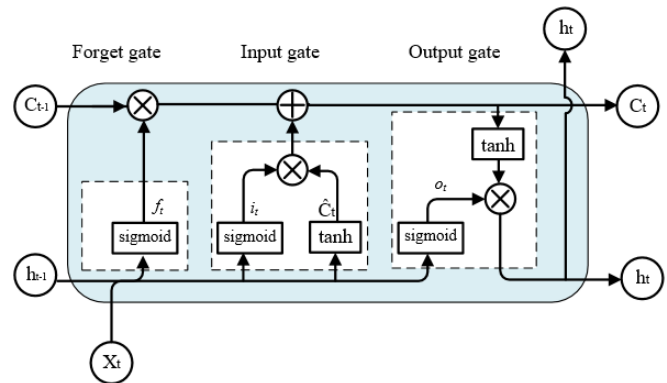


Fig. 1. Structure of LSTM.

B. WOA

The Whale Optimization Algorithm (WOA) is a novel swarm intelligence method inspired by the bubble net foraging strategy of humpback whales. This algorithm demonstrates superior performance compared to traditional optimization approaches. WOA simulates this behavior through two main search strategies: exploration and exploitation. During exploration, whales move randomly in search of prey, while in the exploitation phase, they navigate toward the prey in a spiral pattern to find the optimal solution. The algorithm uses a series of mathematical equations to simulate these behaviors, enabling it to effectively search for optimal solutions in complex and high-dimensional spaces [20]. In WOA, each candidate solution corresponds to a position within the search space, and the algorithm optimizes the objective function by mimicking the whale's hunting behavior. The algorithm operates through three main phases [21].

1) *Encircling prey*: When a whale detects the position of its prey, it adjusts its position based on the current best solution. During this process, the whale moves closer to the optimal solution by a certain proportion. The position of the whale is updated as described in Eq. (7).

$$\begin{cases} D = |C \cdot X^*(t) - X(t)| \\ X(t+1) = X^*(t) - A \cdot D \\ A = 2a \cdot r - a \\ C = 2 \cdot r \end{cases} \quad (7)$$

In Eq. (7), D represents the distance between the whale individual and its prey. t represents the iteration number; A and C are coefficient vectors; X and X^* are the current whale position and the current best whale position; a decreases linearly from 2 to 0 during the iteration; r is a random number in the range $[0, 1]$.

2) *Bubble-net attacking*: There are two mechanisms designed for Bubble-net attacking: shrinking encircling mechanism and spiral updating position.

a) *Shrinking encircling mechanism*: This mechanism is implemented using the parameters in Eq. (7). During the iterations, the behavior is achieved by linearly decreasing the value of a from 2 to 0, while A fluctuates within the range $[-a, a]$. When A is a random value between $[-1, 1]$, the whale's position is updated to lie somewhere between its original position and the current optimal position.

b) *Spiral updating position*: Initially, the distance between the whale and its prey is calculated. Then, a spiral equation is derived to simulate the whale's spiral movement, the specific equation is as follows.

$$\begin{cases} D = |X^*(t) - X(t)| \\ X(t+1) = D \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) \end{cases} \quad (8)$$

In Eq. (8), b is the spiral shape constant; l is a random value within the range $[-1, 1]$.

When the whale approaches the prey, its behaviors of shrinking encircling and spiral position updating occur simultaneously. To simulate this bubble-net attack, it is assumed that the humpback whale has a 50% chance of performing either shrinking encircling or spiral position updating. $X(t+1)$ is shown in Eq. (9).

$$X(t+1) = \begin{cases} X^*(t) - A \cdot D & p < 0.5 \\ D \cdot e^{bl} \cdot \cos(2\pi l) + X^*(t) & p \geq 0.5 \end{cases} \quad (9)$$

In Eq. (9), p is a random number in the range $[0, 1]$.

3) *Search for prey*: At this phase, the whale population performs global exploration. When $|A| > 1$, the whale population ceases to adjust its position according to the current optimal solution. Instead, the position is updated based on a randomly selected whale, with the goal of expanding the search range and seeking the optimal solution to maintain population diversity. Therefore, only a small modification to Eq. (7) is needed to obtain the mathematical model for this stage.

$$\begin{cases} D = |C \cdot X_{rand} - X(t)| \\ X(t+1) = X_{rand} - A \cdot D \end{cases} \quad (10)$$

In Eq. (10), X_{rand} represents the position of a randomly selected whale.

WOA achieves a balance between local and global exploration through its three primary operations, thereby

enhancing global search capabilities and mitigating the premature convergence problem commonly encountered in traditional optimization algorithms. As a result, WOA demonstrates superior performance in solving complex optimization problems [22].

III. PROPOSED IWOA-LSTM PREDICTION MODEL

A. IWOA-LSTM Model

The performance of an LSTM network is highly influenced by the selection of appropriate hyperparameters. However, traditional hyperparameter tuning methods often rely on manual expertise, which may not ensure optimal results across diverse scenarios [23]. To address this issue, this research proposes an improved Whale Optimization Algorithm (IWOA) integrated with the LSTM model to automate the hyperparameter optimization process.

IWOA combines Whale Optimization with Logistic Chaos Mapping to improve population diversity during initialization and prevent premature convergence to local optima, providing a more effective starting point for the optimization process. Additionally, IWOA incorporates a dynamic adjustment mechanism based on fitness variations, enabling the automatic fine-tuning of the mutation factor. This mechanism enhances the algorithm's global search capability and increases optimization efficiency.

In the IWOA-LSTM model, IWOA optimizes key LSTM hyperparameters, including the number of hidden units and the learning rate. By balancing global exploration and local exploitation, IWOA accelerates convergence and fine-tunes LSTM parameters, ultimately improving prediction accuracy. This approach minimizes manual intervention, enhancing the adaptability and performance of the LSTM network.

B. Chaotic Map Initialization

In WOA, the initialization of the population is a crucial factor that influences the optimization performance. Traditional random initialization can result in an uneven distribution of initial solutions, potentially reducing the algorithm's efficiency. To address this challenge, chaotic mapping is employed to adjust the control parameters of WOA, thereby improving the balance between exploration and exploitation. This technique enhances the algorithm's global convergence rate by generating a more dispersed and diversified initial population, thus reducing the likelihood of premature convergence to local optima.

In this study, chaotic mapping is introduced during the initialization phase of WOA, specifically utilizing the logistic chaotic mapping proposed by Prasad [24] and Yousri [25]. This mapping exhibits both random and deterministic characteristics, which facilitates the adjustment of WOA's control parameters. The initialization equation for chaotic mapping is as follows.

$$x_{i,j} = x_{i,j} * (b_j - a_j) + a_j \quad (11)$$

In Eq. (11), $x_{i,j}$ represents the position of the i -th whale in the j -th dimension. a_j and b_j denote the boundaries of the j -th dimensional space. The initialized $x_{i,j}$ is generated using the

Logistic chaotic mapping, and its iteration equation is as follows.

$$x = r * x * (1 - x) \quad (12)$$

In Eq. (12), r is the control parameter. In this study, its value is set to 4. This initialization method ensures the diversity of the population distribution, providing a strong starting point for subsequent iterations.

C. Dynamic Adjustment Mechanism

In the optimization process of WOA, applying an appropriate mutation operation after each individual's position update significantly enhances the algorithm's global search capability [26]. The mutation factor plays a pivotal role in this process. By dynamically adjusting the mutation factor based on changes in fitness and iteration step size, the algorithm effectively balances global exploration and local exploitation, thereby improving both convergence efficiency and solution accuracy.

To enable the dynamic adjustment of the mutation factor, this study proposes a fitness-based adjustment mechanism. This mechanism takes into account the current range of fitness changes, the step size, and the global optimal fitness value. By modulating the mutation factor, the algorithm achieves enhanced diversity and convergence performance. The dynamic adjustment factor is mathematically defined in Eq. (13).

$$\mu_t = \mu_{t-1} * \left(1 + \beta * \frac{|f_t - f_{t-1}|}{f_{max}} + \gamma * \sqrt{|f_t - f_{t-1}|} - \alpha * \frac{t}{T} \right) \quad (13)$$

In Eq. (13), μ_t represents the adjustment factor for the current iteration, while μ_{t-1} is the adjustment factor for the previous iteration. f_t and f_{t-1} are the fitness values for the current and previous iterations, respectively. β , γ , and α are hyperparameters that control the update of the adjustment factor. These parameters are used to measure the linear and nonlinear weights of the fitness change, as well as the effect of the time step size. In this study, they are set to 0.1, 0.2, and 0.05, respectively. t is the current iteration number, and T is the maximum number of iterations. μ_t is constrained within the range (0, 1).

After each position update, to boost the algorithm's global exploration capability, this study introduces a mutation operation based on the dynamic adjustment factor. After calculating the dynamic adjustment factor μ_t , a normal distribution random disturbance $N(0,1)$ is applied to fine-tune the individual positions. This disturbance operation not only enhances population diversity but also effectively prevents individuals from getting trapped in local optima, especially in the later stages of optimization. The specific position update equation is presented in Eq. (14).

$$x_{i,j}(t + 1) = x_{i,j}(t) + \mu_t * N(0,1) \quad (14)$$

This operation enhances the algorithm's ability to avoid local optima by introducing random disturbances, especially during the later stages of population convergence or optimization. The dynamic adjustment factor integrates fitness differences with a square root term, significantly increasing the

variation intensity during the early stages of optimization, when fitness differences are more pronounced. This approach introduces greater randomness into the search process, fostering a more thorough exploration of the global solution space.

As iterations progress, the influence of the adjustment factor is gradually reduced through a time attenuation mechanism, enabling the algorithm to transition from global exploration to local exploitation. This shift allows the algorithm to focus on refining the solution, thereby improving optimization accuracy.

Furthermore, the recursive calculation of the dynamic adjustment factor relies on the value from the previous generation. This design ensures smoother variation intensity, preventing abrupt fluctuations during the optimization process and maintaining stability. With this adaptive adjustment mechanism, the algorithm can adjust its search strategy in real-time based on changes in fitness, enabling it to maintain robust global search capabilities while avoiding premature convergence. Ultimately, this improves both optimization efficiency and solution quality.

D. The Construction Steps of IWOA-LSTM

This study employs the IWOA to optimize the LSTM neural network for predicting financial time series. Fig. 2 illustrates the overall process, with the steps detailed below:

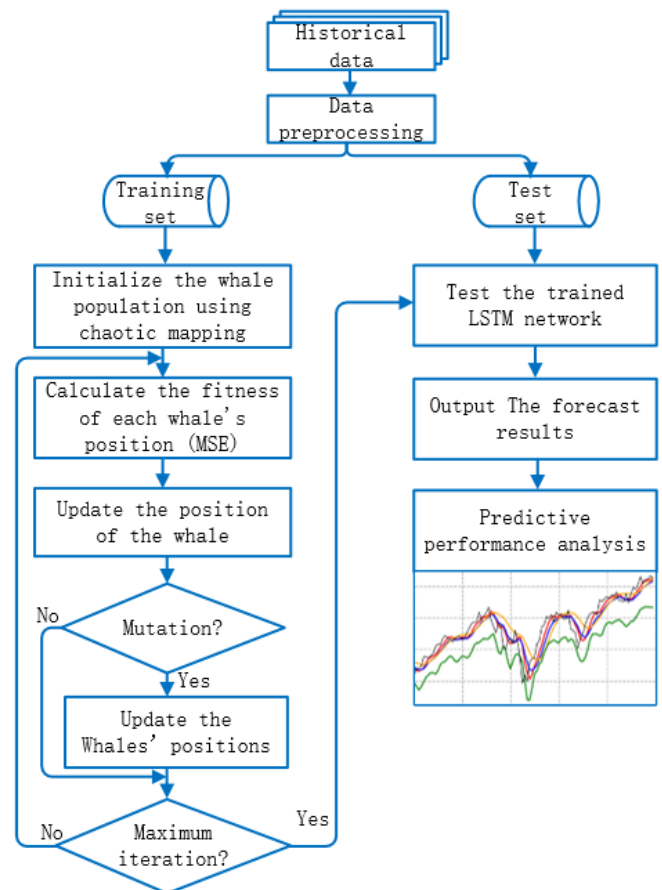


Fig. 2. Flowchart of IWOA-LSTM.

Step 1: Data preprocessing. Normalize the original time series data adopting Min-Max scaling to bring it into the [0, 1], then partition the data into training and test sets.

Step 2: Initialize IWOA and LSTM parameters. For IWOA, set the population size, search space boundaries (for LSTM's hidden units and learning rate), maximum iterations, and dynamic adjustment parameters (β , γ , α). Use the logistic map for chaotic initialization to enhance population diversity. For the LSTM model, set the initial parameters to ensure smooth optimization.

Step 3: Train and optimize the LSTM model using IWOA by minimizing the mean squared error (MSE) through iterative updates. Evaluate the fitness of each whale in the swarm, identifying the best fitness values for both individual whales and the global solution. Use IWOA's position update formula and dynamic adjustment factor to refine whale positions, balancing global exploration and local exploitation. Continue the process until the maximum iteration threshold is reached or the global fitness threshold is achieved.

Step 4: Apply the fine-tuned LSTM model to the test data to validate its prediction capabilities.

Step 5: Evaluate the prediction performance by applying the optimized LSTM model to forecast the financial time series data. The results are then measured against five performance indicators to evaluate the model's accuracy and effectiveness.

IV. EXPERIMENTS

All programming works are implemented in the Python 3.9 environment. The computational experiments are performed on a system featuring a 12th Gen Intel(R) Core(TM) i5-1235U CPU, 16 GB of RAM, and operating on Windows 10.

A. Dataset

This study selected five globally representative stock indices as research subjects: the S&P 500 Index (Code: ^GSPC), Dow Jones Industrial Average (Code: ^DJI), FTSE 100 Index (Code: ^FTSE), Nasdaq Composite (Code: ^IXIC), and Shanghai Composite Index (Code: 000001.SS). These indices represent major economies in the Americas, Europe,

and Asia, providing comprehensive insights into the global capital market.

Specifically, the S&P 500 Index and the Dow Jones Industrial Average represent the U.S. capital market comprehensively. The S&P 500 Index includes 500 companies with the largest market capitalizations and significant industry representation, providing a broad view of the market. In contrast, the Dow Jones Industrial Average Index focuses on 30 prominent companies, often referred to as blue-chip stocks, representing major sectors of the U.S. economy. The FTSE 100 Index reflects the performance of the largest British companies by market capitalization and serves as a key benchmark for the European market. The Nasdaq Composite Index, heavily weighted by technology stocks, highlights global trends in technological innovation and growth. Lastly, the Shanghai Composite Index is a vital indicator of mainland China's capital market, encompassing all listed stocks on the Shanghai Stock Exchange.

The experimental data is obtained from <https://finance.yahoo.com/> and spans a period of 10 years, from October 23, 2014, to October 21, 2024. This timeframe encompasses significant global economic events, including the financial crisis triggered by the COVID-19 pandemic in 2020, subsequent recovery phases, and various national policy adjustments, which resulted in substantial market fluctuations. The dataset includes daily recorded basic information, providing a solid foundation for model training and testing. These data facilitate the assessment of the model's adaptability and robustness within a complex and volatile market environment. In addition, to verify the predictive ability of the model when external factors (e.g., macroeconomic indicators, social sentiment) are not introduced, the experimental design uses only historical price data for prediction. This design highlights the model's intrinsic performance in financial time series analysis, avoids interference from external variables, and objectively evaluates the performance of IWOA-LSTM in volatile market environments. The historical trends of the five stock indices are shown in Fig. 3. The detailed statistical analysis of the closing prices for each stock index is presented in Table I.

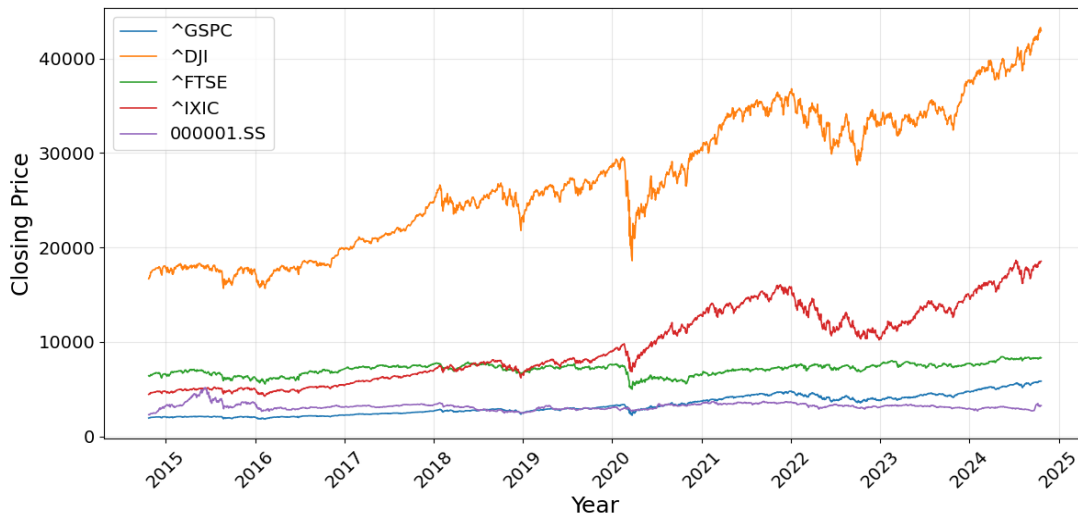


Fig. 3. Historical closing price charts for five stock indices.

TABLE I. DESCRIPTIVE STATISTICS OF STOCK CLOSING PRICES

Name	Count	Max	Min	Range	Mean	Std	Kurtosis	Skewness	Normality Test Statistic	Normality Test P-Value
^GSPC	2515	5864.67	1829.08	4035.59	3280.57	1035.14	-0.88	0.49	403.10	0.00
^DJI	2515	43275.91	15660.18	27615.73	27194.68	7082.78	-1.12	0.13	1129.42	0.00
^FTSE	2524	8445.80	4993.90	3451.90	7109.40	601.09	-0.02	-0.49	92.50	0.00
^IXIC	2515	18647.45	4266.84	14380.61	9573.88	3959.34	-1.11	0.41	1101.31	0.00
000001.SS	2427	5166.35	2290.44	2875.91	3183.23	349.92	5.52	1.45	782.03	0.00

From Fig. 3 and Table I, it is evident that the closing prices of these five stock indices (^GSPC, ^DJI, ^FTSE, ^IXIC, and 000001.SS) exhibit obvious nonlinear characteristics and high volatility, with the price trends showing irregularity and substantial noise. For example, the fluctuation range of ^DJI index is as high as 27615.73, which is significantly higher than that of the other indices, indicating extreme volatility. Although some trend changes are observed in the data, the overall price fluctuation demonstrates strong uncertainty, and the fluctuation pattern cannot be simply summarized as a linear relationship. The price fluctuations of these indices are influenced by a variety of complex factors, displaying both randomness and nonlinear characteristics. Traditional linear models struggle to effectively capture these intricate dynamics.

Conventional forecasting approaches face significant limitations in addressing the nonlinear and complex nature of stock market data, particularly in highly volatile time series. These methods, based on linear assumptions, are inadequate for capturing long-term dependencies and nonlinear trends. To overcome these challenges, this study employs the LSTM model, which excels at handling complex time series data due to its unique architecture and memory mechanism. By capturing long-term dependencies and nonlinear patterns, LSTM outperforms traditional methods, offering improved prediction accuracy and stability in the face of noise and complexity in stock market data.

B. Data Preprocessing

Data preprocessing is a critical component of data analysis, as it significantly influences model performance and prediction accuracy. In this study, the raw stock index data are cleaned to eliminate null and duplicate values, ensuring the integrity and reliability of the dataset. Subsequently, the Min-Max normalization technique is applied to scale the daily closing prices to the range [0, 1]. This normalization step mitigates the impact of varying data dimensions on model training. The equation used for normalization is in Eq. (15).

$$x_i = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (15)$$

In Eq. (15), x_i denotes the normalized data, while x_{max} and x_{min} represent the maximum and minimum value in the financial time series, respectively.

After the prediction is completed, the prediction results need to be transformed using inverse normalization. The inverse normalization equation is in Eq. (16).

$$x = (x_{max} - x_{min}) x_i + x_{min} \quad (16)$$

We prepare the data and normalize it to ensure uniform scaling and provide a stable input for the mode training.

C. Evaluation Metrics

Measuring prediction accuracy is a multifaceted task, and no single evaluation metric applies universally across different application scenarios. To comprehensively assess the model's prediction performance, this study employs five widely used evaluation metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Coefficient of Determination (R^2), and Explained Variance Score (EVS).

These evaluation metrics can be classified into error measures and fitting measures. Error measures, including RMSE, MAE, and MAPE, primarily assess the differences between predicted and actual values. Fitting measures, including R^2 and EVS, evaluate the model's fit. Specifically, the closer the R^2 value is to 1, the better the model's fit, while the closer the EVS value is to 1, the stronger the model's ability to explain the data. These five metrics together offer a comprehensive and accurate reflection of the model's prediction accuracy and fitting ability. The five evaluation metrics are expressed by the following equations:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (17)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (18)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (19)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (20)$$

$$EVS = 1 - \frac{Var(y_i - \hat{y}_i)}{Var(y)} \quad (21)$$

Among them, y_i stands for the true value, \hat{y}_i represents the predicted value, \bar{y}_i denotes the mean of the financial time series, and $Var()$ refers to the variance.

To construct the input data for the LSTM model, this study employs sliding window method for time series and evaluates the impact of different time step settings on model's effectiveness. Specifically, the sliding window method uses fixed time steps to extract sequential features from stock data.

Four different time step settings are tested: 10, 20, 30, and 60. The MSE and training time of the model are calculated for each setting.

The closing price data are divided into training and test sets based on an 80:20 ratio, using various time step configurations (time_steps = 10, 20, 30, and 60). The LSTM model is trained on the training set and generates predictions for the test set. For each time step configuration, the training time (in seconds) and the MSE on the test set are recorded. Table II presents an overview of the experimental results.

TABLE II. IMPACT OF LSTM TIME STEP ON PREDICTION PERFORMANCE

Time Steps	MSE	Training Time (s)
10	0.000463	9.974107
20	0.000577	12.020250
30	0.000701	13.437092
60	0.000464	20.137358

TABLE III. MODELS COMPARED IN THIS RESEARCH

Model	Description of model parameters
IWOA-LSTM	time_steps=10, epochs=50, batch_size=32, population_size=5, units ∈ [10, 100], learning_rate ∈ [0.0001, 0.01]
WOA-LSTM	time_steps=10, epochs=50, batch_size=32, population_size=5, units ∈ [10, 100], learning_rate ∈ [0.0001, 0.01]
LSTM	time_steps=10, units=50, learning_rate=0.001, epochs=50, Dropout=0.2, batch_size=32
RNN	time_steps=10, units=50, learning_rate=0.001, epochs=50, Dropout=0.2, batch_size=32

D. S&P500 Forecasting

This study first selects the S&P 500 index (^GSPC) as the target for prediction. The dataset contains 2,515 records, spanning from October 23, 2014, to October 21, 2024, with daily closing prices. For model training, a retrospective period of 10 days is used, meaning that the closing prices from the past 10 days are employed in the prediction process. To thoroughly assess the prediction performance of each model, this study adopts five different evaluation indicators. Each model is evaluated based on these five indicators, with the best performance metrics marked in bold. The prediction comparison results for the various models on the S&P 500 index are presented in Table IV.

TABLE IV. COMPARISON RESULTS OF PREDICTIONS BETWEEN VARIOUS MODELS OF ^GSPC

Model	RMSE	MAE	MAPE	R2	EVS
IWOA-LSTM	48.1826	37.8081	0.9377%	0.9935	0.9936
WOA-LSTM	55.6435	44.2542	0.9638%	0.9913	0.9916
LSTM	70.4543	55.8142	1.2216%	0.9860	0.9863
RNN	93.7222	76.6777	1.5764%	0.9752	0.9869

As observed from the results in Table IV, the effectiveness of the IWOA-LSTM model is significantly superior to that of the other models across all evaluation metrics. Specifically, the IWOA-LSTM model outperforms WOA-LSTM in RMSE, MAE, MAPE, R², and EVS by 13.41%, 14.57%, 0.03%, 0.22%, and 0.20%, respectively. Furthermore, IWOA-LSTM shows

The experimental results reveal significant variations in both the MSE and the training duration of the model across different time step configurations. Specifically, when the time step is 10 days, the MSE of the model is 0.000463, and the training time is approximately 9.97 seconds. When the time step is increased to 60 days, the MSE slightly increases to 0.000464, while the training time rises significantly to about 20.14 seconds. Although the MSE at 60 days is similar to that at 10 days, the longer training time notably affects computational efficiency. Therefore, considering the need to balance prediction accuracy with computational efficiency, a 10-day time step is selected as the optimal configuration.

Next, the IWOA algorithm is applied to optimize the number of LSTM units and the learning rate. By leveraging IWOA, the model can dynamically fine-tune these hyperparameters, enhancing both prediction accuracy and computational efficiency. To evaluate the effectiveness of IWOA optimization, this study compares it with other popular models, including WOA-LSTM, LSTM, and RNN. Table III shows the models compared in this study and their parameters.

improvements over the LSTM model by 31.61%, 32.26%, 0.28%, 0.76%, and 0.74%, and is 48.59%, 50.69%, 0.64%, 1.88%, and 0.68% better than the RNN model in the same metrics.

In addition to evaluating prediction accuracy, this study also investigates the computational efficiency of the IWOA-LSTM model by comparing its runtime performance with that of the standard WOA during hyperparameter optimization and model training. Fig. 4 presents the comparison of their runtime performance.

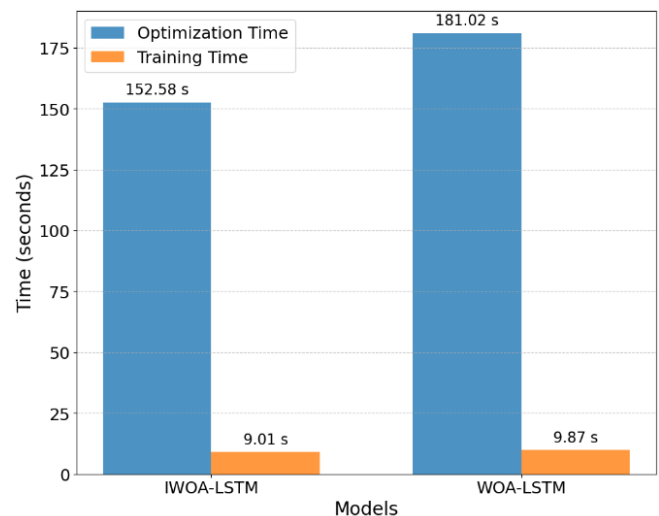


Fig. 4. Runtime comparison.

Fig. 4 compares the optimization and training times of the IWOA-LSTM and WOA-LSTM models, highlighting the performance differences. The optimization time for IWOA-LSTM is 152.58 seconds, while WOA-LSTM takes 181.02 seconds, indicating that IWOA-LSTM is more efficient in parameter optimization. This improvement is attributed to the use of logistic chaotic mapping and a dynamic adjustment factor mechanism, which enhance search precision and reduce unnecessary iterations. For training time, IWOA-LSTM takes 9.01 seconds, while WOA-LSTM takes 9.87 seconds. Although the difference is small, IWOA-LSTM shows a slight advantage, proving that improved optimization efficiency does not slow down the training process. The results show that IWOA-LSTM improves operational efficiency while maintaining predictive performance and is able to converge on complex datasets relatively quickly.

Overall, the results of these experiments on the ^GSPC stock index show that the IWOA-LSTM model performs well across all evaluation metrics. In addition to its excellent performance on these metrics, the IWOA-LSTM model also shows an advantage in hyperparameter optimization time compared to the original WOA, reflecting an improvement in the model's computational efficiency during the optimization process. This highlights the effectiveness of optimizing the LSTM hyperparameters, which not only leads to more accurate predictions but also improves computational efficiency.

Next, to provide a more intuitive view of the fitting performance of each model, the final fitting comparison results are presented in Fig. 5 and Fig. 6.

Fig. 5 demonstrates the fitting results of the ^GSPC index prediction based on different models. The curves in the figure illustrate the prediction results of the IWOA-LSTM, WOA-LSTM, LSTM, and RNN models, respectively, and are compared with the actual price curves. It can be observed that the prediction curves of the IWOA-LSTM model are closest to the actual prices, especially in most time intervals, where the prediction curves of the IWOA-LSTM almost coincide with the actual price curves, showing very high prediction accuracy. In contrast, although the WOA-LSTM model also shows relatively accurate predictions, the discrepancy between the predicted and actual values becomes more pronounced during periods of high volatility. Nonetheless, despite slight deviations in more volatile markets, the overall performance remains robust, suggesting that the model is able to effectively capture the general trend. The LSTM and RNN models show relatively lower prediction accuracies, especially in regions of high price volatility, where the predicted curves of both models deviate more from the actual prices. This suggests that the LSTM and RNN models are not as effective as the IWOA-LSTM and WOA-LSTM models in capturing complex patterns.

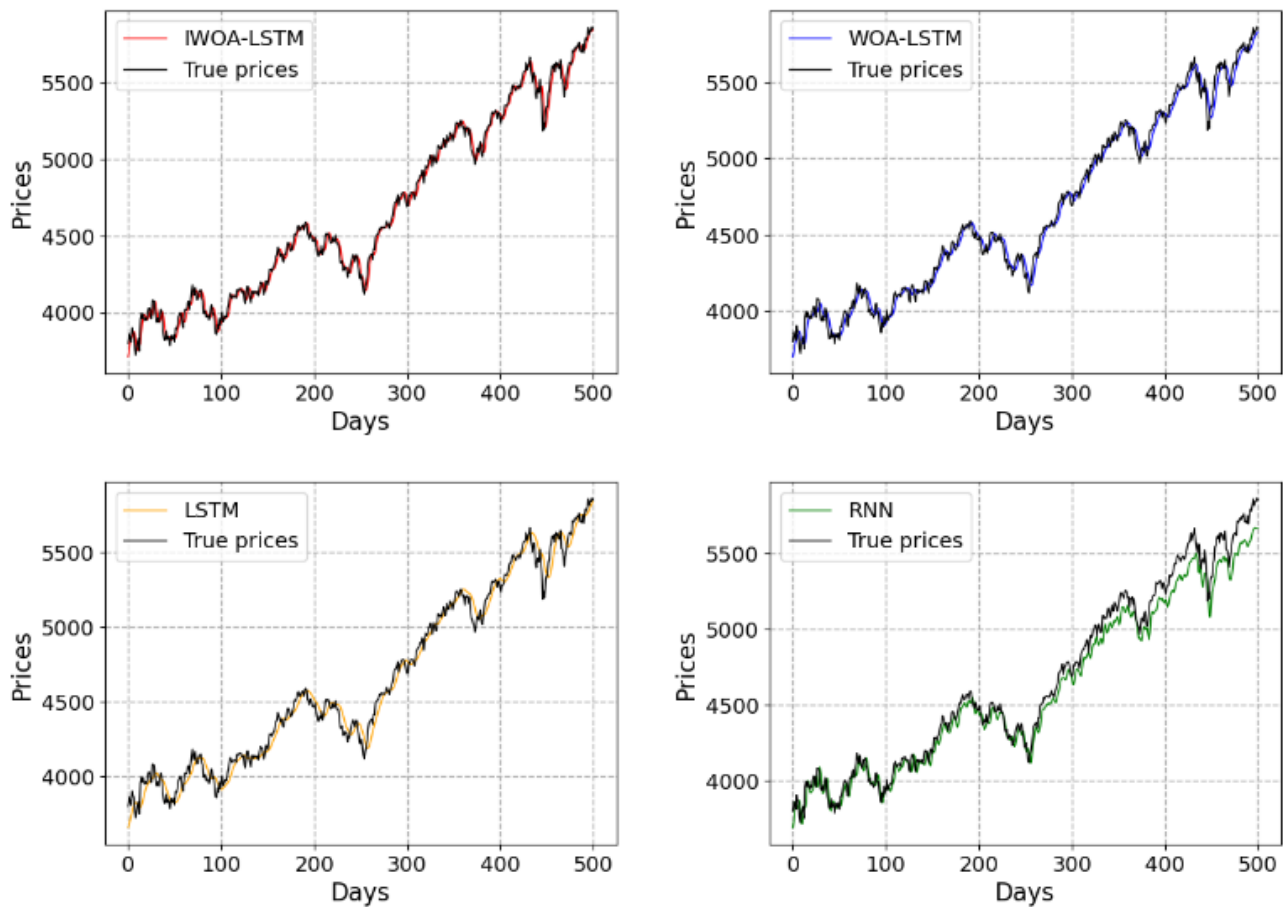


Fig. 5. Individual forecasts of four models on the ^GSPC index.

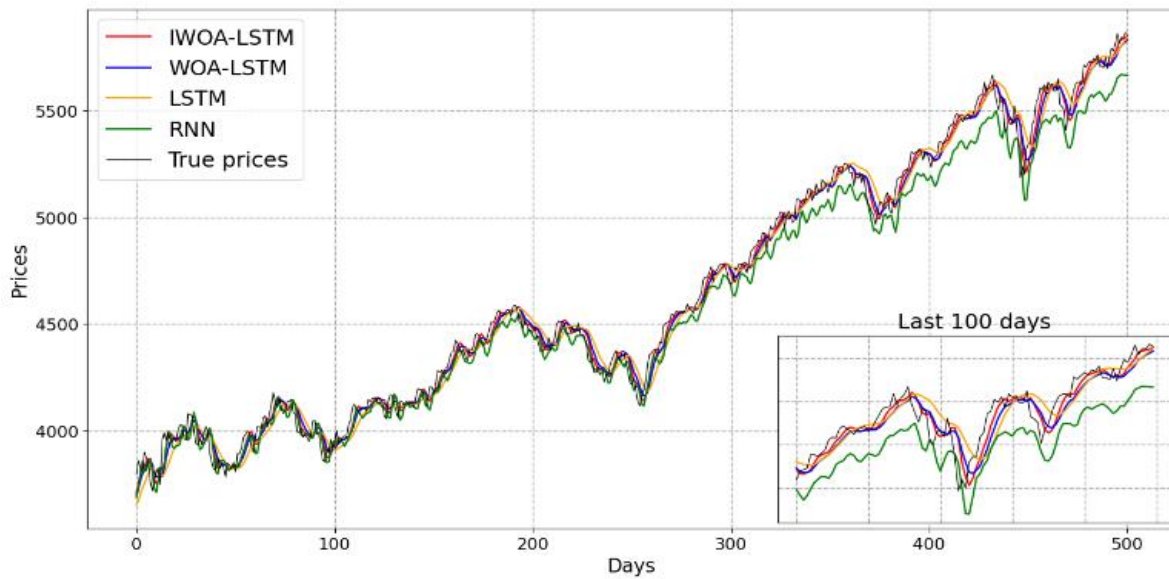


Fig. 6. Comparison of forecasts from four models on the ^GSPC index.

The illustration in Fig. 6 shows the forecast results for the last 100 days. During this period, the IWOA-LSTM prediction curve continues to outperform the other models, especially in short-term price fluctuations, where IWOA-LSTM more accurately captures the trend of stock prices. In contrast, the fitting effect of the RNN model is poor, particularly during periods of significant stock price fluctuations, where its prediction error is quite noticeable.

Overall, IWOA-LSTM performs particularly well in the ^GSPC index prediction task. It is better at capturing the complex dynamic patterns in the time series and providing high-precision prediction results, which demonstrates its advantages in financial time series forecasting.

To further validate the predictive ability of the IWOA-LSTM model, this study compares it with other improved WOA algorithms proposed by Shao [27] and Guan [28]. Apart from the improved algorithms discussed in the paper, all other parameter settings are consistent with those used in this study. Additionally, this study incorporates PSO-LSTM [29] and GA-LSTM [30], two advanced swarm intelligence optimization models. A comparative analysis is performed to evaluate the prediction performance of different optimization methods on the ^GSPC index. The comparison of prediction performance results is presented in Table V.

TABLE V. COMPARISON OF PREDICTION PERFORMANCE OF IWOA-LSTM MODEL AND OTHER IMPROVED LSTM MODEL

Model	Optimization Details	RMSE	MAE	R2
IWOA-LSTM (Proposed in this study)	Logistic chaotic mapping; dynamic adjustment factor mechanism	48.1826	37.8081	0.9935
IWOA-LSTM [27]	Tent chaotic mapping; adaptive weight	57.7395	47.2308	0.9906
WOA-BiLSTM [28]	Whale optimization algorithm optimized Bidirectional LSTM	65.0222	53.2598	0.9881
PSO-LSTM [29]	Particle swarm optimized LSTM	49.1197	38.8116	0.9931
GA-LSTM [30]	Genetic algorithm optimized LSTM	52.7665	42.0509	0.9921

From Table V, the IWOA-LSTM model achieves better prediction results compared to other models. With the integration of the logistic chaotic map and dynamic adjustment mechanism, IWOA-LSTM demonstrates improved prediction accuracy and stability. Although the other models have been enhanced, their accuracy and effectiveness in predicting financial time series do not fully match those of IWOA-LSTM. This suggests that IWOA-LSTM offers certain advantages in forecasting financial time series.

E. The Robustness and Reliability Verification

In order to conduct a deeper validation of the IWOA-LSTM model's robustness and reliability, this study performs experiments on other four famous stock indices: ^DJI, ^FTSE,

^IXIC, and 000001.SS. The corresponding experimental results are presented in Table VI to Table IX and Fig. 7 to Fig. 10.

TABLE VI. COMPARISON RESULTS OF PREDICTIONS BETWEEN VARIOUS MODELS OF ^DJI

Model	RMSE	MAE	MAPE	R2	EVS
IWOA-LSTM	344.4983	277.7575	0.7700%	0.9870	0.9888
WOA-LSTM	436.9144	357.2306	0.9766%	0.9791	0.9829
LSTM	483.1204	379.3579	1.0572%	0.9744	0.9749
RNN	632.4826	513.0577	1.3752%	0.9562	0.9732

TABLE VII. COMPARISON RESULTS OF PREDICTIONS BETWEEN VARIOUS MODELS OF ^FTSE

Model	RMSE	MAE	MAPE	R2	EVS
IWOA-LSTM	73.6150	54.8546	0.7112%	0.9502	0.9502
WOA-LSTM	73.9682	55.0725	0.7130%	0.9497	0.9500
LSTM	183.0727	157.7665	1.9980%	0.6917	0.8732
RNN	77.3710	60.2541	0.7754%	0.9449	0.9526

TABLE IX. COMPARISON RESULTS OF PREDICTIONS BETWEEN VARIOUS MODELS OF 000001.SS

Model	RMSE	MAE	MAPE	R2	EVS
IWOA-LSTM	38.1677	25.8819	0.8397%	0.9371	0.9372
WOA-LSTM	42.3903	28.5810	0.9294%	0.9224	0.9229
LSTM	56.3180	37.6384	1.2245%	0.8631	0.8631
RNN	43.9960	30.4456	0.9893%	0.9164	0.9175

TABLE VIII. COMPARISON RESULTS OF PREDICTIONS BETWEEN VARIOUS MODELS OF ^IXIC

Model	RMSE	MAE	MAPE	R2	EVS
IWOA-LSTM	208.4710	167.3047	1.1916%	0.9921	0.9923
WOA-LSTM	259.5208	211.9485	1.4697%	0.9878	0.9899
LSTM	324.4165	264.5356	1.8428%	0.9810	0.9836
RNN	327.8000	260.9647	1.7722%	0.9806	0.4920

As shown by the comparison results in Table VI to Table IX and Fig. 7 to Fig. 10, the IWOA-LSTM model demonstrates superior prediction accuracy and a better fit than other models in predicting four different stock indices (^DJI, ^FTSE, ^IXIC, and 000001.SS). The IWOA-LSTM model achieves lower errors and improved accuracy across multiple evaluation metrics (RMSE, MAE, MAPE, R², EVS). In particular, its RMSE and MAE values indicate better accuracy compared to WOA-LSTM.

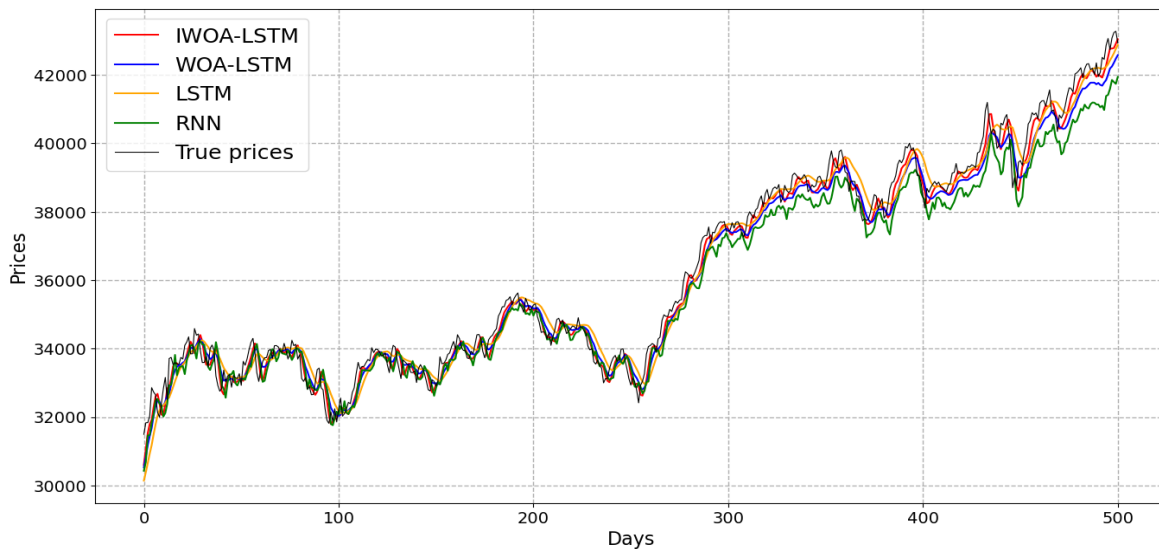


Fig. 7. Comparison of forecasts from four models on the ^DJI index.

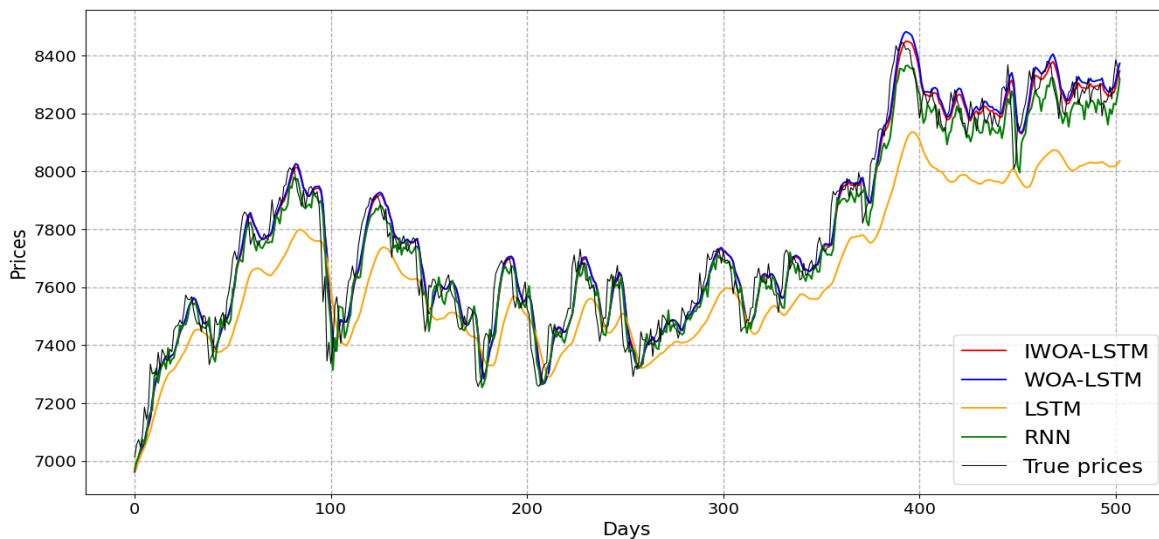


Fig. 8. Comparison of forecasts from four models on the ^FTSE index.

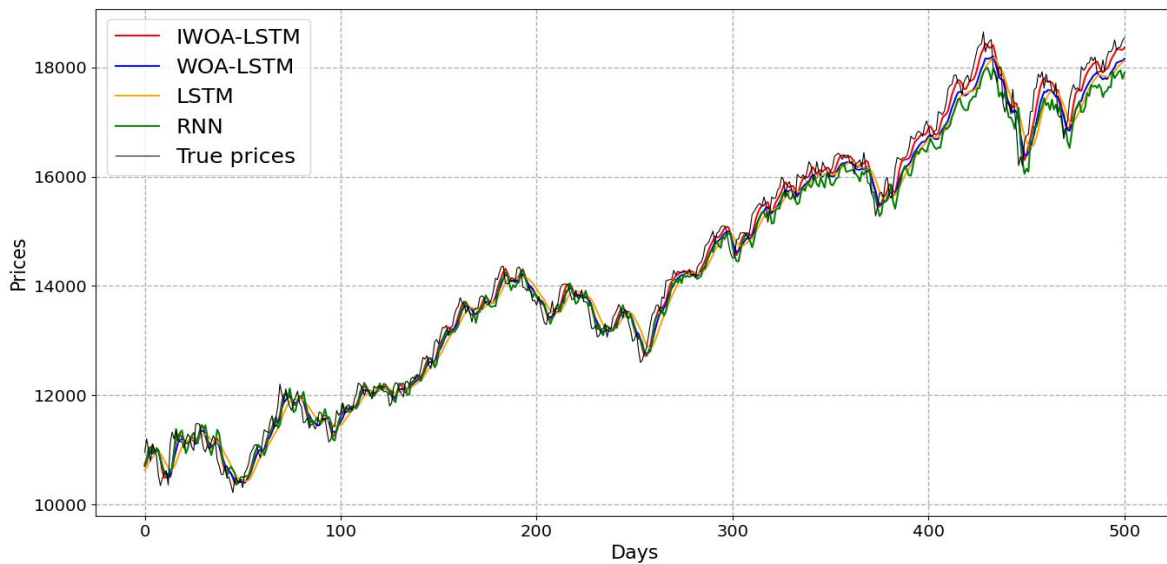


Fig. 9. Comparison of forecasts from four models on the ^IXIC index.

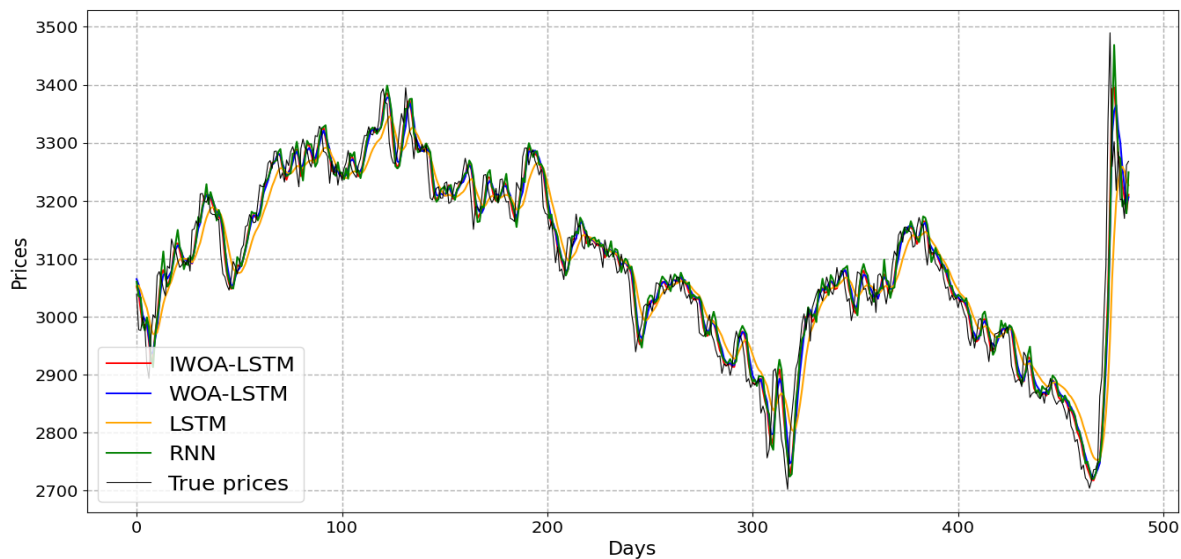


Fig. 10. Comparison of forecasts from four models on the 000001.SS index.

When compared with the LSTM and RNN models, the IWOA-LSTM model shows clear advantages, particularly in capturing the volatility of financial markets and the complexity of time series data. It demonstrates strong adaptability in the ^DJI, ^FTSE, and ^IXIC and also performs well in predicting the Shanghai Composite Index (000001.SS), with RMSE, MAE, and MAPE values lower than those of the other comparative models. Overall, the IWOA-LSTM model improves prediction accuracy and stability through optimized hyperparameter settings, demonstrating strong potential for application in financial time series forecasting, particularly in stock index prediction.

V. CONCLUSION

A novel model integrating the Improved Whale Optimization Algorithm (IWOA) with Long Short-Term Memory (LSTM) is proposed to enhance the accuracy of stock

market index forecasting. This model employs chaotic map initialization and a dynamic adjustment mechanism to optimize the LSTM network's parameters, thereby improving prediction which confirms performance. Chaotic assignment improves the global search capability of the algorithm and enables a thorough exploration of the solution space. Additionally, the dynamic adjustment factor increases WOA efficiency, optimizes LSTM hyperparameters, and improves prediction accuracy and stability.

The study focuses on five representative stock market indexes (^GSPC, ^DJI, ^FTSE, ^IXIC, and 000001.SS) and evaluates the model's performance adopting five essential evaluation metrics: RMSE, MAE, MAPE, R^2 , and EVS. These metrics comprehensively evaluate the model's prediction ability from the perspectives of error magnitude, absolute and relative error, goodness of fit, and variance explained. Experimental results show that the IWOA-LSTM model

outperforms WOA-LSTM, LSTM and RNN in all five metrics. This highlights its superior accuracy, robustness and stability.

Although this study presents an innovative approach to stock index prediction and offers valuable contributions, the IWOA-LSTM model has certain limitations. Its performance depends heavily on the range of hyperparameter values, which need to be set based on empirical experience to ensure optimal results. Additionally, although the IWOA-LSTM model reduces runtime compared to the traditional WOA algorithm, the computational cost remains relatively high. Furthermore, the validation is conducted on only five representative stock indices, limiting the sample size.

Future research will focus on expanding the dataset to include additional stock indices and broader financial market data to assess the model's generalizability and robustness. The model will also be applied to other time series tasks, such as energy forecasting and medical trend analysis, to evaluate its cross-industry adaptability and performance. Additionally, advanced optimization techniques, including the integration of metaheuristic algorithms and hybrid optimization methods, will be explored to enhance feature selection and hyperparameter tuning. Modal decomposition methods will be studied to decompose and reconstruct time series data, reducing noise and improving the model's ability to detect market fluctuations, thereby enhancing prediction accuracy and stability. Further research will also explore the correlations between the stock market and other financial markets to analyze the potential impact of external market fluctuations. Moreover, external features, such as macroeconomic indicators and social sentiment, will be incorporated to assess their influence on predictions and improve the model's adaptability in complex market environments.

ACKNOWLEDGMENT

The research is funded by Guangdong Province Key Discipline Research Capacity Enhancement Project (No. 2022ZDJS146); Dongguan Science and Technology of Social Development Program (No. 20221800902782).

REFERENCES

- [1] M. A. Rahim, M. Mushafiq, S. D. Khan, R. Ullah, S. Khan, and M. Ishaque, "Technical analysis-based unsupervised intraday trading DJIA index stocks: is it profitable in long term?," *Applied Intelligence*, vol. 55, no. 2, 2025, pp. 1–12.
- [2] W. Chen, C. K. Yeo, C. T. Lau, and B. S. Lee, "Leveraging social media news to predict stock index movement using RNN-boost," *Data & Knowledge Engineering*, vol. 118, 2018, pp. 14–24.
- [3] M. M. Al Haromainy, D. A. Prasetya, and A. P. Sari, "Improving performance of RNN-based models with genetic algorithm optimization for time series data," *TIERS Information Technology Journal*, vol. 4, no. 1, 2023, pp. 16–24.
- [4] X. Zhang, N. Gu, J. Chang, and H. Ye, "Predicting stock price movement using a DBN-RNN," *Applied Artificial Intelligence*, vol. 35, no. 12, 2021, pp. 876–892.
- [5] A. Q. Md, S. Kapoor, C. J. AV, A. K. Sivaraman, K. F. Tee, H. Sabireen, and N. Janakiraman, "Novel optimization approach for stock price forecasting using multi-layered sequential LSTM," *Applied Soft Computing*, vol. 134, 2023, p. 109830.
- [6] Z. Yang, Z. Zeng, K. Wang, S. S. Wong, W. Liang, M. Zanin, and J. He, "Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions," *Journal of Thoracic Disease*, vol. 12, no. 3, 2020, pp. 165–174.

- [7] J. Q. Wang, Y. Du, and J. Wang, "LSTM based long-term energy consumption prediction with periodicity," *Energy*, vol. 197, 2020, p. 117197.
- [8] Z. Li, H. Yu, J. Xu, J. Liu, and Y. Mo, "Stock market analysis and prediction using LSTM: A case study on technology stocks," *Innovations in Applied Engineering and Technology*, 2023, pp. 1–6.
- [9] P. Singh, M. Jha, M. Sharaf, M. A. El-Meligy, and T. R. Gadekallu, "Harnessing a hybrid CNN-LSTM model for portfolio performance: A case study on stock selection and optimization," *IEEE Access*, vol. 11, 2023, pp. 104000–104015.
- [10] Y. Yu, X. Si, C. Hu, and J. Zhang, "A review of recurrent neural networks: LSTM cells and network architectures," *Neural Computation*, vol. 31, no. 7, 2019, pp. 1235–1270.
- [11] A. G. Nikolaev and S. H. Jacobson, "Simulated annealing," in *Handbook of Metaheuristics*, 3rd ed., G. T. Rado and H. Suhl, Eds. New York: Academic, 2010, pp. 1–39.
- [12] Y. Ji, A. W. C. Liew, and L. Yang, "A novel improved particle swarm optimization with long-short term memory hybrid model for stock indices forecast," *IEEE Access*, vol. 9, 2021, pp. 23660–23671.
- [13] X. Chen, L. Cheng, C. Liu, Q. Liu, J. Liu, Y. Mao, and J. Murphy, "A WOA-based optimization approach for task scheduling in cloud computing systems," *IEEE Systems Journal*, vol. 14, no. 3, 2020, pp. 3117–3128.
- [14] Y. Zhang and S. Yang, "Prediction on the highest price of the stock based on PSO-LSTM neural network," in *2019 3rd International Conference on Electronic Information Technology and Computer Engineering (EITCE)*, Oct. 2019, pp. 1565–1569.
- [15] M. Clerc and J. Kennedy, "The particle swarm-explosion, stability, and convergence in a multidimensional complex space," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 1, 2002, pp. 58–73.
- [16] H. Xin and H. Yu, "WOA-LSTM CSI 500 forecast model based on Baidu Index," in *International Conference on Business Intelligence and Information Technology*, Singapore, Dec. 2023, Springer Nature Singapore, pp. 139–147.
- [17] M. W. Hasan, "Building an IoT temperature and humidity forecasting model based on long short-term memory (LSTM) with improved whale optimization algorithm," *Memories-Materials, Devices, Circuits and Systems*, vol. 6, 2023, p. 100086.
- [18] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, 1994, pp. 157–166.
- [19] A. Graves, "Supervised sequence labelling," in *Springer Berlin Heidelberg*, 2012, pp. 5–13.
- [20] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, 2016, pp. 51–67.
- [21] F. S. Gharehchopogh and H. Gholizadeh, "A comprehensive survey: Whale Optimization Algorithm and its applications," *Swarm and Evolutionary Computation*, vol. 48, 2019, pp. 1–24.
- [22] M. H. Nadimi-Shahraki, H. Zamani, Z. A. Varzaneh, and S. Mirjalili, "A systematic review of the whale optimization algorithm: theoretical foundation, improvements, and hybridizations," *Archives of Computational Methods in Engineering*, vol. 30, no. 7, 2023, pp. 4113–4159.
- [23] Z. Che, C. Peng, and C. Yue, "Optimizing LSTM with multi-strategy improved WOA for robust prediction of high-speed machine tests data," *Chaos, Solitons & Fractals*, vol. 178, 2024, p. 114394.
- [24] D. Prasad, A. Mukherjee, and V. Mukherjee, "Temperature dependent optimal power flow using chaotic whale optimization algorithm," *Expert Systems*, vol. 38, no. 4, 2021, p. e12685.
- [25] D. Yousri, D. Allam, and M. B. Eteiba, "Chaotic whale optimizer variants for parameters estimation of the chaotic behavior in Permanent Magnet Synchronous Motor," *Applied Soft Computing*, vol. 74, 2019, pp. 479–503.
- [26] E. Mezura-Montes and C. A. C. Coello, "Constraint-handling in nature-inspired numerical optimization: past, present and future," *Swarm and Evolutionary Computation*, vol. 1, no. 4, 2011, pp. 173–194.
- [27] C. Shao and J. Ning, "Construction and application of carbon emission prediction model for China's textile and garment industry based on

- improved WOA-LSTM,” *J. Beijing Inst. Fash. Technol. (Nat. Sci. Ed.)*, vol. 43, 2023, pp. 73–81.
- [28] X. Gua, X. Gong, X. Yang, et al., “基于 WOA-BiLSTM 模型的短期光伏出力预测 [Short-term photovoltaic output forecasting based on WOA-BiLSTM model],” *Electric Power and Energy*, vol. 44, no. 6, 2023, pp. 613-616+653.
- [29] W. Jia, Y. Zhang, Z. Wei, Z. Zheng, and P. Xie, “Daily reference evapotranspiration prediction for irrigation scheduling decisions based on the hybrid PSO-LSTM model,” *Plos One*, vol. 18, no. 4, 2023, p. e0281478.
- [30] X. Liang, “Stock Market Prediction with RNN-LSTM and GA-LSTM,” *SHS Web of Conferences*, vol. 196, 2024, p. 02006, EDP Sciences.

Multinode LoRa-MQTT of Design Architecture and Analyze Performance for Dual Protocol Network IoT

Rizky Rahmatullah¹, Hongmin Gao², Ryan Prasetya Utama³, Puput Dani Prasetyo Adi⁴,
Jannat Mubashir⁵, Rachmat Muwardi⁶, Widar Dwi Gustian⁷, Hanifah Dwiyanti⁸, Yuliza⁹
School of Integrated Circuits and Electronics, Beijing Institute of Technology, Beijing, China^{1,2,5}
National Research and Innovation Agency, Indonesia^{3,4,7,8}
School of Optics and Photonics, Beijing Institute of Technology, Beijing, China⁶
Department of Electrical Engineering, Universitas Mercu Buana, Jakarta, Indonesia^{6,9}

Abstract—LoRaWAN networks and large places do not support Wi-Fi for multiple points. An architecture that offers dual networks to alter their supporting networks is needed for IoT device installation. The novelty in this research is that designing an architecture for multimode LoRa-MQTT with a mechanism for testing LoRa data transmission with different delays and Wireshark for testing Wi-Fi network QoS on MQTT is necessary. This hour-long LoRa network experiment shows that the End-Node can only receive one data at a time. One data set will be received if several data sets are obtained due to conflict. The second experiment showed data barely reached 70%. The signal strength or RSSI, and the node that sent the data initially decide the data received from a given node, some seconds apart, towards tested QoS with excellent packet loss, 21 ms delay, 50,616 bytes/s throughput, and 0.1426 jitter. Avoid data conflicts and loss by utilizing fewer nodes or adding end nodes in this experiment. The network service is excellent. According to this study, LoRa and MQTT can work well together. This approach could solve Internet of Things communication concerns, especially in large places that are LoRaWAN-inaccessible and Wi-Fi networks are limited.

Keywords—LoRa; MQTT; multinode; QoS; LoRaWAN

I. INTRODUCTION

Improvements in telecommunications technology are currently developing rapidly, and new telecommunications systems are starting to emerge, such as the LoRa (Long Range) protocol [1]-[3] and also MQTT [4],[5], which are used in various fields. LoRa is currently believed to be a telecommunications technology that can be used for long-distance communications because LoRa uses CSS (Chirp Spread Spectrum) modulation technology [6], which makes it possible to send low-power and long-distance data. LoRa is often used in IoT (Internet of Things) devices [7]-[9] to send data from node to node. The advantage of LoRa is that it uses non-licensed frequencies such as the AS920-923 standard, which is the LoRa frequency standard in Asia ranging from 920MHz to 923MHz so that any user can freely use this frequency.

LoRa is a technology located at the physical layer network for limited node-to-node long-distance communication [10]. It cannot be used for open networks such as Wi-Fi or GSM [11], [12], which allows data to be sent to servers. An open network is important for the Internet of Things so that users can monitor or control IoT devices. So, LoRaWAN [13] is a solution because

LoRaWAN works at the data link layer [14]. LoRaWAN allows IoT devices to send data to servers and already has been tested with a multi-client model [15]. However, the LoRaWAN network remains limited in its capabilities due to incomplete implementation across all sites, For the future, LoRaWAN must be able to be combined with cross-communication technology at different frequencies, for example at 2.4 GHz, with a large bandwidth it is possible to transfer data well such as traffic light data [16] which requires video in the data transmission process, and also Face Recognition [17] which is based on images. Some techniques such as doing split images when transmitting data are also still a challenge using LoRa's limited bandwidth. LoRaWAN Technology must also continue to be improved in terms of server security, for now, LoRaWAN Server uses AES [18] as server security, In the future, LoRaWAN technology will continue to be developed along with the development of telecommunication systems such as Lacuna Space which uses LPWAN Satellite LEO technology in its development.

Moreover, if the Internet of Things device is configured for the LoRaWAN network, this must be considered since it may provide difficulty. The site of the Internet of Things device installation lacks functionality for the LoRaWAN network. Moreover, such a broad area only facilitates the Wi-Fi network for many places. This research proposes a solution to the location issue that is incompatible with the LoRaWAN network. Moreover, a limited number of locations have Wi-Fi access. On the other hand, MQTT is usually used to communicate media of the Internet of Things. This solution is realized by integrating LoRa and MQTT to facilitate data transmission between the user and the server, or vice versa. Two communications are needed, namely LoRa and Wi-Fi [19], to send data to the MQTT broker.

The main contribution of this work is to demonstrate and analyze the LoRa-MQTT multi-node communication technology. In this research, the design of architecture LoRa-MQTT on multi-node IoT devices to send and receive data so that users can monitor data in real-time [20] and control [21] the devices installed on the IoT device. Apart from that, LoRa and Wi-Fi networks that use the MQTT protocol [22] will be analyzed using Wireshark [23] to determine the quality of service (QoS) [24]-[26] on the system. As a result, the percentage of end-node data reception sent by multi-nodes will be calculated later. Other than that, measurements will be taken of the data received by the MQTT Broker to know the system's reliability in the end.

II. METHOD AND ARCHITECTURE DESIGN

This section presents the architectural design for multimode and end-node, an IoT device consisting of an ESP32 and LoRa microcontroller [27],[28]. Two programming settings are applied to each node, namely to send data from node to end node and receive data from end nodes. Likewise, the end-node has two settings, namely receiving from each node, then the end-node will publish the data to the MQTT-broker [29]-[31]. During the publishing process, the network will be tested using Wireshark to see the quality of service on the network. Quality of service consists of packet loss, the total number of packets lost from all the data left, as in Eq. (1). Then delay is the time required for data to cover the distance from origin to destination, as in Eq. (2). Next, throughput is the actual bandwidth measured in units. A specific time is used to transfer data of a certain size, such as Eq. (3) Jitter is a variation of delay caused by the queue length in data processing and reassembly of data packets at the end of transmission due to previous failures such as Eq.(4) Delay variations will refer to all samples in Wireshark. The second setting is to subscribe to data by the end node, and the end node will send the data to the node according to the address. As shown in the block diagram in Fig. 1. In carrying out the multi-communication technique of LoRa end-nodes, a method is needed so that there is no redundancy in transmission which causes packet errors [6] such as ADR, LBT, LR-FHSS, etc, This is an essential solution.

Furthermore, Eq. (1)-Eq. (4) are the basic formulas used in making the analysis in this research and finding the right results

from the LoRaWAN analysis by comparing the theoretical and practical results so that it becomes the right solution during the process of transmitting data analysis.

$$Packet\ loss = \frac{(Packet\ transmittite - Packet\ Received)}{Packet\ transmittite} \times 100(1)$$

$$Delay = \frac{Total\ Delay}{Total\ Packets\ Received} \quad (2)$$

$$Throughput = \frac{Packet\ Received}{Time\ Transmitted} \quad (3)$$

$$Jitter = \frac{Total\ Delay\ Variation}{Total\ packets\ received - 1} \quad (4)$$

This research uses six nodes as IoT devices that will send data to the end nodes. Each node, including end nodes, has an address and channel, as shown in Fig. 1. The address of each node is different; each LoRa can be set using numbers ranging from 0 to 65535 and the same channel; this channel represents the frequency used. In this research, channel 72 is used, which represents the 922MHz frequency. Each node has the same program contents: the destination address, destination channel, and data. Data nodes consist of two types, namely analog data and digital data. Analog data represents data from sensors, and digital data represents data from actuators, which only consists of 1 and 0 or 1 for active and 0 for inactive. Address settings, channels, and examples of programming content, along with examples of data with semicolon separators sent by each node, can be seen in Table I.

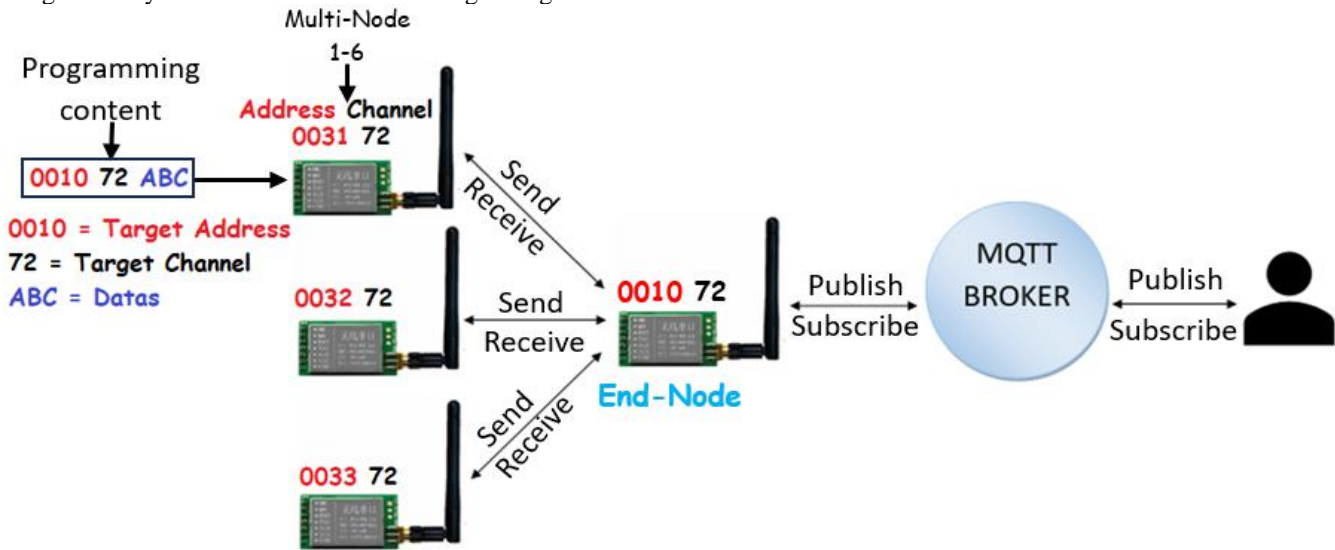


Fig. 1. The design architecture of multinode LoRa-MQTT.

TABLE I. CONFIGURATION OF LoRA NODES

Node Name	Address	Channel	Target Address	Target Channel	Programming Content	Information
Node 1	31	72	10	72	0, 10, 72, 10;20;30	Key = 0 (NOT USE) Target Address=10 Target Channel=72 String Data = 10;20;30
			10	72		
Node 2	32	72	10	72		
			10	72		
Node 3	33	72	10	72		
			10	72		
Node 4	34	72	10	72		
			10	72		
Node 5	35	72	10	72		
			10	72		
Node 6	36	72	10	72		
			10	72		
End-Node	10	72	31, 32, 33, 34, 35, 36	72	From 10;20;30 to be [data1=10;data2=20;data3=30]	convert string data to JSON

The data flow process from multi-node to end-node to the user is as follows:

- 1) Data is sent from each node to the end node in the form of string data type using the LoRa network with 10 addresses and 72 channels or 922MHz frequency.
- 2) Then, the data will be converted into JSON from the end node.
- 3) Next, it is published to the MQTT Broker using a Wi-Fi network with the topic smartfarm/Address_Node/sensors with the information in Table III.
- 4) Users will subscribe to data from MQTT Broker on the same topic as number 3.

Next, the user process sends command data to move the actuator on the node as follows:

- 1) The user will send data in JSON form to the MQTT broker with the topic smartfarm/Address_Node/actuator with the information in Table III.
- 2) Then, subscribed by the end-node on the same topic as number 1.
- 3) Then, the end node converts the JSON data to String.

The end node sends data to the user's destination node according to the Address_Node on the topic.

III. RESULT AND DISCUSSION

This section presents the results and analysis of several experiments, which are divided into two. The first is an analysis of LoRa communications, which explains the results of experiments sending data from multi-node to end-node with variations in delay, and the second is an analysis of the quality of service on Wi-Fi networks with the MQTT protocol using Wireshark.

A. LoRa Communication Analysis

Define abbreviations and acronyms the first time they are used. This section presents an analysis of the experiment of sending data from six nodes to the end node. Several experiments have been carried out, namely sending data simultaneously at one time and sending data with different time delay variations at each node, as in Table II and detail conflict in Table IV.

In the first experiment, each node sends data to the end node with the same delay, namely 1 second. Only node 1 is sent during the sending process, and only data from one node can be received. This is due to data reception conflicts with other nodes. The data received from a particular node depends on the node that sent the data first with a difference of a few seconds and the signal strength or RSSI (Receive Signal Strength Indicator), as shown in Fig. 2.

TABLE II. RESULT AND ANALYZES OF VARIATION DELAY FOR SENDING DATA

No Experiment	Delay settings	Information data received by the end-node					
		Node 1	Node 2	Node 3	Node 4	Node 5	Node 5
1	Node 1,2,3,4,5,6 = 1 second	Data Received	no data	no data	no data	no data	no data
2	Node 1 = 3 minutes Node 2 = 5 minutes Node 3 = 7 minutes Node 4 = 9 minutes Node 5 = 11 minutes Node 6 = 13 minutes	Data Received with note : Data reception conflict 13 times in 1 hour with other nodes	Data Received with note : Data reception conflict 6 times in 1 hour with other nodes	Data Received with note : Data reception conflict 3 times in 1 hour with other nodes	Data Received with note : Data reception conflict 7 times in 1 hour with other nodes	Data Received with note : Data reception conflict 2 times in 1 hour with other nodes	Data Received with note : Data reception conflict 1 times in 1 hour with other nodes
3	Node 1 = 13 minutes Node 2 = 14 minutes Node 3 = 15 minutes Node 4 = 16 minutes Node 5 = 17 minutes Node 6 = 18 minutes	Data Received with note : There is no conflict within 1 hour, but after more than 1 hour there will be a conflict in receiving data with other nodes					

TABLE III. MQTT TOPIC INFORMATION

No	Topic of MQTT	Information
1	smart farm/Address_Node/sensors	Topic for data from nodes (analog or digital data). If the data is from Address 30 then Address_Node=30
2	smart farm/Address_Node/Threshold	Topic for threshold data for sensor values to move actuators. If the data is from Address 30 then Address_Node=30
3	smart farm/Address_Node/actuator	Topic command data from the user. If the user's destination is address 30 then address_Node = 30

Each node was given different data transmission delay variations in the second experiment. Within 1 hour, there were quite frequent reception conflicts. The data received by the end node will depend on the split-second difference and RSSI. From the result of the experiment, using the Greatest Common Factor (GCF), can identify which nodes are going through conflict along with which minutes as shown in Table IV, and the percentage of receiving data is 70%.

In the third experiment, delay variations were given from node 1 to node 6, starting from 13 to 18 minutes. Within 1 hour of sending data to the end node, there was no conflict in receiving data; however, after more than 1 hour, there was a conflict in receiving data, and conflicts occurred more frequently over time because the time interval delay is much longer. This multi-node system cannot communicate data at the same time. To avoid conflict in the receiver, future studies can set each node to send data separately between milliseconds and seconds.

B. Analyzes Network of MQTT Protocol

Furthermore, when the end node receives data in the string type from each node, it will convert the string data into JSON and immediately send it to the MQTT Broker. This section will present network analysis with the MQTT protocol using Wireshark to capture network data with uplinks and downlinks, as in Table V. This experiment was carried out to see the network stress level by sending data every five seconds with a data load of 118 bytes. Then, the Wireshark capture results will be processed to produce Quality of Service as in Table VI, which consists of delay, throughput, packet loss, and jitter using Eq. (1) - Eq. (4).

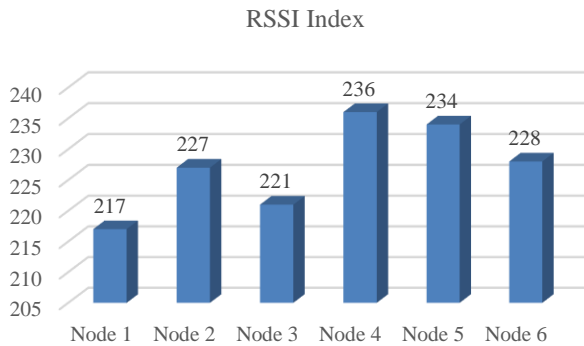


Fig. 2. RSSI Index from node 1 to node 6.

Moreover, from the perspective of the ITU G.1010 standard, Table VI provides an overview of the quality of service on the network that was utilized to access the MQTT Broker on the smart farm server. Because there is no packet loss, the delay value in this table falls into the category of very good. This indicates that there is no loss of data during the transmission process from the end node to the MQTT Broker. After that, the delay is considered to be very good because it is still less than 150 milliseconds; this is because the network conditions are good and the weather conditions are sunny.

The next step in the research process is to conduct experiments with a variety of weather conditions to enhance the architectural architecture of the LoRa-MQTT protocol. Therefore, the jitter is satisfactory because it is less than 75 milliseconds; it would be ideal if the jitter were zero milliseconds. In addition to this, it is essential to conduct research with two end nodes for every six nodes to get a better understanding of the design and reduce the number of data conflicts.

TABLE IV. DETAIL CONFLICT OF NODE WITHIN 1 HOUR OBSERVATION

Number of Send Data	Node1 (3 min)	Node 2 (5 min)	Node 3 (7 min)	Node 4 (9 min)	Node 5 (11 min)	Node 6 (13 min)	Conflict
1							
2							
3	x						
4							
5		x					
6	x						
7			x				
8							
9	x			x			+
10		x					
11					x		
12	x						
13						x	
14			x				
15	x	x					+
16							
17							
18	x			x			+
19							
20		x					
21	x		x				+
22					x		
23							
24	x						
25		x					
26						x	
27	x			x			+
28			x				
29							
30	x	x					+
31							
32							
33	x				x		+
34							
35		x	x				+
36							
37							
38							
39	x					x	+
40		x					
41							
42	x		x				+
43							
44					x		
45	x	x		x			+
46							
47							
48	x						
49			x				

50		x					
51	x						
52						x	
53							
54	x			x			+
55		x			x		+
56			x				
57	x						
58							
59							
60	x	x					+
Total Data Sent	20	12	8	6	5	4	55
Total Data Receive	39						
Percentage of Receive	7.090.909.091						

C. Analyzes Network of MQTT Protocol

When the end node receives data in the string type from each node, it will convert the string data into JSON and immediately send it to the MQTT Broker. This section will present network analysis with the MQTT protocol using Wireshark to capture network data with uplinks and downlinks, as in Table V. This experiment was carried out to see the network stress level by sending data every five seconds with a data load of 118 bytes. Then, the Wireshark capture results will be processed to produce Quality of Service as in Table VI, which consists of delay, throughput, packet loss, and jitter using Eq. (1)-(4).

TABLE V. UPLINK AND DOWNLINK FROM ENDNODE TO MQTT BROKER SMARTFARM SERVER

No.	Time	Source	delay
1	18:42:26,477690000	10.72.0.121	00:00:05,044
2	18:42:31,521880000	10.72.0.121	00:00:05,004
3	18:42:36,526048000	10.72.0.121	00:00:05,049
4	18:42:41,574734000	10.72.0.121	00:00:05,067
5	18:42:46,641786000	10.72.0.121	00:00:05,046
6	18:42:51,687533000	10.72.0.121	00:00:05,061
7	18:42:56,749428000	10.72.0.121	00:00:05,023
8	18:43:01,772036000	10.72.0.121	00:00:05,068
9	18:43:06,839599000	10.72.0.121	00:00:05,025
10	18:43:11,864858000	10.72.0.121	00:00:05,068
11	18:43:16,932637000	10.72.0.121	00:00:05,052
12	18:43:21,985399000	10.72.0.121	00:00:05,018
13	18:43:27,002951000	10.72.0.121	00:00:05,004
14	18:43:32,007443000	10.72.0.121	00:00:05,050
15	18:43:37,057111000	10.72.0.121	00:00:05,008

TABLE VI. VALUE AND CATEGORY FOR PACKET LOSS, DELAY, THROUGHPUT, AND JITTER

	Packet Loss	Delay	Throughput	Jitter
Value	0%	21ms	50.616 bytes/s	0.1426ms
Category	Very Good	Very Good	No Category	Good

From the perspective of the ITU G.1010 standard, Table 6 provides an overview of the quality of service on the network that was utilized to access the MQTT Broker on the smartfarm server. Because there is no packet loss, the delay value in this table falls into the category of very good. This indicates that there is no loss of data during the transmission process from the end node to the MQTT Broker. After that, the delay is considered to be very good because it is still less than 150 milliseconds; this is because the network conditions are good and the weather conditions are sunny.

The next step in the research process is to conduct experiments with a variety of weather conditions to enhance the architectural architecture of the LoRa-MQTT protocol. Therefore, the jitter is satisfactory because it is less than 75 milliseconds; it would be ideal if the jitter were zero milliseconds. In addition to this, it is essential to conduct research with two end nodes for every six nodes to get a better understanding of the design and reduce the number of data conflicts.

IV. CONCLUSION

The issue is that the IoT device's location lacks compatibility with the LoRaWAN network, and extensive areas require Wi-Fi network connectivity for many position points. The research defines the multi-node LoRa-MQTT architecture and conducts a performance comparison of the two protocols. An experiment was conducted in the LoRa network to transmit data from six nodes to the end nodes with differing delays. The results of this experiment, conducted over one hour of observation, indicate that the End-Node can receive just one data packet at a time. If multiple data points are received, a conflict will occur, resulting in the reception of only one data point. The second trial revealed that the data obtained scarcely attained 70%. The data obtained from a certain node is contingent upon the initial node that

transmitted the data, with a few seconds of interval, in addition to the signal strength or RSSI. The subsequent part of the research involves identifying a mechanism, modulation, adjustment of delay, or procedure that facilitates data transmission. Contemplate employing a queueing method. In the investigation of the MQTT protocol network, there is 0% packet loss classified as very excellent, a delay of 21ms also categorized as very good, a throughput of 50,616 bytes/s, and a jitter of 0.1426 classified as good. This study demonstrates that the integration of LoRa with the MQTT protocol can yield a highly successful solution. This concept can address Internet of Things obstacles, including communication problems in extensive platforms that lack LoRaWAN accessibility, with Wi-Fi networks available only at certain locations.

FUTURE RESEARCH

LoRaWAN technology that is developing at this time, not only setting the Multi-Communication and transmission analysis process using MQTT but needs development from the Terrestrial Communication side to Non-Terrestrial Communication before data is received on the LoRaWAN Server. This is used in the future to reduce Packet Error and the presence of obstacles and Interferences.

ACKNOWLEDGMENT

Thanks to Riset dan Inovasi untuk Indonesia Maju (RIIM) (G2) National Research and Innovation Agency (BRIN) for providing research funding support. Thanks also to the School of Integrated Circuits and Electronics, Beijing Institute of Technology, and colleagues at the telecommunications research center and electronics research center who have contributed to the completion of this research. The author hopes that this research can be further developed and utilized as well as possible by partners.

REFERENCES

- [1] Q. L. Hoang, W. -S. Jung, T. Yoon, D. Yoo and H. Oh, "A Real-Time LoRa Protocol for Industrial Monitoring and Control Systems," in IEEE Access, vol. 8, pp. 44727-44738, 2020, doi: 10.1109/ACCESS.2020.2977659.
- [2] Q. L. Hoang and H. Oh, "A Real-Time LoRa Protocol Using Logical Frame Partitioning for Periodic and Aperiodic Data Transmission," in IEEE Internet of Things Journal, vol. 9, no. 16, pp. 15401-15412, 15 Aug. 2022, doi: 10.1109/JIOT.2022.3162019.
- [3] H. P. Tran, W. -S. Jung, T. Yoon, D. -S. Yoo and H. Oh, "A Two-Hop Real-Time LoRa Protocol for Industrial Monitoring and Control Systems," in IEEE Access, vol. 8, pp. 126239-126252, 2020, doi: 10.1109/ACCESS.2020.3007985.
- [4] N. -Z. Wang and H. -Y. Chien, "Design and Implementation of MQTT-Based Over-the-Air Updating Against Curious Brokers," in IEEE Internet of Things Journal, vol. 11, no. 6, pp. 10768-10777, March 15, 2024, doi: 10.1109/JIOT.2023.3327447.
- [5] F. Buccafurri, V. de Angelis and S. Lazzaro, "MQTT-A: A Broker-Bridging P2P Architecture to Achieve Anonymity in MQTT," in IEEE Internet of Things Journal, vol. 10, no. 17, pp. 15443-15463, 1 Sept. 1, 2023, doi: 10.1109/JIOT.2023.3264019.
- [6] Adi, P. D. P., & Wahyu, Y. (2023). The error rate analyze and parameter measurement on LoRa communication for health monitoring. Microprocessors and Microsystems, 98, 104820. <https://doi.org/10.1016/J.MICPRO.2023.104820>.
- [7] G. Y. Odongo, R. Musabe, D. Hanyurwimfura, and A. D. Bakari, "An Efficient LoRa-Enabled Smart Fault Detection and Monitoring Platform for the Power Distribution System Using Self-Powered IoT Devices," in IEEE Access, vol. 10, pp. 73403-73420, 2022, doi: 10.1109/ACCESS.2022.3189002.
- [8] S. Lee, J. Lee, J. Hwang and J. K. Choi, "A Novel Deep Learning-Based IoT Device Transmission Interval Management Scheme for Enhanced Scalability in LoRa Networks," in IEEE Wireless Communications Letters, vol. 10, no. 11, pp. 2538-2542, Nov. 2021, doi: 10.1109/LWC.2021.3106649.
- [9] L. -H. Shen, C. -H. Wu, W. -C. Su and K. -T. Feng, "Analysis and Implementation for Traffic-Aware Channel Assignment and Contention Scheme in LoRa-Based IoT Networks," in IEEE Internet of Things Journal, vol. 8, no. 14, pp. 11368-11383, July 15, 2021, doi: 10.1109/JIOT.2021.3051347.
- [10] I. W. Mustika, W. J. Anggoro, E. Maulana, and F. Y. Zulkifli, "Development of Smart Energy Meter Based on LoRaWAN in Campus Area," in 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), Dec. 2020, pp. 209-214. doi: 10.1109/ISRITI51436.2020.9315511.
- [11] N. Datta, A. Malik, M. Agarwal and A. Jhunjhunwala, "Real Time Tracking and Alert System for Laptop through Implementation of GPS, GSM, Motion Sensor and Cloud Services for Antitheft Purposes," 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), Ghaziabad, India, 2019, pp. 1-6, doi: 10.1109/IoT-SIU.2019.8777477.
- [12] H. K. Patel, T. Mody, and A. Goyal, "Arduino Based Smart Energy Meter using GSM," 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), Ghaziabad, India, 2019, pp. 1-6, doi: 10.1109/IoT-SIU.2019.8777490
- [13] E. Sisinni et al., "LoRaWAN Range Extender for Industrial IoT," in IEEE Transactions on Industrial Informatics, vol. 16, no. 8, pp. 5607-5616, Aug. 2020, doi: 10.1109/TII.2019.2958620.
- [14] O. Seller, "LoRaWAN Link Layer," in Journal of ICT Standardization, vol. 9, no. 1, pp. 1-12, 2021, doi: 10.13052/jicts2245-800X.911.
- [15] A. F. Rachmani and F. Y. Zulkifli, "Trial and Evaluation of LoRa Performance for Smart System Multi-Client Model," in 2018 International Seminar on Intelligent Technology and Its Applications (ISITIA), Aug. 2018, pp. 39-44. doi: 10.1109/ISITIA.2018.8710898
- [16] R. Muwardi, H. Zhang, H. Gao, M. Yunita, Y. Wang and Yuliza, "Design of New Traffic System YOLO-LIO: Light-Traffic Intercept and Observation," 2024 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET), Bandung, Indonesia, 2024, pp. 20-25, doi: 10.1109/ICRAMET62801.2024.10809042.
- [17] R. Muwardi, H. Qin, H. Gao, H. U. Ghifarsyam, M. H. I. Hajar and M. Yunita, "Research and Design of Fast Special Human Face Recognition System," 2020 2nd International Conference on Broadband Communications, Wireless Sensors and Powering (BCWSP), Yogyakarta, Indonesia, 2020, pp. 68-73, doi: 10.1109/BCWSP50066.2020.9249452.
- [18] S. Budiyo, L. M. Silalahi, F. A. Silaban, R. Muwardi and H. Gao, "Delivery Of Data Digital High Frequency Radio Wave Using Advanced Encryption Standard Security Mechanism," 2021 International Seminar on Intelligent Technology and Its Applications (ISITIA), Surabaya, Indonesia, 2021, pp. 386-390, doi: 10.1109/ISITIA52817.2021.9502262.
- [19] I. Stoev, S. Zaharieva, A. Borodzheva and G. Staevska, "An Approach for Securing MQTT Protocol in ESP8266 Wi-Fi Module," 2020 XI National Conference with International Participation (ELECTRONICA), Sofia, Bulgaria, 2020, pp. 1-4, doi: 10.1109/ELECTRONICA50406.2020.9305164.
- [20] A. I. Siam et al., "Portable and Real-Time IoT-Based Healthcare Monitoring System for Daily Medical Applications," in IEEE Transactions on Computational Social Systems, vol. 10, no. 4, pp. 1629-1641, Aug. 2023, doi: 10.1109/TCSS.2022.3207562
- [21] L. -D. Liao et al., "Design and Validation of a Multifunctional Android-Based Smart Home Control and Monitoring System," in IEEE Access, vol. 7, pp. 163313-163322, 2019, doi: 10.1109/ACCESS.2019.2950684
- [22] B. Mishra and A. Kertesz, "The Use of MQTT in M2M and IoT Systems: A Survey," in IEEE Access, vol. 8, pp. 201071-201086, 2020, doi: 10.1109/ACCESS.2020.3035849
- [23] Rahmatullah, R., Dani Prasetyo Adi, P., Prasetya, S., Budi Santiko, A., Wahyu, Y., Surya Wicaksana, Bb., Yushady Bissa, S. C., Jana Yanti, R.,

- Adya Pramudita, A. (2023). Analyze Transmission Data from a Multi-Node Patient's Respiratory FMCW Radar to the Internet of Things. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 14(5), 2023. <http://dx.doi.org/10.14569/IJACSA.2023.0140521>
- [24] Rahmatullah, R., Kadarina, T, M., Dkk. Design and Implementation of IoT-Based Monitoring Battery and Solar Panel Temperature in Hydroponic System. *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)*, Indonesia, 2023, pp. 810-820, doi: 10.26555/jiteki.v9i3.26729.
- [25] E. Shahri, P. Pedreiras and L. Almeida, "A Scalable Real-Time SDN-Based MQTT Framework for Industrial Applications," in *IEEE Open Journal of the Industrial Electronics Society*, vol. 5, pp. 215-235, 2024, doi: 10.1109/OJIES.2024.3373232.
- [26] Y. Xiao, E. Pei, K. Wang, W. Zhou, and Y. Xiao, "Design and Research of M2M Message Transfer Mechanism of Looms for Information Transmission," in *IEEE Access*, vol. 10, pp. 76136-76152, 2022, doi: 10.1109/ACCESS.2022.3189367.
- [27] A. Pradeep, A. Latifov, A. Yodgorov, and N. Mahkamjonkhodzoda, "Hazard Detection using custom ESP32 Microcontroller and LoRa," 2023 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, United Arab Emirates, 2023, pp. 36-40, doi: 10.1109/ICCIKE58312.2023.10131784.
- [28] A. F. Rachmani and F. Y. Zulkifli, "Design of IoT Monitoring System Based on LoRa Technology for Starfruit Plantation," in *TENCON 2018 - 2018 IEEE Region 10 Conference*, Oct. 2018, pp. 1241-1245. doi: 10.1109/TENCON.2018.8650052.
- [29] K. Kosaka, Y. Noda, T. Yokotani and K. Ishibashi, "Implementation and Evaluation of the Control Mechanism Among Distributed MQTT Brokers," in *IEEE Access*, vol. 11, pp. 134211-134216, 2023, doi: 10.1109/ACCESS.2023.3335273.
- [30] A. Velinov, A. Mileva, S. Wendzel, and W. Mazurczyk, "Covert Channels in the MQTT-Based Internet of Things," in *IEEE Access*, vol. 7, pp. 161899-161915, 2019, doi: 10.1109/ACCESS.2019.2951425.
- [31] L. G. A. Rodriguez and D. M. Batista, "Resource-Intensive Fuzzing for MQTT Brokers: State of the Art, Performance Evaluation, and Open Issues," in *IEEE Networking Letters*, vol. 5, no. 2, pp. 100-104, June 2023, doi: 10.1109/LNET.2023.3263556.

Machine Learning-Based Fifth-Generation Network Traffic Prediction Using Federated Learning

Mohamed Abdelkarim Nimir Harir¹, Edwin Ataro², Clement Temaneh Nyah³

Electrical Engineering (Telecommunication option), Pan African University Institute of Basic Sciences Technology and Innovation (PAUSTI), Nairobi, Kenya¹

School of Electrical and Electronic Engineering, Technical University of Kenya, Nairobi, Kenya²

Electrical and Computer Engineering, University of Namibia, Windhoek, Namibia³

Abstract—The rapid development and advancement of 5G technologies and smart devices are associated with faster data transmission rates, reduced latency, more network capacity, and more dependability over 4G networks. However, the networks are also more complex due to the diverse range of applications and technologies, massive device connectivity, and dynamic network conditions. The dynamic and complex nature of the 5G networks requires advanced and accurate traffic prediction methods to optimize resource allocation, enhance the quality of service, and improve network performance. Hence, there is a growing demand for training methods to generate high-quality predictions capable of generalizing to new data across various parties. Traditional methods typically involve gathering data from multiple base stations, transmitting it to a central server, and performing machine learning operations on the collected data. This work suggests a hybrid model of Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), and federated learning applied to 5G network traffic prediction. The model is assessed on one-step predictions, comparing its performance with standalone LSTM and GRU models within a federated learning environment. In evaluating the predictive performance of the proposed federated learning architecture compared to centralized learning, the federated learning approach results in lower Root Mean Square error (RMSE) and Mean Absolute Errors (MAE) and a 2.25 percent better Coefficient of Determination (R squared).

Keywords—5G Mobile network; machine learning; federated learning; parallel hybrid LSTM+GRU; network traffic prediction; centralized learning; dynamic network condition

I. INTRODUCTION

As witnessed the evolution of communication networks into the 5G era, the demand for high-speed, low-latency connectivity is growing exponentially. The development of 5G networks also increases data rates and complexity due to various services and multiplexed device connections; this makes network resource management of a 5G network a complicated task because of the diverse nature of network traffic conditions. The growth in the number of users and devices is increasing traffic exponentially, causing congestion in the network from many angles [1].

Due to most devices now being connected, the conventional 4G networks cannot meet the current demand. The advantages that the 5G network can provide are becoming more visible as its scope continues to expand. Compared to their 4G counterparts, 5G networks provide faster data transmission, reduced delay, expanded coverage, and greater reliability.

Managing 5G networks as they evolve and become more complicated is now one of the greatest difficulties with developing them. One critical element of this management is predicting network traffic, which has advanced greatly through machine learning techniques [1], [2].

With new elements such as millimeter waves, massive MIMO (Multiple Input, Multiple Output), and network slicing, 5G networks are much more sophisticated than their predecessors. This complexity requires sophisticated traffic control methods. Moreover, the various services of 5G, such as enhanced mobile broadband (eMBB), massive machine-type communication (mMTC), and ultra-reliable low-latency communication (URLLC) [3], imply different traffic dynamics that require prediction and control.

The huge increase in the number of devices connected leads to an increase in the volume of mobile traffic and adds stress to the system to cope with the volume of data [4]. Ericsson expects 5G subscriptions projected at approximately 610 million by the end of 2023, meaning that about one-fifth of all mobile subscriptions worldwide would be 5G [5]. The research predicted demand growth will increase to approximately 5.3 billion 5G subscriptions by 2029 [6].

5G networks allow the implementation of edge computing, which brings computation and data storage close to the network edge. By processing data at the edge, latency-sensitive applications might achieve lower network utilization levels while enjoying greater speeds, security, and privacy [7].

Network traffic forecasting was based on statistical models. Techniques such as time series analysis, regression analysis, and Markov models have been employed to forecast network behavior by leveraging historical traffic data. Time series models such as the Autoregressive Integrated Moving Average (ARIMA) are effective tools in identifying seasonal trends and irregular patterns in data over periods [8]. They make it possible to determine traffic cycles that could be predicted at intervals of a day, week, or even a month. Regression models, e.g., linear, polynomial, and multiple regression analysis, can explain such relations as the time of day, human activity, and external factors such as weather. Besides this, these models are good at understanding relationships influencing traffic volume.

Markov models such as the Hidden Markov Models (HMM) and the Markov Chain Monte Carlo (MCMC), on the other hand, apply a probabilistic technique to estimate different stages of the network, thus market appeal and ability to offer better traffic

prediction because of considering the stochasticity of the network traffic [8], [9]. Some of these traditional models have also been used for forecasting traffic.

However, with the growth of the networks, the weaknesses of these models are becoming more of a concern. In most cases, these models obtrude the linearity, which, along with the inability to manage high dimensional data, high rate of change in patterns of networks, or manage an anomaly, which is not a common phenomenon in the network [9]. This highlights the requirement for more sophisticated predictive models. Today, machine learning (ML) methods are believed to help deal with the complexities of 5G traffic, especially concerning the predictive aspect. Unlike traditional models, ML models can handle large volumes of datasets, capture nonlinear relationships, and learn and update to changes occurring in real-time.

The advantage of machine learning models is that they can discover intersectional structures or relationships that can be captured through conventional statistical means such as models by training generally on large datasets [10]. For instance, deep learning models such as Recurrent Neural Networks (RNNs) and long short-term memory (LSTM) appreciate recognizing sequential information in a given network traffic data frame rate, improving precision forecast along the time scale. Similarly, incentive-based resource allocation has been demonstrated to integrate across existing resource conditions by learning sub-network policies with dynamic structures [11].

Federated learning (FL) is projected as one of the many machine learning paradigms appropriate for 5G networks. Federated learning is a novel approach for training models without needing a central server to host raw data on devices [12]. This is highly timely in the case of fifth-generation networks (5G), as data privacy and security are critical owing to the potential proliferation of personal gadgets and the IoT. Federated Learning helps model training with privacy, bandwidth, and data constraints by leveraging edge devices in a distributed manner.

FL advancements enable privacy and more efficient model training, as the processing of network traffic data can occur at the edge of the network [13], [14]. This paper studies how predictive machine learning models based on federated learning can be employed to predict traffic in 5G networks.

The structure of the paper is as follows: Section II reviews the related work, while Section III outlines the architectures of the prediction methods. Section IV delves into the prediction methodology. The experimental results are presented in Section V, and the paper concludes in Section VI.

II. RELATED WORK

Advancements in 5G networks have made efficiently managing their dynamic and complex nature challenging. One possible solution to the above is predicting the network traffic on such a network, which is steadily receiving support in the form of machine learning advances.

The author in study [15], through network Internet traffic analysis and forecasting of input traffic flow parameters to the model, developed a 5G network traffic prediction model that

utilizes recurrent neural networks in their paper. They have employed Gated Recurrent Units (GRU) and (LSTM) to obtain a balance between optimality and viability. Such networks have acquired short-term traffic predictions since feature engineering was introduced to the model to reduce generalization errors and manage missing and corrupted data. Still, there is a need for more research on machine learning application techniques for network management and control in traditional distributed architectures.

In study [1], this paper presents a lightweight hybrid attention deep learning model for traffic prediction in 5G networks. The model integrates depthwise separable convolution with channel and spatial attention techniques to lower prediction costs. With its capacity to conserve computing resources, the model exhibits promise for use in integrated sensing, communication, and computation applications. The temporal and spatial properties of 5G network traffic data are revealed through data analysis, and the suggested model effectively addresses accuracy and complexity concerns using feature extraction and prediction capabilities.

To improve its prediction capabilities for 5G cellular network traffic flow, the authors in study [4] propose a deep learning model based on a Bidirectional Long Short-Term Memory (BiLSTM) architecture with hyperparameter optimization. The stated model demonstrates better prediction accuracy and shorter running time. Thus, it is helpful for real-time applications even though the authors did not discuss the practical limitations of deploying the model. The focus is on possible future research related to resource allocation schemes and IoT cloud architectures. Generally, the findings of the suggested Deep Learning Mobile Traffic Flow Prediction (DLMTFP) technique are encouraging for developing mobile traffic prediction in 5G networks.

In study [16], this paper proposes a Deep-Broad Learning System (DBLS) for traffic flow prediction in 5G cellular wireless networks. It explains that DBLS is suitable for 5G networks because it integrates deep representative and broad learning to provide accurate prediction while keeping the running time low. They showed that DBLS is more accurate and efficient than conventional deep neural networks. It is observed that enhancing the reasonable amounts of enhancement nodes adaptively can enhance the efficiency of the DBLS model and hence lead to high penetration prediction.

According to study [17], the study proposes to predict the traffic of the 5G network and its challenges, owing to the diversity and heterogeneous nature of the 5G traffic. To address these problems, a Smoothed Long Short-Term Memory (SLSTM) model is proposed to enhance prediction accuracy. Adjustments are made to the number of layers and hidden units based on the prediction accuracy, and seasonal time is based on the time series modeling techniques used to smooth the output sequences. This article recommends further research on other factors influencing 5G traffic to make it more applicable in practice.

In study [18], the study engages numerous cross-domain big data resources to construct a spatiotemporal cross-domain neural network model (STC-N) that enables deep learning in wireless cellular network regional traffic prediction. The method consists

of the integration of feature fusion, multi-domain data integration, timestamp-based modeling, and spatiotemporal correlations. The paper also discusses a cross-domain transfer learning approach for improved prediction performance in traffic generation.

It focuses on how cross-domain datasets interact within the prediction model and how it affects the accuracy of the prediction. Nevertheless, the analysis of different kinds and volumes of cross-domain datasets, their synthesis, and association effects on wireless cellular traffic prediction accuracy deserves further attention. Although the reporting in this paper concerns the effect of many cross-domain datasets on prediction accuracy, there is scope for investigating the best combination and weighting of this dataset.

Paper in study [19] describes a novel method of estimating the traffic flow in cellular networks utilizing counters that monitor the performance of LTE radio frequency signals. It investigates a range of machine learning models that can forecast traffic in the network depending on time. It demonstrates that while ensembles such as Gradient Boosting produce the most accurate predictions and spend longer training time, linear models operate faster but depend on preprocessing. The analysis identifies the importance of having a large volume of good quality data necessary to train machine learning models and speaks to the challenges of deployment and solutions whereby autoML may be utilized during retraining, regularization, and feature engineering.

In study [20], this paper considers the issues of mobile network forecasting as applied in a distributed manner, specifically with forecasting traffic for base stations and 5G networks overall. It evaluates different aspects of the centralized and federated learning model, pointing out the strengths of federated learning for better generalization metrics, economies of computational resources, and less carbon dioxide emission. It also highlights the role of model aggregation algorithms and data preprocessing methods in improving the predictive power of the models. The models, which include LSTM and GRU, are quite effective in federated learning scenarios.

The research presented in the paper [21] investigates the efficiency of energy usage in augmented deep-learning model architecture. It discusses federated traffic prediction mechanisms for cellular networks for optimal energy usage. It shows how the difference in the region affects the performance of a wide variety of models, such as Transformer and Length short-term memory-based models. The results demonstrate that while complicated models are more demanding in energy, the expectation is also a high increase in accuracy. This study seeks to raise the understanding of Distributed AI Technologies' environmental impacts and their pose threats to communication systems. It advocates the merits of incorporating sustainability factors into model selection.

In study [22], the paper discusses the problems of implementing FL in vehicular IoT systems, such as variable mobility, limits to communication capability, and risks of non-IID data in combination with the management of resources. Working issues in FL for autonomous driving, intelligent transport systems, and resource-sharing developments are elaborated. The authors define the areas for further explorations,

increasing FL advanced paradigms: scaling up and security on the background of the complex vehicular IoT scenarios.

III. PREDICTION METHOD ARCHITECTURES

A. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a relatively sophisticated variant of the Recurrent Neural Network (RNN) model commonly employed in research today. One of the key aspects LSTM networks address is the long-range dependencies in sequential data, which results in higher performance in many practical applications [23]. Some use cases where LSTM networks have been highly successful include language translation, voice detection, and forecasting. This explains the popularity of LSTM networks in multiple applications and their efficiency in deep learning frameworks directed toward time series data. They encode the RNN memory with three gates alongside cell states, allowing the network to keep and erase information when necessary.

In the standard arrangement, an LSTM block consists of four extra layers and a hidden state in an RNN. Variables include Cell state (C_t), input gate (i_t), output gate (o_t), and forget gate (f_t). Each layer performs a specific operation on the others depending on how the information is created from the training data [24]. Fig. 1 shows the structure of LSTM.

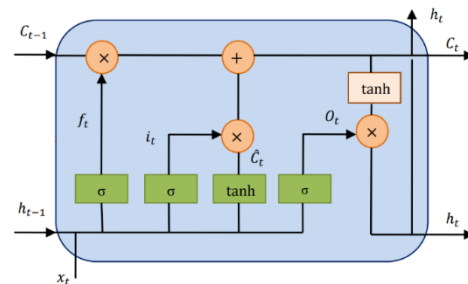


Fig. 1. Architecture of Long Short-Term Memory (LSTM) [23].

The memory of LSTM network networks is represented by the cell state, which is essential to LSTMs. The cell state process resembles a production line or conveyor belt. Except for a few linear interactions like addition and multiplication, the parameter information flows directly across the chain. These interactions determine the status of the information. The information will continue to flow without modifications if no interactions exist. Through the gates, which permit optional information to pass through, the LSTM block modifies or adds information to the cell state [24].

The forget gate eliminates data no longer needed in the cell state. The gate receives two inputs, x_t (the input at that specific moment) and h_{t-1} (the output of the previous cell), which are multiplied by weight matrices before bias is added. After being run through an activation function, the output is binary. When the output for a particular cell state is 0, the information is lost, and when the output is 1, it is saved for later use [25]. The nodal output equations of the LSTM are expressed as follows.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (2)$$

$$C_t^{\sim} = \tan h(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * C_t^{\sim} \quad (4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (5)$$

$$h_t = o_t * \tan h(c_t) \quad (6)$$

These equations describe how an LSTM unit works, distinguishing it from simple RNN, with several gates controlling information flow. The forget gate f_t in (1) determines what portion of the previous cell state C_{t-1} should be preserved, and the input gate i_t in Eq. (2) identifies how much new information from the current input x_t and previously hidden state h_{t-1} is added to the cell state. The candidate cell state C_t^{\sim} in Eq. (3) is determined with a tanh function. The new cell state C_t combines the old cell state and the new information added, as shown in Eq. (4). The output gate o_t in Eq. (5) controls how much of the updated cell state C_t is passed on to the hidden state h_t , influencing the output at this time step. The hidden state h_t in Eq. (6) is finally computed, where the output gate o_t is applied to the tanh of the new cell state, passing it on to the next time step, thus helping the LSTM keep long-term dependencies in sequential data.

Various parameters guide the internal mechanism of the network. W_f , W_i , W_c , and W_o are the weight matrices multiplied by the forget gate, input gate, candidate cell state, and output gate, respectively, and they are used for both the previous hidden state, h_{t-1} , and current input x_t . Similarly, b_f , b_i , b_c , and b_o are the biases related to these gates and states, added to the product sum of the inputs to bias the output. The sigmoid function σ is employed in the forget gate f_t , input gate i_t , and output gate o_t to compress values between 0 and 1, indicating how much influence they should have (how much to forget, retain, and output, respectively).

B. Gated Recurrent Unit (GRU)

GRU employs a gating mechanism to regulate the information passing through the network. The gates in LSTM determine which information to keep and which to discard at every step, enabling the network to learn long-range dependencies better. The GRU has two main components: the update and the reset gates.

The update gate decides how much new information to write to the memory now, and the reset gate decides how much old information to forget. The basic idea of GRU is that the network hidden state will be updated only by selecting time steps using gating mechanisms. The gates control what information joins and leaves the network. The GRU has two gating mechanisms: reset gate and update gate. The update gate specifies the proportion of the new input to add to the hidden state, and the reset gate specifies the extent to which the previous hidden state should be erased. The GRU output is computed based on the updated hidden state [23]. The architecture is shown in Fig. 2.

The update gate calculation is the first step in a GRU. It uses the current input and the previous hidden state to decide how much to update the previous hidden state; the sigmoid is used here [24]. Here are the GRU nodal output equations.

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (7)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (8)$$

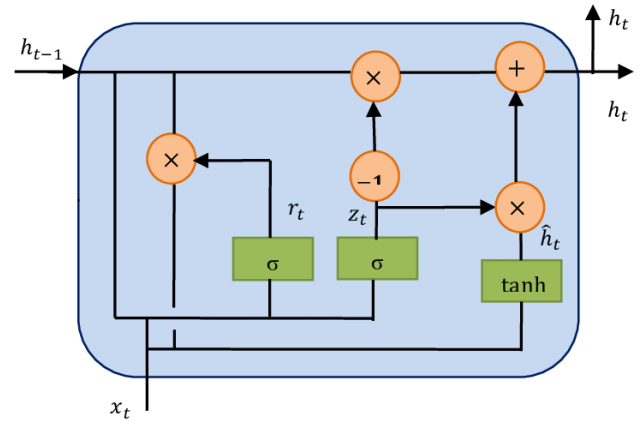


Fig. 2. Architecture of Gated Recurrent Unit (GRU) [23].

$$h_t^{\sim} = \tan h(W_h \cdot [r * h_{t-1}, x_t] + b_h) \quad (9)$$

$$h_t = z_t * h_{t-1} + (1 - z_t) * h_t^{\sim} \quad (10)$$

The GRU equations have several parameters that dictate how certain elements within the input data behave over the time steps. Where update gate z_t defined in Eq. (7), uses weights W_z , biases b_z , previously hidden state h_{t-1} , and current input x_t to determine how much of the past hidden state will pass to the next step. In the same way, Eq. (8) also applies a reset gate r_t , which uses weights W_r , biases b_r , and a combined h_{t-1} and x_t to decide how much to "forget" of the past for computing the hidden state of the candidate [26].

The candidate hidden state h_t^{\sim} is calculated from weights W_h , biases b_h , reset gate r_t applied to h_{t-1} and current input x_t , processed with the tanh function as shown in Eq. (9): Finally, the new hidden state h_t in Eq. (10) is expressed as a weighted sum of the candidate hidden state h_t^{\sim} (scaled by $1 - z_t$) and the previous hidden state h_{t-1} (scaled by z_t). The weights W_z , W_r , and W_h and the biases b_z , b_r , and b_h are learned during training and determine how inputs are decoded past and current information at each time step to adjust and combine the input information [27].

IV. PROPOSED METHOD DESCRIPTION

A. LSTM+GRU Parallel Network

In the proposed parallel hybrid model, the same input is applied to both LSTM and GRU layers. This enables the model to capture two different temporal representations simultaneously. This is especially beneficial because it combines the strengths of both architectures; thus, while the LSTM capability fortifies long-term dependencies, GRU computational efficiency and faster convergence make it an essential strength for more robust feature extraction in time series prediction tasks.

The input data processed through the LSTM and GRU branches come out as outputs from these branches. The outputs are then concatenated to form a combined feature representation [27]. After passing through dense layers, the final prediction is based on this combined representation. Fig. 3 shows the structure of the parallel hybrid model LSTM+GRU.

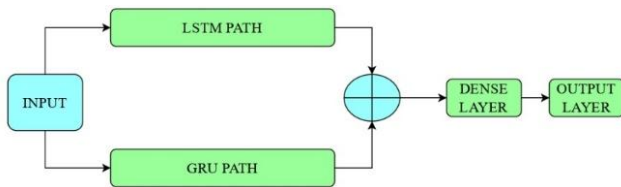


Fig. 3. Architecture of parallel hybrid model of LSTM+GRU.

As shown in Fig. 3, the model architecture consists of an input layer that receives the 5G traffic input data. The input is fed into both the LSTM and GRU paths simultaneously. Here, the LSTM (Long Short-Term Memory) network can capture long-term dependencies in time series data, which is important for identifying trends and patterns in 5G over long periods. However, the GRU (Gated Recurrent Unit) path takes the same input with fewer, faster-to-train steps in the architecture, allowing for efficient short-term dependency capture [27], [28]. Though both networks are good at processing sequential data, both contain different complementary strengths, such as LSTM being a long-term memory network and GRU being a memory-efficient network with fewer parameters.

As illustrated in Fig. 3, after processing the input via LSTM and GRU paths, the outputs are concatenated (as shown by the circle in Fig. 3). This merging step is performed to combine the information learned by the LSTM and GRU networks. This concatenated output is fed through a dense layer, and this layer helps further process the combined feature representation extracted from the LSTM and GRU branches [28]. Finally, the dense layer is connected to the output layer, which provides the model prediction.

B. Federated Learning

Federated learning is an advanced machine learning approach that enables decentralized model training without sharing raw data between multiple devices or nodes. It is suitable for privacy-preserving scenarios, such as 5G network traffic prediction [20].

In a centralized learning setting, data brought in from different sources, such as base stations or user devices, would be aggregated in one place during model training, which can raise concerns about the privacy and security of data. With federation learning, each device or node learns a model based on its data at each device or node, and only the changes in the model (such as the weights or the gradients) are sent to the central server [21]. Afterward, this central server employs these updates to enhance the performance of the global model. Furthermore, in the context of 5G network traffic prediction, federated learning allows individual base stations or edge devices in the network to collaborate on training a predictive model without exchanging their raw traffic data. This keeps the users and network-sensitive data secure while providing realistic traffic pattern predictions [12].

In 5G networks, federated learning presents a valuable approach due to the large geographical distribution of data from numerous devices. Through the distributed learning of models locally trained on diverse data, federated learning can also improve the prediction about network congestion, traffic

demand, and resource allocation for a particular network in the future while maintaining data privacy and low communication overhead in the network [21], [22]. Fig. 4 shows a round of the federated learning process.

The federated learning process involves multiple clients or base stations (BS) and a central server, as shown in Fig. 4. In step 1, the central server sends the global model to all clients. Step 2: Clients then update their local models by locally training the model with their private data. In step 3, clients return updated model parameters to the server (aggregator) without sharing raw data. These local model updates are then sent to a central server, which uses an aggregation function (that is, namely Federated Averaging) to aggregate these local model updates to produce an updated global model in Step 4. After updating the model, it repeatedly redistributes the new model to the clients for more training iterations. In this decentralized manner, clients update the global model in a privacy-preserving way by sending model updates rather than datasets.

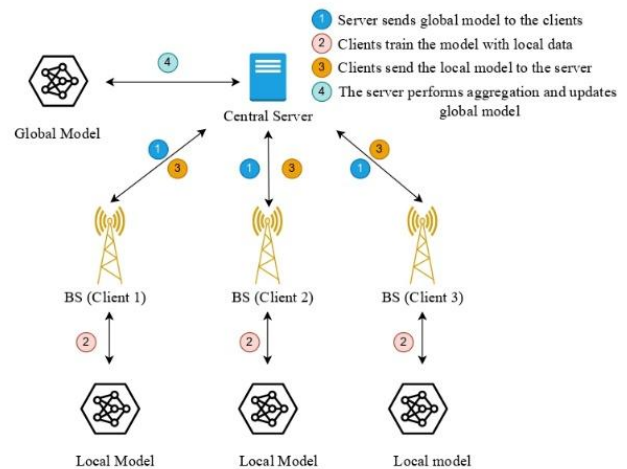


Fig. 4. Federated learning process [20].

Federated Averaging (FedAvg) is widely used due to its simplicity and effectiveness in handling non-iid (non-identically distributed) data across clients. This is common in real-world scenarios where different devices may have access to diverse datasets [12]. FedAvg also minimizes the communication overhead by reducing the frequency of interactions between the clients and the central server, making it well-suited for distributed environments.

C. Implementation of the LSTM+GRU Hybrid Model

The flowchart in Fig. 5 illustrates the steps in implementing and evaluating a hybrid LSTM+GRU model for predicting 5G network traffic, including data preparation, model training, and model testing.

1) *Data preprocessing*: This stage addresses various data quality issues, such as outliers, missing values, and data splits. Missing values were handled based on the percentage of missing data in each feature. Features with more than 50% missing values were removed, while those with less than 50% were imputed using the mean of the column.

Outliers were managed using the Interquartile Range (IQR) capping method, which can limit the impact of extreme values and improve model robustness.

The data was split into training and testing sets with three different ratios (80:20, 85:15, and 90:10). In this study, the 90:10 ratio was used as it provided the best results. The 90-10 splits mean we used 90% of the data to train the model and 10% to test it. This ensures that most of the data is used for model training and that a different portion is set aside to evaluate it.

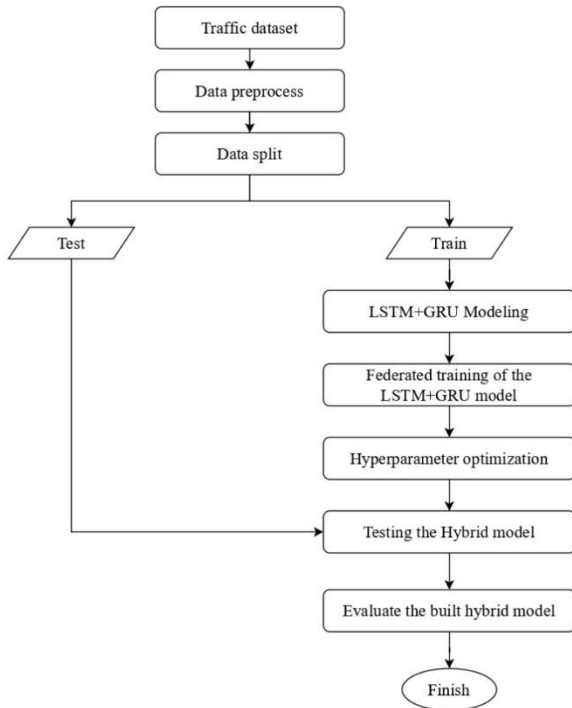


Fig. 5. Flowchart of modeling LSTM+GRU model for 5G network traffic prediction.

2) *LSTM+GRU modeling*: The hybrid LSTM-GRU model was implemented in Python and supported by TensorFlow libraries (i.e., Keras and TensorFlow Federated) using a Google Colab platform. The model had an LSTM layer of 128 units, then a GRU layer with the same number of units, and ReLU as an activation function. These layers were used in parallel (i.e., passed to an add() function). After that, the added output passes through a Dense layer with 64 units and ReLU activation. Finally, a Dense layer with 1 unit was added for the output. The Adam optimizer with a learning rate of 0.001 was used to optimize the model, and L1 and L2 regularizers (set to 0.05) were applied for overfitting prevention. The model was trained for 90 epochs with a batch size of 64.

3) *Federated training LSTM+GRU model*: The server starts the computation in federated training, and clients (base stations) join as participants. A subset of these clients are selected to receive the current global model from the server and use their local data to train [20], [21]. Once local training is done, the clients send the updated models and historical data (loss values and evaluation metrics) to the server. The server

then aggregates the locally trained models, updates the global model, and repeats the process for several federated rounds, as shown in Fig. 4. Hybrid models combining LSTM and GRU have been proposed before [27], [28]. To the best of our knowledge, this work is the first to implement a parallel LSTM+GRU network and training using federated learning for time series prediction. The study created a flexible framework to make it more realistic for network traffic prediction scenarios. After the model is trained, it is tested on the held-out testing data to see how well it predicts 5G network traffic. This step checks how well the LSTM+GRU model can predict the traffic.

4) *Hyperparameter optimization*: Hyperparameter tuning is an important part of neural network development and is usually done through trial. Table I shows the model-specific hyperparameters.

TABLE I. MODEL HYPERPARAMETERS

LSTM, GRU	LSTM+GRU
Activation: ReLU	Activation: ReLU for both branches
Output layer: linear activation	Output layer: linear activation
No. of units: 128	No. of units: 128 for both branches
Dense layer: 64 units	Dense layer: 64 units
Optimizer: Adam	Optimizer: Adam
Regularizer L1, L2: 0.06	Regularizer L1, L2: 0.05
Learning rate: 0.001	Learning rate: 0.001
Drop out: 0.4	Dropout: 0.2
Local Epochs: 3	Local Epochs: 3
Batch size: 64	Batch size: 64
Federated rounds: 10	Federated rounds:10

A validation run was done for each model to finalize the hyperparameters that gave the best performance and fit before training the final model. The training data was split 90-10 for validation during the validation process. Keras search is used for hyperparameter optimization to get the parameters shown in Table I.

5) *Evaluation metrics*: To evaluate and analyze the network model prediction results, the evaluation metrics used are the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Coefficient of Determination (R²). The corresponding mathematical formulas are presented in Eq. (11,12,13). RMSE in Eq. (11) is particularly effective at measuring the model dispersion, where a lower RMSE indicates a higher concentration level and greater accuracy.

MAE in Eq. (12), measures the absolute differences between the predicted and actual results by taking the absolute values and then calculating the mean. A lower MAE signifies a smaller prediction error. R-squared in Eq. (13) is widely used as an optimal measure for assessing linear regression models, as it translates the prediction accuracy into a value between 0 and 1, offering an intuitive representation of the model accuracy [29]. When the model fit is ideal, the R-squared value approaches 1.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2} \quad (11)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |f_i - y_i| \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - f_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (13)$$

V. EXPERIMENTAL RESULTS

A. Dataset Description

The paper uses a 5G trace dataset from an Irish mobile telecommunication operator, as outlined in study [30]. It considers file downloading in a dynamic environment and uses the Download traffic bandwidth data produced from file downloads in a dynamic environment as the target variable. The data samples are aggregated for different days of the experimental period.

- 10,974 Samples collected from 2019/12/14/10:16:30 to 2019/12/17/08:16:23
- 4,106 Samples collected from 2020/01/16/07:26:43 to 2020/01/12:16:29
- 12,511 Samples collected from 2020/02/13/13:03:24 to 2020/02/27/20:50:06

Preprocessing steps were executed to clean and prepare the raw data for analysis. The preprocessing steps included feature normalization, missing value treatment, outlier treatment, and data samples collected on different days aggregated into one dataset. The integration finally resulted in 27,591 samples in the final dataset. Ten features were selected: GPS coordinates (longitude and latitude), timestamp, uplink bitrate, download bitrate and its download state, velocity, and several cellular signal indicators RSRQ (Reference Signal Received Quality), RSRP (Reference Signal Received Power), SNR (Signal-to-Noise Ratio), and CQI (Channel Quality Indicator). Cellular signal indicators are critical as they glimpse the network physical layer. For 5G systems, these features are pertinent since they correlate to how signal quality affects bandwidth and throughput. Velocity and geolocation information permits an exploration of how network performance may vary with mobility versus location environments; such considerations are crucial to many applications in 5G, where users typically move around a lot.

The dataset was gathered from the initial deployment phase of 5G, and it includes key performance indicators (KPIs) such as throughput, channel conditions, and context-related metrics. These metrics remain fundamental to understanding network performance, regardless of technological advancements. As 5G builds on similar foundational principles, the data provides insights that still apply today.

B. Model Comparison

This study evaluates the performance of the proposed hybrid LSTM+GRU model against standalone LSTM and GRU models in the context of 5G network traffic prediction within the federated learning framework. Both LSTM and GRU units were specifically designed to capture temporal dependencies in sequential data; however, LSTM particularly excels in modeling long-term behaviors, while GRU gives a computationally efficient alternative for short-term dependencies with a simpler structure. The hybrid model is the parallel combination of these

structures, thus permitting richer feature extraction by exploiting both strengths. Performance scores of the prediction models for 5G network traffic are shown in Table II.

TABLE II. MODELS PERFORMANCE IN 5G NETWORK TRAFFIC PREDICTION

Models/Measures	RMSE	MAE	R2
LSTM	0.2360	0.3696	0.830
GRU	0.2349	0.3656	0.833
LSTM+GRU	0.2291	0.3556	0.845

Table II highlights the superiority of the hybrid LSTM+GRU model in predicting 5G network traffic. The hybrid model achieved the lowest RMSE of 0.2291 and MAE of 0.3556, demonstrating its ability to minimize large and average prediction errors effectively. Furthermore, its R^2 value of 0.845, the highest among the models, indicates that it explains 84.5% of the variance in the data, making it the most accurate and generalized model for capturing both long-term and short-term traffic patterns. Although the stand-alone GRU model performed better than the LSTM model, the hybrid model always gave better results.

The hybrid model predictive performance is further corroborated by visualizations in Fig 6, 7, and 8, where its predictions closely align with the actual data, showing minimal deviations. This superior accuracy can be attributed to the combined architecture of LSTM and GRU. Despite the hybrid model superior performance, it is computationally more expensive due to its integrated architecture, which increases the number of parameters and requires more memory and processing power. The training time is also longer, as the model must optimize both LSTM and GRU layers.

However, in scenarios where computational resources are available, the hybrid model offers a worthwhile trade-off, as its enhanced accuracy and generalization make it ideal for applications like network optimization or capacity planning. The GRU model is a simpler yet effective alternative for environments with resource constraints or needing faster prediction. The standalone LSTM model, however, appears less suited for 5G traffic prediction due to its lower overall performance and difficulty adapting to the data's highly dynamic nature.

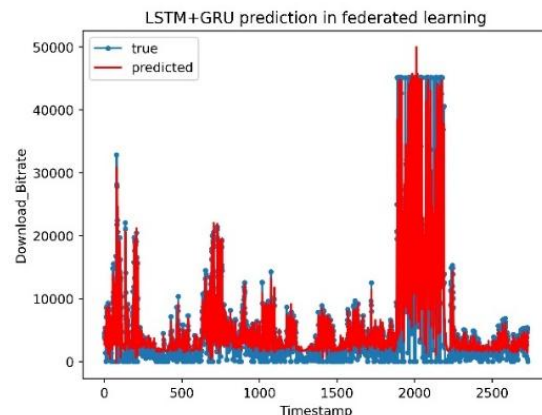


Fig. 6. Prediction of federated LSTM+GRU on test data.

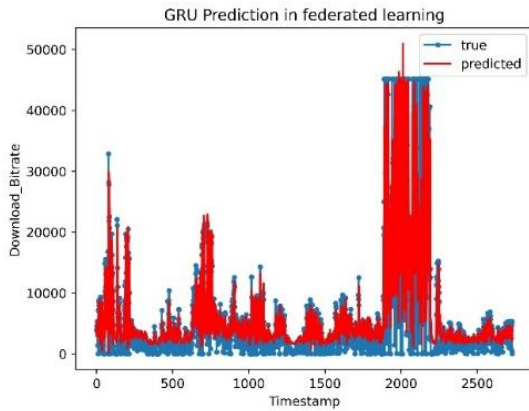


Fig. 7. Prediction of federated GRU on test data.

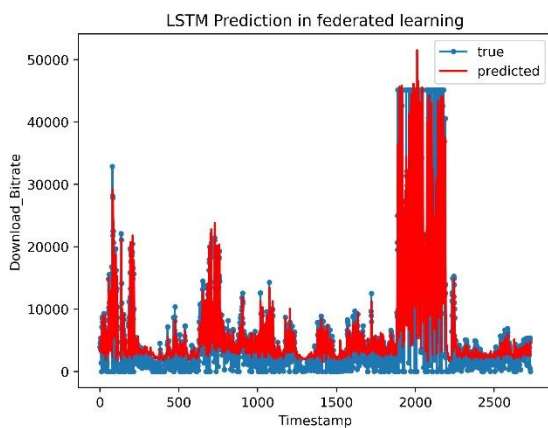


Fig. 8. Prediction of federated LSTM on test data.

C. Learning Setting Comparison

The time conducted an exhaustive set of experiments on the 5G network traffic dataset to analyze the performance of deep learning models with a specific focus on the effectiveness of the federated learning framework. The study compared the LSTM+GRU hybrid model performance in centralized versus federated learning settings. A centralized learning environment involves training the model on the complete dataset on a single server, while federated learning trains local models at various participants (clients) that aggregate after each round. The model architecture consistency is maintained across both learning environments to achieve comparison equity. The LSTM+GRU hybrid model was applied in both scenarios. In the centralized learning setup, the model was trained for 90 epochs, letting the model traverse the entire dataset 90 times. Ten federated rounds and three local epochs on each client were executed for the federated learning setup. Since federated learning involves multiple clients, the total practical epochs across all clients is 90, obtained as (rounds \times clients \times local epochs). The two results obtained under different learning frameworks were compared upon completing the experiments, as shown in Table III.

TABLE III. PERFORMANCE OF FEDERATED LEARNING AND CENTRALIZED IN 5G NETWORK TRAFFIC PREDICTION

Models/Measures	RMSE	MAE	R2
Centralized	0.2438	0.3687	0.826
Federated	0.2291	0.3556	0.845

Table III highlights a comparative analysis between centralized and federated learning in predicting 5G network traffic, emphasizing the advantages of federated learning. Federated learning achieved a reduced RMSE of 0.2291 and MAE of 0.3556, outperforming centralized learning, which yielded an RMSE of 0.2438 and MAE of 0.3687. It represents a 2.25% improvement in accuracy, underscoring the benefits of federated learning decentralized architecture. Federated learning ability to aggregate knowledge from diverse client models trained on local data allows it to capture a wider range of traffic patterns. This diversity introduces variations in local models, enhancing the global model ability to learn robust and generalized representations of network traffic behavior. In contrast, centralized learning lacks this diversity, relying on a single dataset, which limits its ability to generalize across varying traffic conditions.

One key benefit of federated learning is its scalability and privacy-preserving nature. Training models locally and aggregating updates at the server level avoids transferring raw data, making it an ideal solution for scenarios requiring strict data confidentiality, such as 5G networks. However, federated learning introduces complexity in synchronizing and aggregating models across multiple clients, which can increase computational complexity. Despite this, the distributed nature of federated learning ensures that the system remains scalable and capable of handling the demands of large-scale 5G networks while offering improved predictive accuracy.

Fig. 9 and Fig. 10 provide insights into the model predictions under centralized and federated learning setups. Both setups show the LSTM+GRU hybrid model performing well across a range of bitrate values. However, the performance in regions with lower bitrates reveals a notable challenge. The predictive accuracy drops near zero bitrates, indicating that the model struggles to detect meaningful patterns in this data range. This drop in performance is likely due to sparse or noisy data in these regions, where signal characteristics are less distinct. Such underfitting in low-bitrate areas highlights a common limitation in machine learning models when dealing with sparse or low-intensity data.

Addressing this challenge would involve strategies such as augmenting the training dataset to include more low-bitrate cases, ensuring the model encounters these scenarios during training. Another approach could involve using specialized techniques that enhance the model sensitivity to sparse data regions, such as weighted loss functions or regularization techniques tailored for imbalanced datasets. These enhancements would help mitigate underfitting and improve the model robustness, enabling more accurate predictions across the

entire bitrate spectrum. By doing so, federated learning could further solidify its position as a scalable and effective solution for 5G network traffic prediction, particularly when paired with architectures like the LSTM+GRU hybrid model that captures complex patterns.

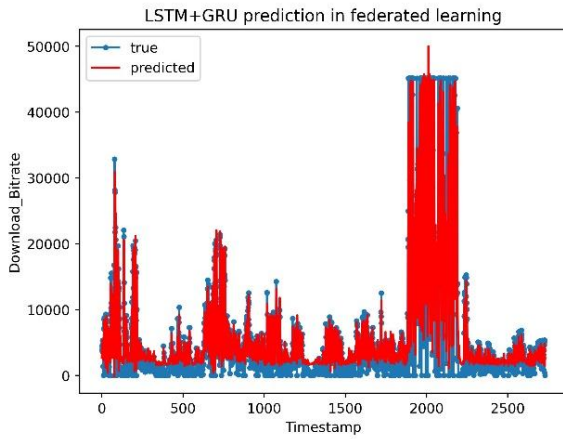


Fig. 9. LSTM+GRU model prediction in federated learning.

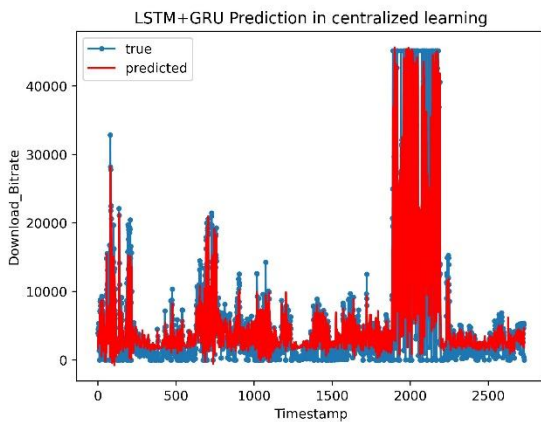


Fig. 10. LSTM+GRU model prediction in centralized learning.

D. Data Splitting Comparison

The preprocessed data was divided into three different ratios (80:20, 85:15, and 90:10). The three different ratios were compared in federated learning and centralized learning, and the ratio 90:10; 90% of the data for training and the remaining 10% of the data for the test yielded the best results compared to the other two ratios, as shown in Table IV, this experiment demonstrated that training on a larger proportion of data allows the model to capture more patterns and nuances in the data, ultimately leading to a better understanding of the underlying structure and relationships. The model can generalize well and perform with lower prediction errors.

TABLE IV. PERFORMANCE OF THE MODEL IN DIFFERENT SPLITS OF THE DATASET

Learning Setting	Federated Learning			Centralized Learning		
	10%	15%	20%	10%	15%	20%
Test Size/Measure						
RMSE	0.2291	0.3430	0.3553	0.2438	0.3444	0.3471
MAE	0.3556	0.4309	0.4448	0.3687	0.4363	0.4429
R2	0.845	0.818	0.805	0.826	0.817	0.811

VI. CONCLUSION

Building high-quality traffic prediction models with effective generalization is an inherently complex task, given the diverse data patterns that characterize 5G network traffic. This paper studies the challenge of predicting 5G network traffic using a hybrid LSTM+GRU model along with a federated learning approach. The hybrid model outperformed the standalone LSTM and GRU models, thus proving its capability to capture both long- and short-term dependencies within the data. At the same time, the federated learning approach adds another dimension to privacy by letting the system learn from varied data on different clients without compromising data privacy. In addition, it produced lower prediction errors with better generalization than centralized learning; thus, it would be an efficient and scalable solution under resource allocation optimization towards network performance enhancement and quality-of-service improvement in complex 5G environments while preserving data confidentiality.

ACKNOWLEDGMENT

The authors extend their gratitude to the African Union Commission for their financial contribution and also to the members of the Pan African University Institute for Basic Sciences, Technology, and Innovation (PAUSTI), Nairobi, Kenya.

REFERENCES

- [1] Su, J., Cai, H., Sheng, Z., Liu, A., & Baz, A. (2024). Traffic prediction for 5G: A deep learning approach based on lightweight hybrid attention networks. *Digital Signal Processing*, 146, 104359. <https://doi.org/10.1016/j.dsp.2023.104359>.
- [2] Saha, Sajal. "Toward building an intelligent and secure network: An internet traffic forecasting perspective" (2023). Electronic Thesis and Dissertation Repository. 9556. <https://ir.lib.uwo.ca/etd/9556>.
- [3] (Alekseeva, Stepanov et al. 2021) Chih-Lin and J. Huang, "RAN revolution with NGFI (xHaul) for 5G," 2017 Optical Fiber Communications Conference and Exhibition (OFC), Los Angeles, CA, USA, 2017, pp. 1-4.
- [4] M. Alias, N. Saxena and A. Roy, "Efficient cell outage detection in 5G HetNets using Hidden Markov Model," in *IEEE Communications Letters*, vol. 20, no. 3, pp. 562-565, March 2016.
- [5] Yue, B., Fu, J., & Liang, J. (2018). Residual recurrent neural networks for learning sequential Representations. *Information*, 9(3), 56. <https://doi.org/10.3390/info9030056>.
- [6] E. Selvamanju and V. B. Shalini, "Deep learning based mobile traffic flow prediction model in 5G cellular networks," 2022 3rd International Conference on Smart Electronics and Communication (ICOSEC).

- [7] Mahajan, Smita & Ramachandran, Harikrishnan & Kotecha, Ketan. (2022). Prediction of network traffic in wireless mesh networks using hybrid deep learning model. *IEEE Access*. PP. 1-1. 10.1109/ACCESS.2022.3140646.
- [8] Shafiq, M. O., Zhu, Z., Yu, X., & Liu, A. (2018). A survey on network traffic prediction techniques in communication networks. *Journal of Network and Computer Applications*, 100, 113-125.
- [9] Deng, S., Zhao, H., Fang, W., Yin, J., Dustdar, S., & Zomaya, A. Y. (2020). Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence. *IEEE Internet of Things Journal*, 7(8), 7457-7469.
- [10] Wang, S., Zhang, X., Zhang, Y., Wang, L., Yang, J., & Wang, W. (2019). A survey on mobile edge networks: Convergence of Computing, Caching, and Communications. *IEEE Access*, 7, 171676-171719.
- [11] Qiao, G., Yu, G., & Ding, Z. (2020). Deep Reinforcement Learning for Cooperative Content Caching in 5G Networks. *IEEE Transactions on Wireless Communications*, 19(6), 3977-3991.
- [12] Zhu, H., Gao, L., Li, W., & Yu, X. (2020). Federated Learning for 5G Vehicular Internet of Things: Opportunities and Challenges. *IEEE Wireless Communications*, 27(2), 12-18.
- [13] Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., & Ivanov, V. (2019). Towards federated learning at scale: System design. *Proceedings of the 2nd SysML Conference*, 1-15.
- [14] Wang, J., Kang, J., Lin, X., & Wang, S. (2019). When edge meets learning: Adaptive control for Resource-constrained distributed machine learning. *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, 63-71.
- [15] A. G. Reddy, S. Sinha, P. A. Reddy and C. Vimala, "Network traffic prediction for 5G network," 2023 International Conference on Recent Advances in Electrical, Electronics, Ubiquitous Communication, and Computational Intelligence (RAEEUCCI), Chennai, India, 2023, pp. 1-5, doi: 10.1109/RAEEUCCI57140.2023.10134497.
- [16] M. Chen, X. Wei, Y. Gao, L. Huang, M. Chen and B. Kang, "Deep-broad learning system for traffic flow prediction toward 5G cellular wireless network," 2020 International Wireless Communications and Mobile Computing (IWCMC), Limassol, Cyprus, 2020, pp. 940-945, doi: 10.1109/IWCMC48107.2020.9148092.
- [17] Gao Z. 5G Traffic prediction based on deep learning. *Comput Intell Neurosci*. 2022 Jun 24;2022:3174530. doi: 10.1155/2022/3174530. PMID: 35785055; PMCID: PMC9249458.
- [18] Q. Zeng, Q. Sun, G. Chen, H. Duan, C. Li and G. Song, "Traffic prediction of wireless cellular networks based on deep transfer learning and cross-domain data," in *IEEE Access*, vol. 8, pp. 172387-172397, 2020, doi: 10.1109/ACCESS.2020.3025210.
- [19] D. Alekseeva, N. Stepanov, A. Veprev, A. Sharapova, E. S. Lohan, and A. Ometov, "Comparison of machine learning techniques applied to traffic prediction of real wireless network," in *IEEE Access*, vol. 9, pp. 159495-159514, 2021, doi: 10.1109/ACCESS.2021.3129850.
- [20] Perifanis, V., Pavlidis, N., Koutsiamanis, R., & Efraimidis, P. S. (2022). Federated learning for 5G base station traffic forecasting. *ArXiv*. <https://doi.org/10.1016/j.comnet.2023.109950>.
- [21] V. Perifanis et al., "Towards energy-aware federated traffic prediction for cellular networks," 2023 Eighth International Conference on Fog and Mobile Edge Computing (FMEC), Tartu, Estonia, 2023, pp. 93-100, doi: 10.1109/FMEC59375.2023.10306017.
- [22] Z. Du, C. Wu, T. Yoshinaga, K. -L. A. Yau, Y. Ji, and J. Li, "Federated learning for vehicular Internet of things: Recent advances and open issues," in *IEEE Open Journal of the Computer Society*, vol. 1, pp. 45-61, 2020, doi: 10.1109/OJCS.2020.2992630.
- [23] Ali, M. A., Zhuang, H., Ibrahim, A.Rehman, O., Huang, M, and Wu, A. A machine learning approach for the classification of kidney cancer subtypes using miRNA. *Genome Data. Appl. Sci.* 2018, 8, 2422; doi:10.3390/app8122422.
- [24] Dridi, Hinda & Ouni, Kais. (2020). Towards robust combined deep architecture for speech recognition: Experiments on TIMIT. *International Journal of Advanced Computer Science and Applications*. 11. 10.14569/IJACSA.2020.0110469.
- [25] Okut, Hayrettin'' Deep Learning: Long-Short Term Memory''(2021).
- [26] Yue, B., Fu, J., & Liang, J. (2018). Residual recurrent neural networks for learning sequential representations. *Information*, 9(3), 56. <https://doi.org/10.3390/info9030056>.
- [27] E. Haque, S. Tabassum and E. Hossain, "A Comparative analysis of deep neural networks for hourly temperature forecasting," in *IEEE Access*, vol. 9, pp. 160646-160660, 2021, doi: 10.1109/ACCESS.2021.3131533.
- [28] Toledo, M.L., & Rezende, M.N. (2020). Comparison of LSTM, GRU and hybrid architectures for usage of deep learning on recommendation systems. *Proceedings of the 4th International Conference on Advances in Artificial Intelligence*.
- [29] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice* (2nd ed.). OTexts.
- [30] Raca, D., Leahy, D., Sreenan, C.J., & Quinlan, J.J. (2020). Beyond Throughput: The Next Generation a 5G Dataset with Channel and Context Metrics.

CN-GAIN: Classification and Normalization-Denormalization-Based Generative Adversarial Imputation Network for Missing SMES Data Imputation

Antonius Wahyu Sudrajat¹, Ermatita^{2*}, Samsuryadi³

Doctoral Program in Engineering Science, Universitas Sriwijaya, Palembang Indonesia¹

Faculty of Computer Science, Universitas Sriwijaya, Palembang Indonesia^{2,3}

Faculty of Computer Science and Engineering, Universitas Multi Data Palembang, Palembang, Indonesia¹

Abstract—Quality data is crucial for supporting the management and development of SMES carried out by the government. However, the inability of SMES actors to provide complete data often results in incomplete dataset. Missing values present a significant challenge to producing quality data. To address this, missing data imputation methods are essential for improving the accuracy of data analysis. The Generative Adversarial Imputation Network (GAIN) is a machine learning method used for imputing missing data, where data preprocessing plays an important role. This study proposes a new model for missing data imputation called the Classification and Normalization-Denormalization-based Generative Adversarial Imputation Network (CN-GAIN). The study simulates different patterns of missing values, specifically MAR (Missing at Random), MCAR (Missing Completely at Random), and MNAR (Missing Not at Random). For comparison, each missing value pattern is processed using both the CN-GAIN and the base GAIN methods. The results demonstrate that the CN-GAIN model outperforms GAIN in predicting missing values. The CN-GAIN model achieves an accuracy of 0.0801% for the MCAR category and shows a lower error rate (RMSE) of 48.78% for the MNAR category. The mean error (MSE) for the MAR category is 99.60%, while the deviation (MAE) for the MNAR category is 70%.

Keywords—Missing values; GAIN method; normalization-denormalization; imputation; UMKM data

I. INTRODUCTION

Indonesia's SMES (Micro, Small, and Medium Enterprises) are essential in increasing economic growth and regional income. This drives the Indonesian government to continue to develop SMEs through several schemes, including providing business capital, increasing business capacity through training, and so on. As a basis for developing SMES, the government requires SMES characteristic data as a basis for decision-making. Business Intelligence is a technology to support government work in managing SMES data. Data integration is an essential foundation of business intelligence.

Extract Transform Load (ETL) is an essential process in data integration where data processing is carried out. In the data integration process, many problems will affect data quality. One of the challenges in this process is handling missing values. Missing values are problems that arise in the ETL process, more

precisely in the data extraction step. [1]. The quality of the underlying data largely determines the quality of the extracted knowledge. Therefore, data quality is a significant concern in data analysis, and data quality is a prerequisite for obtaining quality knowledge. Missing value problems occur due to missing values from an attribute caused by errors when collecting data, system errors ([2], [3]), errors in data entry, refusal or inability of respondents to provide accurate answers [4] and merging of unrelated data [5]. Missing value is a fundamental problem in data science [6].

In some applications, missing values cannot be tolerated and must be replaced with concrete values [7]. Related studies have shown that missing value imputation is beneficial and is a better option than data deletion [8]. Missing data imputation means replacing or correcting the missing data with reasonable values to achieve completeness [9]. Missing data imputation is essential because decision-making errors will occur when an incomplete data set is supported [10]. Some important impacts of handling missing data include the accuracy of statistical analysis, better interpretation, reduction of bias, and improvement of data quality ([3], [11]).

The missing value imputation approach can be broadly categorized into traditional methods and Machine Learning (ML) based algorithm methods. Traditional methods include mean [12], median, linear regression [13], and mode. Some ML-based methods include Algorithms Clustering[14], K-Nearest Neighbor (KNN) [15], Support Vector Machine (SVM) [16], Decision Trees (DT) [17], [18], Random Forest (RF)[19] dan Generative Adversarial Networks (GAN) ([20], [21], [22], [23], [24]). The ability to optimize and extract relationships between data points is an advantage of machine learning-based methods [7]. GAN is an ML method that has attracted researchers' attention in recent years. Missing values are a significant problem in data mining, big data analysis, and ML-based decision-making flows, as the final mining or analysis results can be adversely affected when incomplete data is not imputed correctly [25]. Improvement efforts have been made in several studies that underlie the GAN method, including the research presented in [26] proposes improvements in a new method, namely Generative Adversarial Imputation Nets (GAIN) [27]. In this method, the generator accurately imputes

missing data, and the discriminator aims to distinguish between observed and imputed components. Further improvements to GAIN are carried out by research [7], where the idea is to use deconvolution on the generator and discriminator (DEGAIN). This method makes improvements by adding deconvolution to eliminate correlations between data. Improvements to the imputation method are made based on the characteristics of the data structure ([28], [29]) and the characteristics of the data values. At the same time, research that focuses on the characteristics of data values is still rarely done. The characteristics of the data values are an important initial step to perform accurate imputation. High differences in data values will result in inaccurate results in data processing.

In this study, we optimized the GAIN method [27], a GAN-based algorithm, by developing an enhanced version referred to as CN-GAIN. The CN-GAIN method improves upon GAIN by incorporating data preprocessing tasks as an initial step before imputation, taking into account the characteristics of the existing data. These preprocessing steps include data classification using the k-means method and normalization/denormalization using a robust scaler. The purpose of data classification is to categorize the data based on its inherent characteristics [30]. Meanwhile, normalization and denormalization ensure that no data values disproportionately dominate the dataset. We evaluated the performance of our proposed method using a dataset of SMES from a district in South Sumatra Province. The evaluation included measuring accuracy and several error metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE). We compared the performance of CN-GAIN with the standard GAIN algorithm.

This study proposes a new method for handling missing values, with the basic method being GAIN. This paper is structured as follows. Section II discusses related works in handling missing data, especially those based on the GAIN method and the improvement efforts made. Section III explains the efforts made by researchers in handling missing values through a series of stages and the use of methods so that they can produce more accurate imputation values. Section IV carries out each planned stage, which is very important in research to determine the proposed method's results. Section V discusses the results of each stage and the results obtained. Section VI presents the conclusions of the research and future research plans.

II. LITERATURE REVIEW

Missing value is a widespread problem encountered in many data collection cases. The missing value is a value that is not stored for a variable in the desired observation [30]. Missing data is grouped into three categories, namely: (1) Missing Completely at Random (MCAR), where the data is lost entirely at random (no dependence on any variable). (2) Missing at Random (MAR), where the missing data depends on the observed variables. (3) Missing Not at Random (MNAR), where the missing data depends on the observed variables and unobserved variables [31]. Data imputation is a common way to deal with missing values where missing values are replaced by applying various methods [32]. Missing value imputation is a valuable solution in cleaning datasets, and generative machine

learning methods can produce data that is completely indistinguishable from reality. Research in missing data imputation has mainly focused on adapting existing methods to suit specific datasets and operational environments, improving model adaptability and accuracy.

Generative Adversarial Net (GAN) is an artificial intelligence algorithm designed to solve generative modeling problems. GAN algorithms can be used to fill in gaps in missing data [33]. Many attempts have been made to improve GAN-based missing data imputation ([34], [35], [36][37]). One of the GAN-based imputation methods is Generative Adversarial Imputation Nets (GAIN). In GAIN, the generator component (G) takes a real data vector, imputes missing values conditioned on the actually observed data, and gives a complete vector. Then, the discriminator component (D) gets the complete vector and tries to determine which elements are actually observed and which are synthesized [31].

Several researchers have made improvements to the GAIN method. For example, [38] focused on repairing missing data in single-cell datasets using the basic GAIN method. The proposed method is Single-Cell Generative Adversarial Imputation Nets (scGAIN). Furthermore, the research conducted [39] proposed the Generative Adversarial Multiple Imputation Network (GAMIN) method for duplicate data imputation with data loss rate more than 80%. This method is applied to image data. Optimization of the GAIN method is also carried out by [40] by performing a pre-training procedure to learn the potential information contained in the data and classifying the data using synthetic pseudo-labels which are then named Pseudo-label Conditional Generative Adversarial Imputation Networks (PC-GAIN).

In research conducted by [41] proposed the Deconvolutional Generative Adversarial Imputation Network (DEGAIN) method, which makes improvements by adding deconvolution to eliminate correlations between data. The research [42] proposed a new missing data imputation model based on data clustering with the basic GAIN method as input data. The data set used is electricity consumption with the MCAR missing value type. This imputation method is then named Clustering and Classification-based Generative Adversarial Imputation Network (CC-GAIN). CC-GAIN aims to enhance imputation accuracy by considering both time-series and pattern features in building electricity consumption data.

TABLE I. PREVIOUS RESEARCH

Ref.	Year	Method	Dataset	Type Data
[38]	2019	scGAIN	Tow dataset: Simulated and rela-word dataset	Numeric
[39]	2020	GAMIN	MNIST and CelebA	Image
[40]	2021	PC-GAIN	UCI repository and MNIST dataset	Numerical, categorical and image
[41]	2023	DEGAIN	Letter and SPAM	Image
[42]	2024	CC-GAIN	Electricity consumption data	Numeric

Table I summarizes previous research and is the basis for this research. Based on the research that has been described

previously, no previous research has implemented data preprocessing steps such as data classification, normalization, and denormalization before imputing missing data with the GAIN method. Incorporating these preprocessing steps can better prepare the data for the imputation process and potentially improve the overall performance of the method.

III. RESEARCH METHODOLOGY

Data completeness is one of several dimensions measured in determining data quality. As a data quality dimension, data completeness means the data set is free from missing values (MV or NA). Fig. 1 shows the flow carried out in this study. In this study, the steps taken are collecting data sets, creating data sets (simulation), and applying data sets to the proposed method, namely CN-GAIN and its basic method, GAIN. The last is to evaluate the data set through the prediction process.

A. Data Set

The data set used in this study is the SMES data set in South Sumatra, which was collected from 2017-2020 by the Dinas Koperasi dan UKM South Sumatra. This data is collected per period using several mechanisms, including through distributed data sheets or direct data collection by officers. There are 3301 SMES data records that were successfully collected. The fields include the type of business, manpower, investment_value, production_capacity, production_value, and bb_bp_value.

The performance of the proposed CN-GAIN imputation model is tested using SMES data as described in Table II with predetermined attributes. Table III shows the characteristics of the SMES dataset.

TABLE II. DATA SET UMKM

Type of business	manpower	Investment_value	Production_capacity	Production_value	bb_bp_value
Tempe	3	5000	75000	6000	20000
Tempe	3	5000	30000	30000	10000
Tempe	6	5000	75000	75000	25000
Tahu	1	5000	714000	285600	95200
Tahu	2	5000	36000	108000	36000
Tahu	2	2500	48000	14400	4800
Tahu	1	1500	90000	45000	15000
Batu bata	1	800	300000	165000	55000
Batako	3	85000	300000	750000	25000
Meuble	3	3000	1920	42350	14416
Meuble	3	10000	888	58000	19333
Meuble	3	15000	1800	52500	17500
Bengkel	3	10000	960	240000	8000
...

TABLE III. CHARACTERISTICS OF UMKM DATASET

Field	Data type	Description
Type of business	String	Types of SMES businesses
Manpower	Integer	Number of workers
Investment_value	Integer	Business investment value
Production_capacity	Integer	Production capacity
Production_value	Integer	Production value
bb_bp_value	Integer	Raw material value

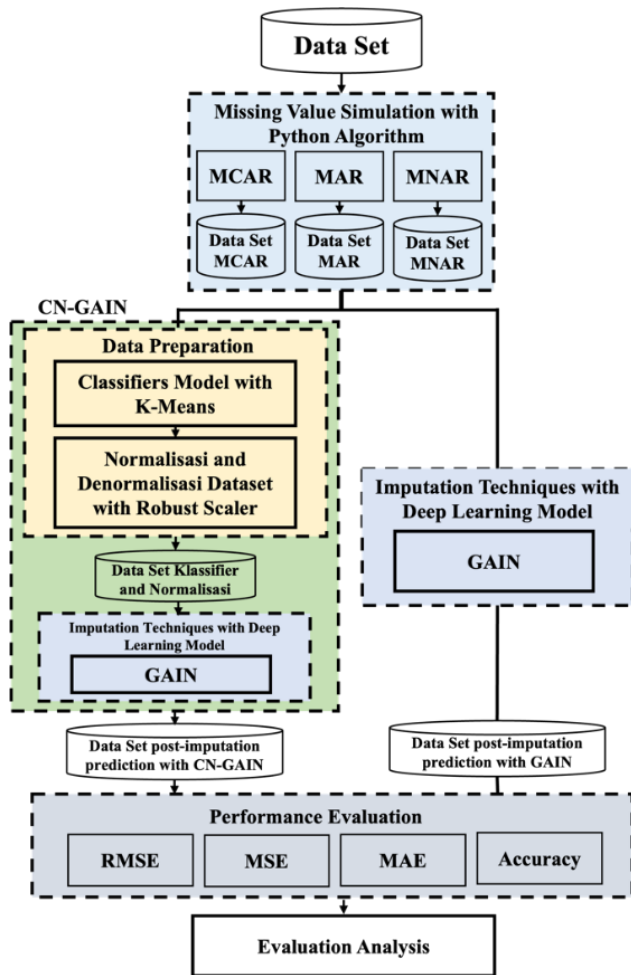


Fig. 1. Experimental flowchart.

B. Missing Value Simulation

Based on the dataset that has been obtained, the next step is to simulate the missing data. In this experiment using Python, the data set was randomly simulated into the MCAR, MNAR, and MAR categories. Algorithm 1 is a pseudo-code of the missing value simulation. By using this algorithm, complete SMES data is then removed randomly. Several Python libraries are used in this simulation, namely wget and numpy.

Algorithm 1. Pseudo-code of Missing Values simulation

```

import numpy as np
import pandas as pd
from utils import *
import torch
Function produce_NA(X, p_miss, mecha="MCAR",
opt=None, p_obs=None, q=None):
    If mecha is "MAR":
        mask = Generate MAR
    Else if mecha is "MNAR":
        mask = Generate MNAR
    Else:
        mask = Generate MCAR
    Return value
End Function
    
```

This function generates missing values in a dataset. The function relies on several libraries: numpy for numerical computations, pandas for data manipulation, and torch for deep learning tasks. The function takes the following parameters: X for the input dataset, p_miss for the proportion missing values, mecha for the mechanism for generating missing data, which can be MCAR (Missing Completely At Random), MAR (Missing At Random), or MNAR (Missing Not At Random), opt, p_obs, q are optional parameters. The function would return the dataset (X) modified with the generated missing values according to the selected mechanism.

C. Data Preparation

The data preparation stage is carried out to understand the characteristics of the data. This step consists of a collection of techniques applied to the data to improve the data quality before processing the machine learning data. Where the initial step taken is to carry out the Classifier model with K-Means, perform data normalization and data denormalization.

1) *Classifier model with k-means*: The k-means algorithm is the simplest and most commonly used clustering algorithm. This algorithm determines the number of clusters (k) that need to be grouped in a Data Set. The steps in performing clustering with the K-Means method include the following [43]:

- a) Determine the number of centroids
- b) Determine points or centroids randomly

$$D(x, y) = \sqrt{(X_1 - Y_1)^2 + (X_2 - Y_2)^2} \quad (1)$$

- c) Calculate and assign new centroids for each cluster.

$$c = \frac{\sum m}{n} \quad (2)$$

- d) Repeat step c, until there are no further changes.

2) *Normalization and denormalization dataset with robust scaler*: Normalization is done to change the value of the attribute in the dataset to have a uniform scale or range. Normalization is used to improve accuracy in classification. [44]. In this study, the normalization technique used is robust scaler. The Robust Scaler formula is stated in Eq. (3):

$$X_{scaled} = \frac{X - median(X)}{IQR(X)} \quad (3)$$

Denormalization using the Robust Scaler involves reversing the scaling process to convert the scaled data back to its original values. This is particularly useful when you want to interpret the results of your model in the context of the original data.

$$X_{Original} = (X_{scaled} \times IQR(X)) + Media(X) \quad (4)$$

D. Imputation Techniques with GAIN Model

Generative Adversarial Network (GAN) is an ML framework trained with two neural networks, namely the generator and the discriminator. The generator aims to create synthetic data that resembles real data, while the discriminator aims to distinguish between real and generated samples [46]. The GAN method is designed to generate images, GANs have been applied in various fields, including natural language processing and speech processing. The goal of GAN is to

generate images that are very similar to real images. GANs are designed for adversarial training of the generator (G) and the discriminator (D), where G is trained to create data that is most similar to the real data, and D is trained to classify the data generated by G. In the process of imputing missing data, GANs generate values that are similar to the real values by modeling the distribution of data surrounding the missing values.

1) *The generator*: The generator is trained for missing data imputation. In the generator model G, the input value is X_a and the matrix M and the noise variable Z are obtained.

$$X_m = X_a \odot M + Z \odot (1 - M) \quad (5)$$

$$X_f = X_a \odot M + G(X_m) \odot (1 - M) \quad (6)$$

2) *The discriminator*: D is used to distinguish the imputed data through G. Unlike GAN, it distinguishes between true and false data from certain constituent elements, not all generated data.

3) *Hint generator*: The hint (H) generator provides some information on the mask M to guide the training of D. This can prevent G and D from learning unintended distributions:

$$H = B \odot M + 0.5 \odot (1 - B) \quad (7)$$

E. Performance Evaluation

This phase is used to evaluate the performance of the proposed method. In this study, accuracy measurement was conducted, and three error metrics were employed to assess performance: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Squared Error (MSE). The rationale for using multiple error metrics is to comprehensively compare errors, as each metric offers different advantages, disadvantages, and features. The mathematical formula for calculating accuracy is presented in Eq. (8):

$$Accuracy = \frac{TN+TP}{TN+FN+FP+TP} \quad (8)$$

Root Mean Square Error (RMSE) calculates the error between the real value and the estimated value (imputed value) to measure the accuracy of imputation ([45], [46]). The difference between the predicted (hypothetical) value and the actual value. RSME is mathematically expressed as shown in Eq. (9):

$$RSME = \sqrt{\frac{\sum_{i=1}^n (X_{i}^{actual} - X_{i}^{imputed})^2}{n}} \quad (9)$$

Mean Absolute Error (MAE) is a matrix for calculating positive and negative deviations between \hat{y}_i predicted and actual values ([46], [47]). MAE is mathematically expressed as shown in Eq. (10):

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (10)$$

Mean Squared Error (MSE) calculates the average error. The average value is close to zero but not negative [46].

Mathematically, MSE is defined based on Eq. (11):

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (11)$$

IV. EXPERIMENTAL RESULTS

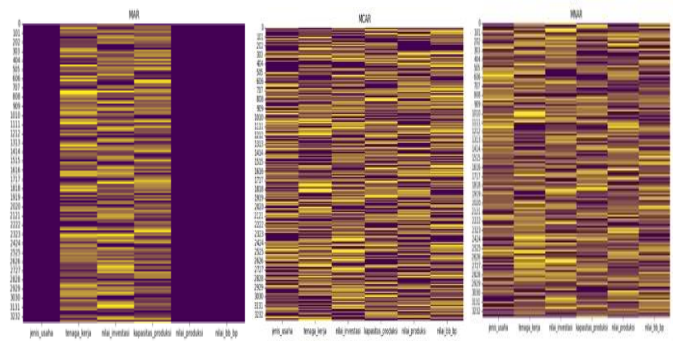
This section presents the results of each step that has been explained previously, namely the results of the missing value simulation, data pre-processing and the proposed CN-GAIN algorithm model.

A. Missing Values Simulation

Based on the python algorithm that has been created previously, where the data set is categorized into three categories of missing values, namely: MAR, MCAR and MNAR. Table IV is the result of the missing value algorithm process that has been run. While Fig. 2 is a visualization of the percentage of missing values based on the missing value category.

TABLE IV. PERCENTAGE OF MISSING VALUES IN EACH ATTRIBUTE

No	Field	MAR		MCAR		MNAR	
		MV	%	MV	%	MV	%
1	Type of business	0	0	129 3	39.1 7	130 4	39.5 0
2	Manpower	135 3	40.9 9	131 2	39.7 5	134 9	40.8 7
3	Investment_value	135 1	40.9 3	132 6	40.1 7	135 0	40.9 0
4	Production_capacity	133 4	40.4 1	132 1	40.0 2	131 7	39.9 0
5	Production_value	0	0	130 9	39.6 5	132 7	40.2 0
6	bb_bp_value	0	0	133 8	40.5 3	134 8	40.8 7



(a) MAR (b) MCAR (c) MNAR

Fig. 2. Missing value visualization.

B. Classification with K-Means

Classification is carried out on datasets based on business type. jenis_usaha column appears to contain information about the type of business, but there are many unique values with variations in capitalization and wording (e.g., "Bengkel motor" and "Bengkel Motor"). Additionally, some values are particular (e.g., "Bengkel"). Hence, classification is needed. To classify these values, K-means is employed. From the existing SMES data, 10 classifications were obtained, as shown in Table V.

TABLE V. CLASSIFICATION OF UMKM DATA

Cluster	Businesses
0	Crafts
1	Vehicle Repair and Maintenance
2	Brick-making
3	Furniture, Woodworking
4	Fish Products
5	Other product manufacturing
6	Tempeh and Tofu Production
7	Agriculture and Farming
8	Snack Production
9	Beverage

C. Architecture of the Proposed CN-GAIN Model

The proposed CN-GAIN model for UMKM data imputation is based on the GAIN method. This model consists of five main modules, namely: 1) clustering, 2) normalization, 3) denormalization, 4) generator, 5) discriminator.

Algorithm 1. Pseudo-code of CN-GAIN

```

from sklearn.preprocessing import RobustScaler
import numpy as np
import pandas as pd
import tensorflow as tf
from tensorflow.keras.layers import Input, Dense
from tensorflow.keras.models import Model
Determine the classification
Draw dataset, number of clusters k
  For i = 1, ..., do
    Label (i) ← clustering(k)
  end for
Normalisasi
scaler = Initialize RobustScaler()
data_scaled = Fit the scaler to the dataframe (df) and transform the data
Denormalization
median_values = Retrieve the median of each feature from the scaler
iqr_values = Retrieve the interquartile range (IQR) of each feature from the scaler
  For each feature in data_scaled:
    original_data = (data_scaled * iqr_values) + median_values
Update Generator and Discriminator
Discriminator optimization
While training loss has not converged do
Draw samples from the dataset
For j = 1, ..., samples do
 $X_m(j) \leftarrow G(X_a(j), z(j))$ 
 $X_f(j) \leftarrow m(j) \odot (X_a(j) + (1 - m(j)) \odot X_m(j))$ 
 $h(j) \leftarrow b(j) \odot m(j) + 0.5(1 - b(j))$ 
 $y(j) \leftarrow C(X_f(j), l(j))$ 
end for
update D using adam optimizer
Generator Optimization
Draw samples from the dataset
For j = 1, ..., do
 $h(j) \leftarrow b(j) \odot m(j) + 0.5(1 - b(j))$ 
 $y(j) \leftarrow C(X_f(j), l(j))$ 
end for
update D using adam optimizer

```

Data was normalized using a Robust Scaler by subtracting the median and dividing the Inter Quartile Range (IQR). Then, data was deformedalized using Robust Scaler. The optimization

begins with discriminator D, which is tuned using a fixed generator and classification via mini-batch sD. Independent samples of Z and B, represented as z(j) and b(j), generate h(j) and Xf(j), respectively. Additionally, y(j) is derived from Xf(j) and l(j). The discriminator is optimized using Xf(j), h(j), and y(j) across all mini-batches. Then, generator G will be optimized by mini-batch s(G) while the discriminator D is updated, and the classification C remains fixed. h(j) and y(j) are computed for all mini-batches and used in optimizing G.

V. RESULT AND DISCUSSION

The dataset prepared and simulated into missing data categories (MAR, MCAR, and MNAR) is then used with the proposed steps and techniques in the data processing stage.

A. Result

The performance of the missing value imputation of the proposed CN-GAIN model is compared with the baseline model, namely the GAIN Model. Each model was tested on three missing value data categories: MAR, MCAR, and MNAR. Where each category has a different level of missing value, in this study, the percentage of missing values was created randomly using the library in python for each type of missing value category. This is different from what was done in the study [42], where the percentage of missing values ranges from 10% to 90%.

The proposed CN-GAIN model has a better accuracy rate and a low error rate. An important step is applying classification, normalization, and denormalization of data before the imputation process with the GAIN model. As a comparison, researchers also applied the GAIN model to compare accuracy and error rates. Table VI shows the results of the proposed model trial with the base model after imputation.

TABLE VI. PERFORMANCE EVALUATION OF PROPOSED CN-GAIN AND GAIN

Metrics	MAR		MCAR		MNAR	
	CN-GAIN	GAIN	CN-GAIN	GAIN	CN-GAIN	GAIN
Accuracy	0.997	0.997	1.00	0.9992	1.00	0.9992
RSME	0.013	0.016	0.0077	0.0097	0.0042	0.0082
MSE	0.0007	0.0015	0.0007	0.0018	0.0006	0.0020
MAE	0.0007	0.177	0.0007	0.066	0.017	0.448

Based on the results of the trials conducted for accuracy, the MAR missing value category has the same value, which is 0.9970. for the MCAR category, the CN-GAIN method has a better accuracy level of 1.00 while the GAIN method is 0.9992. as well as for the value in the MNAR missing value category. For performance seen from the Root Mean Square Error (RMSE) has a better value, whereas the MNAR missing value category has a better value, which is 0.0042. While the performance for the Mean Absolute Error (MAE) shows a better value than this method is MAR, which is 0.0007. Finally, the performance of the Mean Squared Error (MSE) category of MNAR missing value has the best value of the proposed model, which is 0.0006. Fig. 3 compares the CN-GAIN and the GAIN methods as the basic methods for three types of missing values: MAR, MCAR, and MNAR.

Evidence of the proposed model's improvement is demonstrated by the increased performance percentage of the CN-GAIN model compared to the base model, GAIN. Fig. 4 is a visualization of the presentation of CN-GAIN performance on each type of missing value compared to the basic GAIN model. The imputation value's accuracy level occurred in the MCAR missing value category type of 0.0801% and MNAR of 0.0800, while for the MAR category type there was no increase. For the percentage of actual and imputed values (RMSE), the MNAR missing value category type has a better improvement of 48.780%.

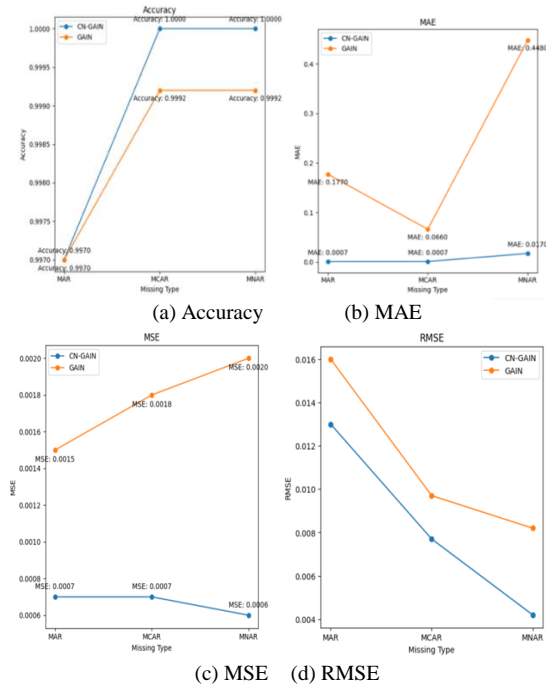


Fig. 3. Performance evaluation of different methods.

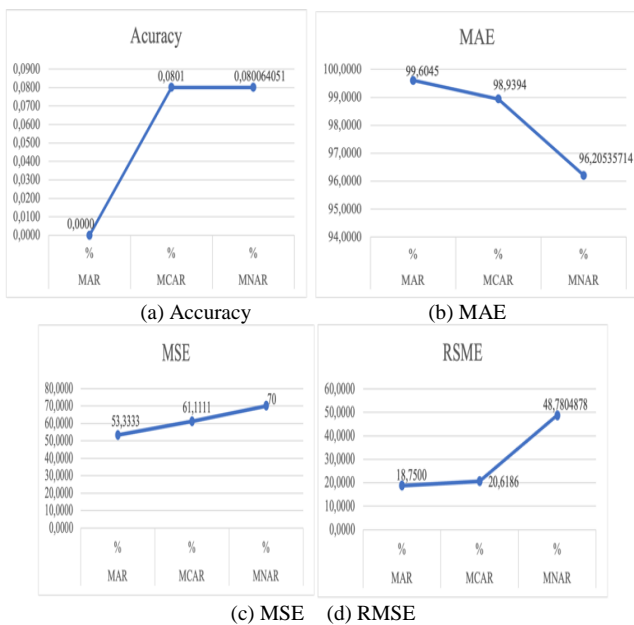


Fig. 4. Performance percentage of CN-GAIN compared to GAIN method on missing value type.

B. Discussion

Based on the results of the trials conducted, the CN-GAIN model has a better level of accuracy than its basic model, the GAIN model. While for the error rate the CN-GAIN method also has a lower error rate. The results (Table V) show that of the three types of missing values simulated, the CN-GAIN model has a relatively high increase in accuracy; only the MAR type of missing value has the same value when the accuracy level is measured. These results prove that the CN-GAIN model as a framework for handling missing values is proposed to be implemented. The classification steps carried out in data preparation make the data set classified according to its group and normalization-denormalization makes the data set in a value range that is not too far apart. This step improved the imputation process, even though there was no change or significant improvement in one type of missing value, especially in terms of accuracy in the MAR type of missing value.

From the overall results, this study has proven that the framework developed using the proposed classification and normalization-denormalization has a better level of accuracy than its standard. This study also proves that the data preprocessing carried out has a better impact on the quality of the data obtained after the value prediction is carried out.

VI. CONCLUSION AND FUTURE WORK

Incomplete SMES data is a significant challenge for SMES' proper management and development. Effective data imputation is essential to produce quality SMES data. In this study, we propose CN-GAIN, a new missing data imputation method designed to handle data with multiple data characteristics in SMES data. By making efforts to classify, normalize, and denormalize data in the data pre-processing process before imputation using the GAIN method. The CN-GAIN model performs better in predicting missing values, with an accuracy value of 0.0801% for the MCAR category and a lower error rate (RMSE), of 48.78% for the MNAR category. The average error (MSE) is 99.60% for the MAR category, and the deviation value (MAE) is 70% for the MNAR category.

For further research, researchers will test the model on other data sources with more complex data characteristics with more varied data types.

REFERENCES

- [1] M. Souibgui, F. Atgui, S. Zammali, S. Cherfi, and S. Ben Yahia, "Data quality in ETL process: A preliminary study," *Procedia Comput Sci*, vol. 159, pp. 676–687, 2019, doi: 10.1016/j.procs.2019.09.223.
- [2] M. P. Fernando, F. César, N. David, and H. O. José, "Missing the missing values: The ugly duckling of fairness in machine learning," *International Journal of Intelligent Systems*, vol. 36, no. 7, pp. 3217–3258, Jul. 2021, doi: 10.1002/int.22415.
- [3] D. Li, H. Zhang, T. Li, A. Bouras, X. Yu, and T. Wang, "Hybrid Missing Value Imputation Algorithms Using Fuzzy C-Means and Vaguely Quantified Rough Set," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 5, pp. 1396–1408, May 2022, doi: 10.1109/TFUZZ.2021.3058643.
- [4] G. Doquire and M. Verleysen, "Feature selection with missing data using mutual information estimators," *Neurocomputing*, vol. 90, pp. 3–11, Aug. 2012, doi: 10.1016/j.neucom.2012.02.031.
- [5] T. Emmanuel, T. Maupong, D. Mpoeleng, T. Semong, B. Mphago, and O. Tabona, "A survey on missing data in machine learning," *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00516-9.

- [6] Z. Chen, S. Tan, U. Chajewska, C. Rudin, and R. Caruana, "Missing Values and Imputation in Healthcare Data: Can Interpretable Machine Learning Help?," 2023.
- [7] R. Shahbazian and I. Trubitsyna, "DEGAIN as tool for Missing Data Imputation," 2023. [Online]. Available: <http://ceur-ws.org>
- [8] M. W. Huang, W. C. Lin, C. W. Chen, S. W. Ke, C. F. Tsai, and W. Eberle, "Data preprocessing issues for incomplete medical datasets," *Expert Syst*, vol. 33, no. 5, pp. 432–438, Oct. 2016, doi: 10.1111/exsy.12155.
- [9] T. Thomas and E. Rajabi, "A systematic review of machine learning-based missing value imputation techniques," *Data Technologies and Applications*, vol. 55, no. 4, pp. 558–585, 2021, doi: 10.1108/DTA-12-2020-0298.
- [10] A. R. Ismail, N. Z. Abidin, and M. K. Maen, "Systematic Review on Missing Data Imputation Techniques with Machine Learning Algorithms for Healthcare," Mar. 01, 2022, Department of Electrical Engineering, Universitas Muhammadiyah Yogyakarta. doi: 10.18196/jrc.v3i2.13133.
- [11] I. Setiawan, R. Gernowo, and B. Warsito, "A Systematic Literature Review on Missing Values: Research Trends, Datasets, Methods and Frameworks," in *E3S Web of Conferences*, EDP Sciences, Nov. 2023. doi: 10.1051/e3sconf/202344802020.
- [12] F. Yulian Pamuji, Ahmad Rofiqul Muslikh, Rizza Muhammad Arief, and Delviana Muti, "Komparasi Metode Mean dan KNN Imputation dalam Mengatasi Missing Value pada Dataset Kecil," *Jurnal Informatika Polinema*, vol. 10, no. 2, pp. 257–264, Feb. 2024, doi: 10.33795/jip.v10i2.5031.
- [13] N. Karmitsa, S. Taheri, A. Bagirov, and P. Makinen, "Missing Value Imputation via Clusterwise Linear Regression," *IEEE Trans Knowl Data Eng*, vol. 34, no. 4, pp. 1889–1901, Apr. 2022, doi: 10.1109/TKDE.2020.3001694.
- [14] A. Dubey and A. Rasool, "Clustering-Based Hybrid Approach for Multivariate Missing Data Imputation," 2020. [Online]. Available: www.ijacsa.thesai.org
- [15] W. Sudrajat and I. Cholid, "K-NEAREST NEIGHBOR (K-NN) UNTUK PENANGANAN MISSING VALUE PADA DATA UMKM," 2023.
- [16] A. Syarif, O. Desti Riana, D. Asiah Shofiana, and A. Junaidi, "A Comprehensive Comparative Study of Machine Learning Methods for Chronic Kidney Disease Classification: Decision Tree, Support Vector Machine, and Naive Bayes." [Online]. Available: www.ijacsa.thesai.org
- [17] S. Nikfalazar, C. H. Yeh, S. Bedingfield, and H. A. Khorshidi, "Missing data imputation using decision trees and fuzzy clustering with iterative learning," *Knowl Inf Syst*, vol. 62, no. 6, pp. 2419–2437, Jun. 2020, doi: 10.1007/s10115-019-01427-1.
- [18] A. Syarif, O. Desti Riana, D. Asiah Shofiana, and A. Junaidi, "A Comprehensive Comparative Study of Machine Learning Methods for Chronic Kidney Disease Classification: Decision Tree, Support Vector Machine, and Naive Bayes." [Online]. Available: www.ijacsa.thesai.org
- [19] A. R. Alsaber, J. Pan, and A. Al-Hurban, "Handling complex missing data using random forest approach for an air quality monitoring dataset: A case study of kuwait environmental data (2012 to 2018)," *Int J Environ Res Public Health*, vol. 18, no. 3, pp. 1–26, 2021, doi: 10.3390/ijerph18031333.
- [20] "Mixed Data Imputation using Generative Adversarial Networks".
- [21] H. Ou, Y. Yao, and Y. He, "Missing Data Imputation Method Combining Random Forest and Generative Adversarial Imputation Network," *Sensors*, vol. 24, no. 4, Feb. 2024, doi: 10.3390/s24041112.
- [22] R. Shahbazian and S. Greco, "Generative Adversarial Networks Assist Missing Data Imputation: A Comprehensive Survey and Evaluation," *IEEE Access*, vol. 11, pp. 88908–88928, 2023, doi: 10.1109/ACCESS.2023.3306721.
- [23] W. Dong et al., "Generative adversarial networks for imputing missing data for big data clinical research," *BMC Med Res Methodol*, vol. 21, no. 1, Dec. 2021, doi: 10.1186/s12874-021-01272-3.
- [24] J. Gao, Z. Cai, W. Sun, and Y. Jiao, "A Novel Method for Imputing Missing Values in Ship Static Data Based on Generative Adversarial Networks," *J Mar Sci Eng*, vol. 11, no. 4, Apr. 2023, doi: 10.3390/jmse11040806.
- [25] M. K. Hasan, M. A. Alam, S. Roy, A. Dutta, M. T. Jawad, and S. Das, "Missing value imputation affects the performance of machine learning: A review and analysis of the literature (2010–2021)," Jan. 01, 2021, Elsevier Ltd. doi: 10.1016/j.imu.2021.100799.
- [26] J. Yoon, J. Jordon, and M. van der Schaar, "GAIN: Missing Data Imputation using Generative Adversarial Nets," Jun. 2018, [Online]. Available: <http://arxiv.org/abs/1806.02920>
- [27] J. Yoon, J. Jordon, and M. van der Schaar, "GAIN: Missing Data Imputation using Generative Adversarial Nets," Jun. 2018, [Online]. Available: <http://arxiv.org/abs/1806.02920>
- [28] A. M. Sefidian and N. Daneshpour, "Missing value imputation using a novel grey based fuzzy c-means, mutual information based feature selection, and regression model," *Expert Syst Appl*, vol. 115, pp. 68–94, Jan. 2019, doi: 10.1016/j.eswa.2018.07.057.
- [29] H. Rosado-Galindo and S. Dávila-Padilla, "Tree-Based Missing Value Imputation Using Feature Selection," *Journal of Data Science*, vol. 18, no. 4, pp. 606–631, Oct. 2020, doi: 10.6339/JDS.202010_18(4).0002.
- [30] M. El-Bakry, A. El-Kilany, S. Mazen, and F. Ali, "Fuzzy based Techniques for Handling Missing Values." [Online]. Available: www.ijacsa.thesai.org
- [31] J. Yoon, J. Jordon, and M. Van Der Schaar, "GAIN: Missing Data Imputation using Generative Adversarial Nets," 2018.
- [32] Z. A. Nadzurah, I. Amelia Ritahani, and A. Nurul, "Performance Analysis of Machine Learning Algorithms for Missing Value Imputation," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 6, 2018.
- [33] I. Goodfellow et al., "Generative adversarial networks," *Commun ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.
- [34] D. Lee, J. Kim, W.-J. Moon, and J. C. Ye, "CollaGAN: Collaborative GAN for Missing Image Data Imputation."
- [35] S. Wang, W. Li, S. Hou, J. Guan, and J. Yao, "STA-GAN: A Spatio-Temporal Attention Generative Adversarial Network for Missing Value Imputation in Satellite Data," *Remote Sens (Basel)*, vol. 15, no. 1, Jan. 2023, doi: 10.3390/rs15010088.
- [36] X. Zheng, Y. Wu, Y. Pan, W. Lin, L. Ma, and J. Zhao, "DPGAN: A Dual-Path Generative Adversarial Network for Missing Data Imputation in Graphs." [Online]. Available: <https://github.com/momoxia/DPGAN>.
- [37] W. Qiu, Y. Huang, and Q. Li, "IFGAN: Missing Value Imputation using Feature-specific Generative Adversarial Networks," in *Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 4715–4723. doi: 10.1109/BigData50022.2020.9378240.
- [38] M. K. Gunady, J. Kancherla, H. Corrada Bravo, and S. Feizi, "scGAIN: Single Cell RNA-seq Data Imputation using Generative Adversarial Networks", doi: 10.1101/837302.
- [39] S. Yoon and S. Sull, "Gamin: Generative adversarial multiple imputation network for highly missing data," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, 2020, pp. 8453–8461. doi: 10.1109/CVPR42600.2020.00848.
- [40] Y. Wang, D. Li, X. Li, and M. Yang, "PC-GAIN: Pseudo-label Conditional Generative Adversarial Imputation Networks for Incomplete Data," Nov. 2020, [Online]. Available: <http://arxiv.org/abs/2011.07770>
- [41] R. Shahbazian and I. Trubitsyna, "DEGAIN as tool for Missing Data Imputation," 2023. [Online]. Available: <http://ceur-ws.org>
- [42] J. Hwang and D. Suh, "CC-GAIN: Clustering and classification-based generative adversarial imputation network for missing electricity consumption data imputation," *Expert Syst Appl*, vol. 255, Dec. 2024, doi: 10.1016/j.eswa.2024.124507.
- [43] W. Sudrajat, I. Cholid, and J. Petrus, "Wahyu Sudrajat et al, Penerapan Algoritma K-Means Untuk
- [44] A. Khoirunnisa and N. G. Ramadhan, "Improving malaria prediction with ensemble learning and robust scaler: An integrated approach for enhanced accuracy," *JURNAL INFOTEL*, vol. 15, no. 4, pp. 326–334, Nov. 2023, doi: 10.20895/infotel.v15i4.1056.
- [45] M. F. Dzulkalnine and R. Sallehuddin, "Missing data imputation with fuzzy feature selection for diabetes dataset," *SN Appl Sci*, vol. 1, no. 4, Apr. 2019, doi: 10.1007/s42452-019-0383-x.

- [46] I. Gad, D. Hosahalli, B. R. Manjunatha, and O. A. Ghoneim, "A robust deep learning model for missing value imputation in big NCDC dataset," *Iran Journal of Computer Science*, vol. 4, no. 2, pp. 67–84, Jun. 2021, doi: 10.1007/s42044-020-00065-z.
- [47] J. H. Li et al., "Comparison of the effects of imputation methods for missing data in predictive modelling of cohort study datasets," *BMC Med Res Methodol*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12874-024-02173-x.

An Agile Approach for Collaborative Inquiry-Based Learning in Ubiquitous Environment

Bushra Fazal Khan, Sohaib Ahmed

Department of Software Engineering, Bahria University, Karachi, Pakistan

Abstract—The use of collaborative inquiry-based learning has been prevalent in educational contexts particularly in science education. Using such collaborative environments, learners can increase their engagement, knowledge and critical thinking skills about science. With the advancement of technologies, ubiquitous learning environments have been designed for facilitating learning in real-time contexts. Over the past few years, agile-based approaches have been implemented at higher education for inquiry-based learning activities. However, there is a lack of studies found that focuses on agile-based approach for ubiquitous collaborative inquiry learning activities at K-12 education level. Therefore, this study presents the ScrumBan Ubiquitous Inquiry Framework (SBUIF), for inquiry-based learning activities at K-12 education level. For this purpose, an application uASK has been developed on the proposed framework, SBUIF. For the evaluation purposes, computer-supported collaborative learning (CSCL) affordances along with micro and meso levels of the M3 evaluation framework has been applied. An experiment was conducted for the evaluation of uASK application in comparison with the Trello application, involving 205 to 127 seventh-grade students. Results demonstrated that uASK learners achieved higher scores as compared with Trello participants. Further, survey results indicated higher levels of engagement, satisfaction, and enjoyment among uASK users. The study concludes that uASK offers significant advantages over Trello in fostering collaborative inquiry-based learning activities in ubiquitous environment.

Keywords—K12 education; agile; ubiquitous; collaborative learning; inquiry based learning

I. INTRODUCTION

Due to the shifting context and evolving learner's needs the educational landscape faces many challenges. By introducing collaborative activities in their education, students not only develop content mastery but also enhance their communication skills, problem-solving abilities, and overall academic motivation [1]. To align students with real-world experiences they will encounter in the workspace thus preparing them for the collaborative nature of many professional environments Collaborative learning fosters many skills in students, encouraging the creation of new knowledge and understanding through dialogue and interaction [2]. An inquiry-based learning approach enables learners to actively explore questions, construct knowledge, and share findings, and has been recognized as an effective method for developing critical thinking skills [3]. Inquiry-based learning (IBL) is considered as an effective educational approach that encourages learners to engage in the learning process more deeply by taking care of their own learning, rather than passively receiving information [4].

With the advancement of technologies, the use of mobile devices in IBL environments presents unique opportunities for learning especially science education [5]. They are: (i) facilitate various levels of inquiry (ii) enhance learners' motivation and engagement (iii) provides seamless learning among various contexts and (iv) leverage between formal and informal science education. The use of mobile devices with sensor technologies can create more interactive and immersive experiences for learners that can actively engage in the learning process. This learning environment is termed as Ubiquitous learning or U-learning [6]. Recent studies highlighted that this learning environment can effectively provide benefits in learners' engagement and critical thinking skills [7][8]. Further, the use of such ubiquitous platforms can enhance student learning by capturing their attention and providing personalized learning experiences [9].

In the literature related to ubiquitous IBL, five types are identified [10]: (i) *Authentic scientific inquiry* in which learners investigate and conclude about real-world or scientific problem (ii) *Abductive science inquiry* in which learners form hypotheses based on research or observations, drawing conclusions through critical thinking (iii) *Collaborative inquiry* in which learners work in groups, understand and solve problems through a repeated process [11] (iv) *Collective whole class inquiry* in which the entire class participates in the inquiry process, collaborating towards a common goal. Learners may think critically and generate ideas for a given problem [12] and (v) *Inquiry with game component* in which learners use game component as an instructional or learning source to address inquiry problems such as combining augmented reality with inquiry environment to enhance their learning experiences [13][14]. This research follows collaborative inquiry for learning science education.

Collaborative inquiry-based learning (CIBL) has gained popularity due to its ability to enhance learners' engagement, comprehend knowledge, and foster critical thinking skills [11]. In these environments, teachers and peers provide an environment, pose problems, and offer support that promotes intellectual growth. When students work together in cooperative groups, they share the process of developing ideas. This collaboration gives them opportunities to think about and expand their own ideas as well as those of their peers.

In these collaborative settings, students can view their peers as resources, not competitors [15]. Group-based exploration allows students to collaboratively exchange ideas and experiences. Differing interpretations of concepts and processes can lead to disagreements, which may ultimately promote intellectual growth through the sharing of knowledge

and the formation of connections [12]. However, such a learning environment faces challenges in terms of processes and resources for collaboration, and requires support from fellow students, which demands energy investment in collaborative efforts.

On the other hand, the implementation of agile methodologies within the educational domain has been a topic of growing interest in both classroom and online learning environments. This approach creates an adaptive and collaborative learning experience, owing to its iterative and incremental nature [16]. Further, Agile has been successfully utilized in academic settings to teach software engineering, primarily through project-based learning approaches. In these scenarios, students work in small groups to produce software as the final learning product, simulating agile practices [17] [18]. However, the existing literature reveals a paucity of empirical research specifically addressing the adoption of any agile methodology in secondary and higher education contexts.

Scrumban, one of the agile methodologies, follows a hybrid project management framework that combines the iterative and incremental nature of Scrum with the continuous flow of Kanban [19]. This framework has been successfully applied in software development and has the potential to be adapted for instructional design and delivery in other learning environments [20]. The integration of Scrumban and inquiry-based learning in ubiquitous environment may create an approach that leverages the strengths of both methodologies. Nevertheless, such integration has not been found in the literature earlier. Therefore, this allows us to explore the use of scrumban methodology as a collaborative framework for inquiry-based learning activities at secondary school students.

This paper presents a framework that combines Scrumban principles with inquiry-based learning for ubiquitous learning environments. To assess the effectiveness of this framework, an application, 'uASK' has been developed and evaluated using the M3 evaluation framework [21] was employed. The paper later discusses the findings and their implications.

II. RELATED WORK

In the literature, there are few applications that use agile based approaches in educational settings. These studies mainly focus on IBL activities in ubiquitous environments as depicted in Table I.

ULMCI is a forum-based web application to assist learners through authentic scientific inquiry [20]. In this study, university students work on solving real-world problems to develop their programming skills, rather than just learning theory. The forum allows learners to interact, discuss, and collaborate on programming challenges. However, no agile method was used for collaboration in this [22]. In another study, two digital platforms are utilized for conducting authentic scientific inquiry for railway engineering university students [23]: 1) Edmodo, a virtual environment for disseminating educational resources and facilitating communication, and 2) Google Docs, a collaborative online writing and editing tool. These platforms were implemented using flipped classroom

approach where learners can independently study core course material outside the classroom, while in-class time is allocated to practical applications and knowledge reinforcement activities.

A research study was conducted in which university students used a digital tool called Trello for learning software engineering [24]. This tool follows Kanban, an agile methodology for managing and organizing learners' activities. Further, inquiry with game component type is applied to make learning process for engaging and interactive. In another instance, scrum methodology is used for teaching Chemistry to 11th grade students [25]. In this study, learners are involved for enhancing their engagements and learning outcomes through manual boards and face-to-face meetings rather any digital tools.

On the other hand, there are few studies that target collaborative inquiry types. In this study, scrumban methodology are followed by university students for learning about project-based web programming courses [26]. The participants in this study used Microsoft Planner, WhatsApp, and Telegram to enhance student engagement, collaboration, and project management skills. Microsoft Planner facilitated task organization and progress tracking, while WhatsApp and Telegram enabled real-time communication [26]. This approach aimed to create a collaborative inquiry earning environment mirroring real-world software development practices, potentially better preparing students for future careers.

In another research, Milićević et al. (2019) integrated Scrum methodology with OpenProject software in e-business project management courses. This approach provides university students with practical experience in agile project management through real-world e-business projects. The combination of Scrum and OpenProject facilitated collaborative inquiry, task allocation, progress tracking, and iterative development [27]. In a similar vein, Parsons et al. (2018) explored an innovative approach to teacher professional development by integrating Scrum and Kanban methodologies, commonly known as Scrumban, and implementing it through the digital platform Trello. This hybrid approach supports and tracks collaboration, allowing all team members to participate in discussions, view the workflow, share files and notes [28]. Further, learning sessions in this research facilitates participants to comprehend practical application of agile and lean concepts to enhance their professional practice.

The relevant studies indicate that the use of agile methodologies in educational context enhance learning experience for the learners [29]. Most of these studies focused on university students except for a study where high school students are involved [25]. However, that study did not use any ubiquitous inquiry type. Therefore, this indicates that there is a significant gap in literature that can implement an agile-based approach for collaborative inquiry-based learning activities in a ubiquitous environment for school students. This gap presents an opportunity in this research to explore a ubiquitous learning environment through an agile-based approach for conducting collaborative inquiry activities.

TABLE I. AGILE APPROACHES USED IN INQUIRY-BASED LEARNING UBIQUITOUS ENVIRONMENTS

Approach/Application	Agile Method	Platform	Learners	Domain Knowledge	Ubiquitous Inquiry Type
ULMCI [22]	-	Forum based application (ULMCI)	University Students	Programming	Authentic scientific inquiry
Agile CSCL [26]	Scrumban	MS Planner, Whatsapp and Telegram	University Students	Web Programming	Collaborative inquiry
IBL-CSCL in flipped classroom [23]	-	Edmodo, and Google Docs	University Students	Railway engineering	Authentic scientific inquiry
GBL Agile [24]	Kanban	Trello	University Students	Software Engineering	Inquiry with game component
Context Based Scrum [25]	Scrum	-	Grade 11	Chemistry Education	-
Scrum Agile Framework [27]	Scrum	OpenProject	University Students	E-Business Project Management	Collaborative inquiry
Agile and Lean learning [28]	Scrumban	Trello	Teachers	Agile and Lean Teaching Concepts	Collaborative inquiry

III. SCRUMBAN UBIQUITOUS INQUIRY FRAMEWORK

This study presents an innovative framework, the ScrumBan Ubiquitous Inquiry Framework (SBUIF), as shown in Fig. 1, which adapts various components from the established methods including Seamless inquiry-based learning framework (SIBLF) [30], and ScrumBan Research Framework (SBRF) [31]. The proposed SIBLF by Song et al. [30] lacks a structured approach, which in turn complicates class management. While SBRF provides structural stability, it does not incorporate inquiry elements for student scaffolding. Additionally, its design for research project management in university education renders it unsuitable for knowledge construction at the K12 level. However, by integrating these elements, SBUIF provides a comprehensive and flexible inquiry framework suitable for use in educational environments for its adaptability to real-world contexts through ubiquitous learning.

The SBUIF is categorized into eight distinct phases, each with its own unique characteristics and goals. Its objective is to create a comprehensive and dynamic learning environment that caters to the diverse needs and preferences of students. By utilizing a combination of learner-centered approaches and technology-enhanced resources, this framework strives to enhance student engagement, collaboration, and the overall quality of the educational experience. This framework also equips students with the necessary skills to engage in scientific inquiry as they develop their ability to apply critical thinking and reasoning to their observations and experiments in a ubiquitous environment. The phases of SBUIF are as under:

A. Build Awareness

The initial phase of "Build Awareness" is essential for establishing the foundation necessary for the effective implementation of the SBUIF approach. This phase focuses on building strong relationships with educators, administrators, and the learning community. This phase develops assessment

procedures that meet learning objectives and encourage stakeholder feedback. These links promote varied viewpoints and insights, improving understanding of learning objectives and context. Further, it emphasizes the need for a thorough literature review to build expertise. This review ensures that learners, teachers, and other stakeholders understand the latest research, best practices, and emerging trends in the field, laying the groundwork for the SBUIF approach.

B. Define Objective

Learners' perspectives, needs, and goals are analyzed in the second phase, "Define Objective". This stage tailors' educational programs to individual learners, making them more effective and interesting. Understanding learners' viewpoints helps instructors to create an environment that motivates and empowers them. The instructional team determines the most relevant and interesting content for the learning objectives by examining the target audience. After determining the target audience, the team can create learning objectives that match learners' needs and interests, making the content engaging and effective. This phase is critical for evaluating the learning goals alignment process to boost student engagement and involvement in their education.

C. Engage

In the third phase, learners past knowledge and experiences are activated and enhanced while meaningful connections are made between peers and instructors. Therefore, it uses collaborative and blended learning approaches to provide an engaging and participatory learning environment where learners can actively contribute knowledge, share thoughts, and have meaningful discussions. This phase can help learners to develop critical thinking and collaborative skills while giving instructors a better understanding of their progress and needs. This level uses advanced technology and a collaborative learning environment to encourage students to participate, share ideas, and learn more.

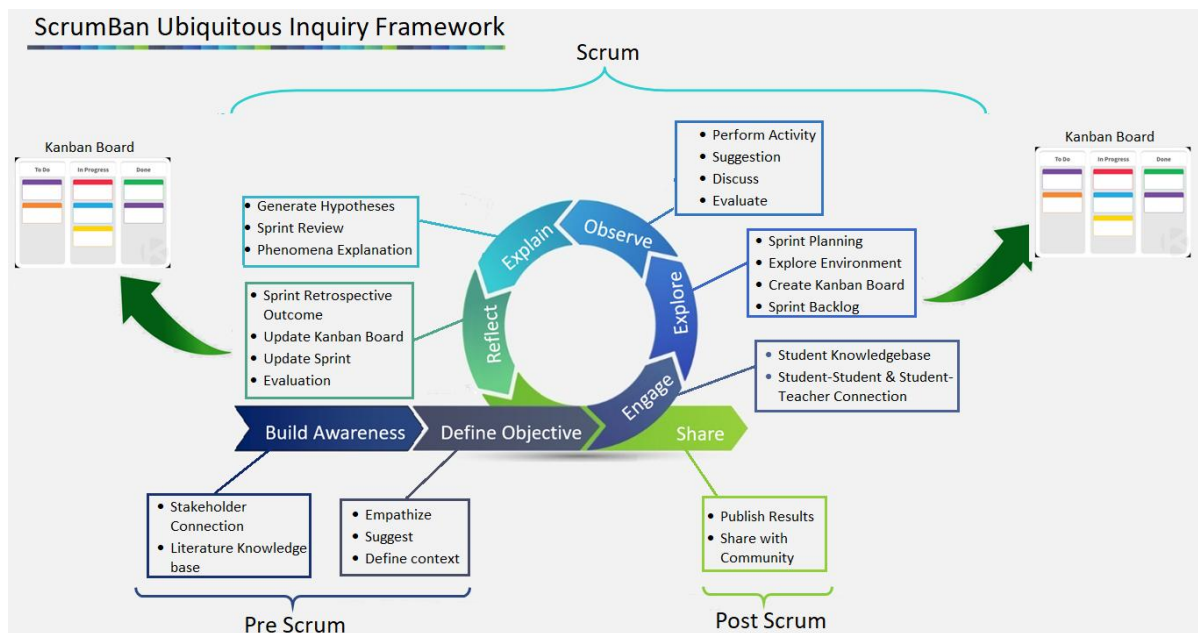


Fig. 1. Proposed ScrumBan ubiquitous inquiry framework adapted from study [30] [32].

D. Explore

Lesson design is systematic and iterative in the fourth phase, "Explore". The instructor and learners plan learning sprints during sprint planning. The sprint planning process helps instructors and learners grasp learning sprint objectives, goals, and expectations, improving learning outcomes. To ensure effective and efficient learning outcomes, educators must carefully develop and define learning sprint objectives, goals, and expectations. Instructors and students evaluate resources, technologies, and learning barriers in the Explore phase. The exploration phase includes creating a Kanban board and sprint backlog to organize and visualize learning. The Kanban board may show the status of each work, helping the team communicate progress and identify areas for improvement. Kanban boards and sprint backlogs boost teamwork and learning. This phase allows instructors to discover and correct student misconceptions and gaps in comprehension, which may be addressed with targeted interventions and additional support to improve learning.

E. Observe

In this phase, the instructor monitors student performance and provides real-time feedback and advice to improve learning and fulfill various requirements. To facilitate learning and meet various needs, the instructor analyzes learners' involvement and performance during learning activities and provides real-time feedback and advice. This phase promotes ongoing improvement through detailed discussions and activity evaluations. This allows the instructor to constantly enhance and alter the teaching method, creating a more inclusive and effective learning environment that boosts learners' engagement and educational quality.

F. Explain

During the "Explain" phase, learners are instructed to

develop hypotheses, assess the advancements achieved throughout the sprint, and obtain detailed explanations of the foundational concepts and principles. This phase is essential for facilitating learners' connections between observed phenomena and underlying theories, thereby enhancing their understanding of the subject matter. This phase offers opportunities for learners to refine hypotheses, challenge assumptions, and enhance insights through interactive discussions and constructive feedback from instructors and peers. Engaging in meaningful exchanges allows learners to explore alternative perspectives, identify new connections, and enhance their critical thinking skills. This collaborative process facilitates a nuanced and comprehensive understanding of the subject matter, thereby enhancing the ability to draw informed conclusions and address complex problems.

G. Reflect

In this phase, a comprehensive review and analysis of the entire learning process is conducted. The sprint retrospective involves a discussion between the instructor and learners regarding the outcomes, effectiveness, and implications of the sprint. The instructor and learners assess learning experiences, pinpoint areas for enhancement, and investigate methods to refine sprint instruction. The pedagogical framework undergoes continuous refinement and optimization through rigorous evaluation and debate, ensuring a highly responsive learning experience tailored to learners' diverse requirements and preferences. The Reflect phase includes the analysis of learning outcomes, the updating of the Kanban board and sprint schedules, and the execution of comprehensive mid-term evaluations. This review and assessment process enables both the instructor and learners to evaluate previous sprints, pinpoint areas for enhancement and modify the Kanban board and sprint plans accordingly.

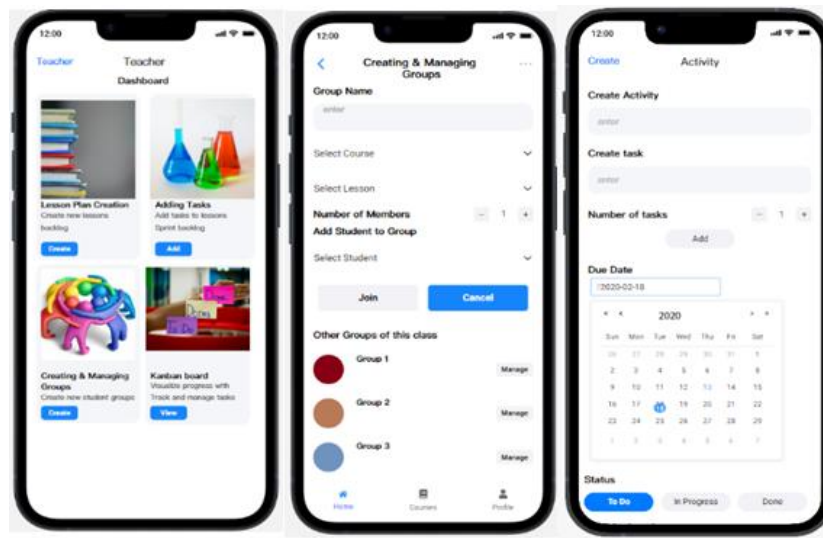


Fig. 2. uASK pre-scrum interfaces.

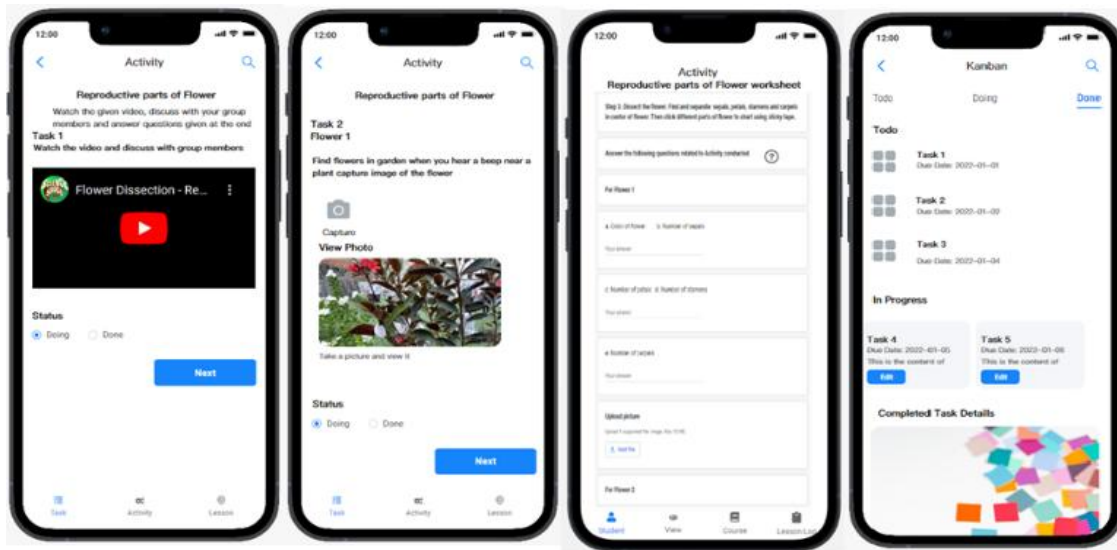


Fig. 3. Scrum activities in uASK.

H. Share

The final phase entails the publication and dissemination of research findings, best practices, and innovative pedagogical approaches developed through the implementation of the student-centered and blended learning framework. This phase of sharing promotes cross-institutional collaboration and knowledge exchange, enabling instructors from various educational institutions to convene and discuss their experiences, best practices, and innovative strategies. The feedback and perspectives obtained from the broader academic community can inform future iterations of the framework, thereby enhancing its effectiveness in addressing the diverse needs and learning styles of learners across various educational contexts. This phase of sharing is essential for the ongoing refinement and enhancement of the overall framework, facilitating the integration of diverse insights and new ideas that improve its responsiveness to the changing needs of learners.

IV. UASK APPLICATION

The proposed ScrumBan Ubiquitous Inquiry Framework (SBUIF) was implemented through the uASK application, which incorporates both Scrum and Kanban methodologies for teaching using agile as a learning pedagogy. Thus, uASK is a scrumban based collaborative learning application that provides a platform for students and teachers to facilitate their learning. The components of uASK application are mentioned as below:

A. Pre-Scrum

In pre-scrum, the teacher will be creating lesson activities and tasks along with building groups as shown in Fig. 2. The activities can be related to science education. Thus, encompassing the pre-scrum phases of build awareness and defining objectives.

B. Scrum

The next phases are categorized as scrum where the students are actively engaged in learning by interacting with each other and the environment. Along with the teacher monitoring the progress through kanban boards. In scrum, inquiry phases of engage, explore, observe, explain & reflect are included. Through these inquiry phases, students can interact with observed phenomena and explore the underlying problems. Fig. 3 shows different screens for the task of reproduction of plants through which students.

C. Post-Scrum

In the post-scrum phase, the results of the activities performed are shown and shared with group and teacher as shown in Fig. 4. The teacher will reflect upon these results and give feedback as required. These results will further be shared with the community for further exploration.



Fig. 4. uASK post-scrum activities.

V. RESEARCH METHODOLOGY

In this research, an application uASK has been developed. The uASK represents a novel application that integrates both Scrum and Kanban methodologies, offering a hybrid approach. To evaluate its effectiveness, an experiment was conducted. For this purpose, uASK was compared with one of the existing project management applications utilized in educational settings: Trello. Trello is a more established application that primarily adheres to Kanban principles. This study aimed to compare these applications in terms of their support to inquiry-based learning methods within an educational context.

A. Participants

The research involved a sample size of 127 seventh-grade students (76 boys and 51 girls) of a local school. The students

were divided into two experimental groups as a quasi-experimental research design was employed. One group was assigned to use Trello, while the other group utilized the designed application, uASK. This experimental design allowed researchers to conduct a side-by-side evaluation of how each application performed in enhancing students' learning experiences through inquiry-based approaches.

B. Experimental Design

The primary objective of this experiment was to assess whether the newly developed uASK application, which implements an innovative framework called the ScrumBan Ubiquitous Inquiry Framework (SBUIF), could demonstrate better performance in comparison to Trello, a platform which have been used in some research [28] in to support agile learning approaches. The SBUIF framework integrated into uASK represents a novel attempt to combine elements of Scrum and Kanban methodologies, tailored specifically for educational contexts.

The experiment design demonstrates a clear contrast between traditional agile-based learning methods and the enhanced SBUIF approach. By incorporating technology-driven support and guidance, the uASK application aims to provide a more structured and supportive learning environment. This comparison allows for an evaluation of the effectiveness of the SBUIF in promoting student engagement, knowledge acquisition, and collaborative learning in a K12 setting.

C. Procedure

The scrum lesson, which served as the primary activity for the experiment, was structured into seven distinct tasks, thus representing a sprint in the agile methodology.

The first task focused on imparting initial lesson knowledge. For experimental group 1, the teacher provided a traditional explanation of the topic. In contrast, experimental group 2 engaged with the subject matter through video content, allowing students to gain knowledge independently as shown in Fig. 5(a).

The second task involved a tour of a botanical garden to collect samples. Experimental group 1 adopted a random sampling approach, while experimental group 2 utilized beacon tags to guide students to specific areas and samples of interest. This technology-enhanced approach aimed to provide a more structured and targeted learning experience as shown in Fig. 5(b).



Fig. 5. Students engaged in learning activities.

For the third task, students were required to dissect samples to identify plant parts. Experimental group 1 relied on group discussions to facilitate this process. Experimental group 2, however, combined group discussions with additional support provided by the uASK application, which offered hints to aid the learning process.

The fourth task involved completing a worksheet provided through the respective applications. Experimental group 1 continued to use group discussions as their primary method for completing the worksheet. Experimental group 2, on the other hand, benefited from both group discussions and application-provided hints, offering a more scaffolded approach to worksheet completion. The fifth task was to form hypothesis on different samples provided by applications.

For the sixth task students are again building a knowledgebase through a video for Experimental group 1 and traditional lecture Experimental group 2 to learn about pollination i.e. wind and insect pollination and the last task is to form hypothesis about different flowers if they are wind or insect pollinated based on images provided by applications.

In the next section further analysis of the results from this experiment is given to draw conclusions about the relative effectiveness of the two approaches. Factors such as student performance, engagement levels, and the learning achieved would need to be assessed to determine the potential benefits of the SBUIF implemented through the uASK application compared to the more traditional Kanban-based approach using Trello.

D. Evaluation Method

Computer-Supported Collaborative Learning (CSCL) has emerged as an influential pedagogical approach that leverages digital tools to enhance group-based learning. According to Jeong and Hmelo-Silver (2016), CSCL offers seven key affordances that can significantly enrich the collaborative learning experience: (i) establishing a joint task, (ii) facilitating communication, (iii) sharing resources, (iv) engaging in productive processes, (v) co-constructing knowledge, (vi) monitoring and regulating learning, and (vii) building groups and communities. These affordances create a structured framework that promotes interaction, collaboration, and shared learning objectives, which are critical components in a group-based digital learning [32]. These affordances have been used in multiple studies as validation measures [33][34]. These have been incorporated with M3 evaluation [35][36][37] to effectively evaluate at different organizational levels.

Mapping of micro and meso evaluation to CSCL affordances is shown in Table II. For all affordances C1 to C7, at micro level, evaluation survey was used for application usability aspects while at meso levels, C4, C5 and C6 affordances are explored through pre-test and post-test scores of students. Further, at this level, C1, C2, C3 and C7 affordances are used to evaluate time logs and observations on users' activities through uASK application.

Thus, integrating CSCL affordances within the M3 Evaluation Framework enabled a multi-level analysis of the learning intervention. This structured approach provided insights into individual learner experiences and the operational environment of the learning system. By addressing these distinct levels, the evaluation framework facilitates a comprehensive understanding of the CSCL system's impact on learning outcomes, offering an evidence-based assessment of the effectiveness of collaborative learning interventions.

TABLE II. EVALUATION METHODS

CSCL affordances	M3 Evaluation Framework	
	Micro Level Method	Meso Level Methods
C1- Establishing a joint task	Application Evaluation survey	Time log & Observation through application e.g. Group formation & Task formation
C2- Facilitating communication		Time log & Observation through application i.e. chat forum
C3- Sharing resources		Time log & Observation through application e.g. videos
C4- Engaging in productive processes		Pre/post-test & Activity evaluation
C5- Co-constructing knowledge		Pre/post-test & Activity evaluation
C6- Monitoring and regulating learning		Pre/post-test & Activity evaluation
C7- Building groups and communities		Time log & Observation through application i.e. chat forum

VI. RESULTS AND ANALYSIS

The study compared the use of Trello with uASK in facilitating collaborative learning during a botany lesson. The study involved 127 grade 7 students, split into two experimental groups (63 for Trello and 64 for uASK), using a structured approach to tasks that tested different collaborative affordances.

A. Micro Level Evaluation

For micro level evaluation, application evaluation survey was used as method for evaluating CSCL affordances in both applications. Table III shows the questions of application evaluation survey with the corresponding affordance. While Fig. 6 shows the Application Evaluation survey results for both Trello and uASK. As micro level focuses on individual learner experience therefore the questions reflect users' experiences with all CSCL affordances (C1 to C7). uASK consistently scored higher across all metrics, especially in "helping with learning" and "ease of sharing information." collaboration and learning enjoyment highlight its success in Finding and Building Groups and Communities and Engaging in Productive Processes. These findings suggest that the application's structure not only enhanced individual engagement but also facilitated group cohesion, which is essential in a collaborative learning context.

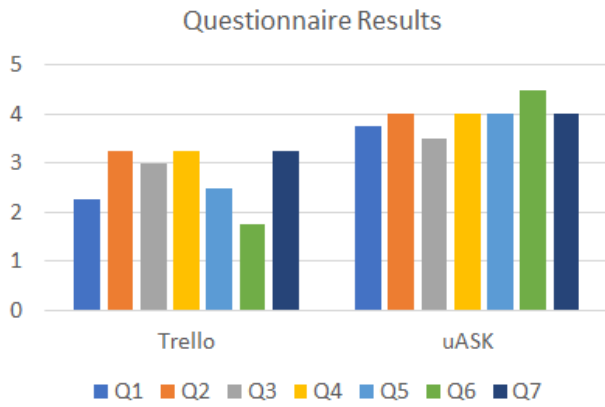


Fig. 6. Analysis of application evaluation survey results.

TABLE III. QUESTIONS ASKED IN QUESTIONNAIRE

Question No	Question asked in survey	CSCL affordances
Q1	The task was well-defined and clear	C1
Q2	There were no challenges in establishing a common understanding of tasks due to application	C1
Q3	The application helped in learning	C5
Q4	The learning experience was enjoyable	C4
Q5	Navigation through the application was easy	C6
Q6	The application helped working in group	C7 & C2
Q7	It was easy to share information and files with our group	C3

B. Meso Level Evaluation

In the meso level evaluation, number of different methods are used. Learning performances of the students during pre and post-tests are evaluated through their scores. Further, students’ engagements are examined through time spent on tasks performed during different activities.

Fig. 7 and Fig. 8 depict Pre Test, Post Test and Activity performance evaluation results that evaluate CSCL affordances C4, C5 & C6. In Fig. 7, the improvement in test scores from pre-test to post-test for both experimental groups. For the Trello group, marks improved from 27.5% to 60.5%, while the uASK group improved from 24.5% to 70.5%. This demonstrates that both groups experienced significant gains in knowledge; however, the uASK group had a more substantial increase.

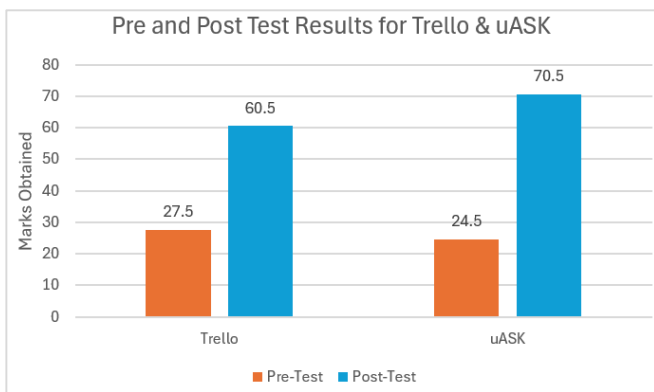


Fig. 7. Results of Pre and post test results for Trello and uASK.

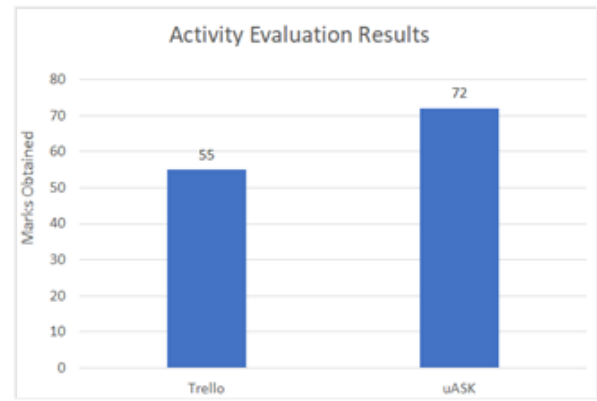


Fig. 8. Activity performance evaluation results.

The marked improvement in the uASK group’s post-test results may be attributed to the integration of both Scrum and Kanban methodologies, which potentially enhanced students’ ability to Engage in Productive Processes and Engage in Co-construction of Learning through a structured yet flexible approach to learning tasks. The provision of guidance via beacon tags and application support might have fostered a deeper understanding, as indicated by the larger knowledge gains in the uASK group. The findings align with Jeong and Hmelo-Silver’s affordances of collaborative learning, emphasizing the effectiveness of uASK’s structured support for Engaging in Productive Processes and Hypothesis Generation (Engaging in Co-construction of Learning).

The results of evaluation of the work done on the worksheet provided through the application are given in Fig. 8. uASK results show marked improvement in the performance of students as it aids in comparison to Trello group.

Then for CSCL affordances C1, C2, C3 & C7 this study used time log observation. Fig. 9 breaks down the time spent on each task by both groups, showing variations in the average time required to complete each activity. The Trello group generally spent more time on each task compared to the uASK group, with especially high time usage in Task 1 & 6 (Knowledgebase creation), Task 2 (botanical garden sample collection), Task 4 (filling the worksheet) & hypothesis generation.

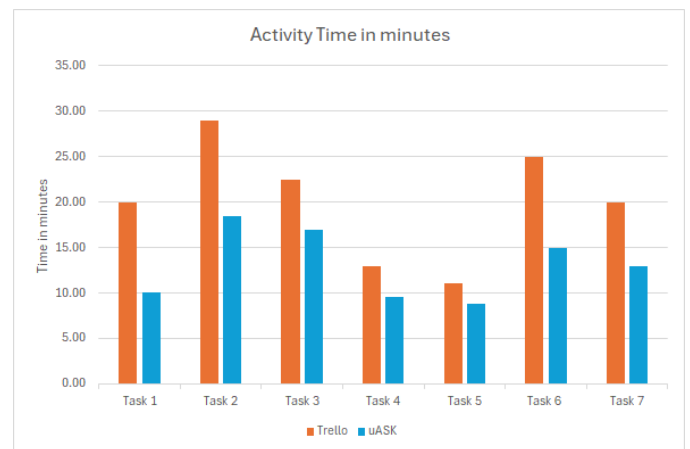


Fig. 9. Time analysis of individual activity in sprint.

The longer time spent by the Trello group suggests that the Kanban-only approach may not provide as efficient task guidance as the combined Scrum-Kanban structure in uASK, which offered hints and beacon-based support. In Task 2, where beacon tags directed students in the uASK group to specific sample areas, the decreased time suggests enhanced Monitoring and Regulation of Learning. This also suggests that uASK's guidance features facilitated quicker and more targeted completion of tasks, reflecting a higher efficiency in Establishing a Joint Task. Tasks requiring in-depth group discussion, such as Task 4, 5 and 7, saw lower completion times with uASK, possibly due to the aid of application prompts which reduced cognitive load and encouraged faster hypothesis generation.

Fig. 10 gives the total time comparison by each group to complete all tasks i.e. entire activity. The Trello group takes more time, while the uASK group completed the tasks in considerably less time. This total time difference reinforces the advantage of the uASK system in streamlining task completion through the ScrumBan inquiry framework. The 34.70%-time efficiency in uASK. The reduced time highlights enhanced Communication and Sharing Resources, as students had quicker access to task instructions and peer input, thereby fostering more effective collaborative engagement. This shorter task duration for uASK demonstrates how the affordance of Monitoring and Regulation of Learning can be operationalized to optimize task efficiency.

Overall, the combination of questionnaire feedback, time logs, and pre/post-test results illustrates how the uASK platform, leveraging the Scrum-Kanban inquiry framework, better supported the collaborative learning affordances identified by Jeong and Hmelo-Silver. The findings emphasize the impact of integrating structured yet flexible methodologies to enhance both the efficiency and depth of collaborative learning among K12 students.

Thus, it is found that uASK showed a larger improvement in post-test scores compared to Trello, indicating that it was more effective in helping students understand the content. The additional features provided by uASK, such as structured hints and guided inquiry, may have contributed to deeper learning.

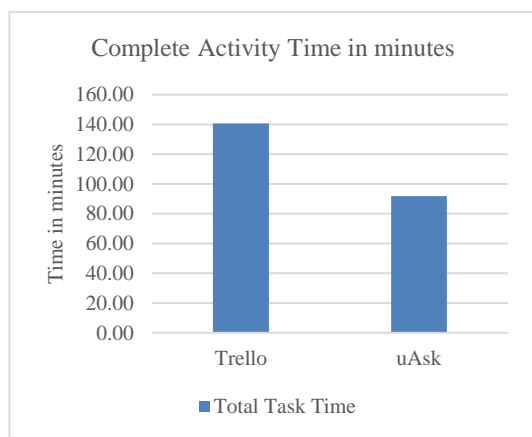


Fig. 10. Time analysis of entire sprint.

Students in the uASK group completed tasks more quickly on average. The combination of Scrum and Kanban principles allowed uASK to provide more structured guidance, reducing time spent on each activity. For instance, beacon tags helped students navigate the botanical garden and locate specific samples more efficiently.

Whereas Trello lacks the structured guidance found in uASK, making it harder for students to stay on track or find specific information without teacher intervention. In the study, this was evident in the longer time taken by the Trello group to complete each task. This lack of support might make it challenging for younger or less-experienced students to engage fully with complex learning tasks.

Additionally, uASK's features, such as hints and prompts, guided students through discussions and worksheet completion. This structured support facilitated Engaging in Productive Processes and Monitoring and Regulation of Learning, as students could focus on task goals without getting stuck or needing excessive teacher intervention. But Trello's Kanban approach is good for organizing tasks but may not provide enough scaffolding for inquiry-based learning, as it does not support features like hypothesis generation, structured hints, or guided navigation. This limitation could result in superficial understanding if students struggle to access deeper content on their own.

Furthermore, Survey results showed higher levels of engagement, satisfaction, and enjoyment in learning for students using uASK. The platform's design likely created a more enjoyable learning experience, as students had access to helpful resources and did not have to rely solely on peer discussion or teacher explanations. This positive feedback suggests that uASK's approach was more user-friendly and better suited to the students' needs. Along with uASK facilitated easier sharing of information and group work, as the platform provided collaborative tools that were more responsive to real-time group needs. The structured prompts likely encouraged smoother communication and better group cohesion. However, Trello's design does not inherently support features for monitoring student progress or self-regulation in the way uASK does. Without built-in prompts or navigation aids, students may find it harder to monitor their learning path or regulate their pace effectively, leading to uneven progress within groups.

VII. CONCLUSION AND FUTURE WORK

In conclusion, this research demonstrates that the uASK application, based on the ScrumBan Ubiquitous Inquiry Framework (SBUIF), offers significant advantages over Trello in fostering collaborative learning among students. The SBUIF platform enabled more substantial improvements in test scores, with the group using SBUIF achieving an increase from 24.5% to 70.5%, compared to Trello's 27.5% to 60.5%. SBUIF's structured support, which included beacon tags, hints, and prompts, improved task efficiency, enabling students to complete activities 34.70% faster than the Trello group. This guidance helped streamline complex tasks, reduced the need for teacher intervention, and supported key collaborative learning processes, such as productive engagement and self-monitoring.

Additionally, survey results revealed higher engagement, satisfaction, and enjoyment among SBUIF users, indicating that its design better aligns with student needs for clear task instructions and collaborative tools. While Trello's simplicity may foster more independent problem-solving, it lacks the structured guidance that younger students or those new to inquiry-based learning may require, making it less effective for complex educational tasks.

However, there are certain limitations to the current study such as it does not explore the effects of SBUIF at macro level for instance how scalable the approach might be for institutions or how it could be integrated into policy. Also, there is a need to conduct longitudinal investigation for multiple sprints for deeper understanding of how the framework performs over extended learning periods.

To conclude, addressing current limitations while integrating AI-driven technologies for adaptive feedback and advanced data analytics will pave the way for a more comprehensive, personalized, and effective learning experience, ultimately enhancing student engagement, interaction, and learning outcomes in the future.

REFERENCES

- [1] Loes, C. N. (2022). The Effect of Collaborative Learning on Academic Motivation. In *Teaching & Learning Inquiry The ISSOTL Journal* (Vol. 10). University of Calgary. <https://doi.org/10.20343/teachlearninqu.10.4>
- [2] Kerimbayev, N., Umirzakova, Z., Shadiev, R. et al. A student-centered approach using modern technologies in distance learning: a systematic review of the literature. *Smart Learn. Environ.* 10, 61 (2023). <https://doi.org/10.1186/s40561-023-00280-8>
- [3] Hava, K., & Koyunlu Ünü, Z. (2021). Investigation of the relationship between middle school students' computational thinking skills and their STEM career interest and attitudes toward inquiry. *Journal of Science Education and Technology*, 30(4), 484-495.
- [4] Liu, C., Zowghi, D., Kearney, M., & Bano, M. (2021). Inquiry-based mobile learning in secondary school science education: A systematic review. *Journal of Computer Assisted Learning*, 37, 1-23.
- [5] Ishaq, K., Azan, N., Rosdi, F., Abid, A., & Ali, Q. (2020). Usefulness of mobile assisted language learning in primary education. *International Journal of Advanced Computer Science and Applications*, 11(1), 383-395.
- [6] Vallejo-Correa, P. Monsalve-Pulido, J., & Tabares-Betancur, M. (2021). A systematic mapping review of context-aware analysis and its approach to mobile learning and ubiquitous learning processes. *Computer Science Review*, 39, 100355, 1-18.
- [7] Wadatan, R., Sovajassatakul, T., & Sriwisathiyakun, K. (2024). Effects of team-based Ubiquitous learning model on students' achievement and creative problem-solving abilities. *Cogent Education*, 11,1, 1-15.
- [8] Adewale, O. S., Agbonifo, O C., Ibam, E.O., Makinde, A.I., Boyinbode, O.K., Ojokoh, B.A., Olabode, O., Omirin, M.S., & Olatunji, S.O. (2024). Design a personalized adaptive ubiquitous learning system. *Interactive Learning Environments*, 32, 1, 208-228.
- [9] Ahmed, S. Javaid, S., Niazi, M. F., Alam, A., Ahmad, A., Baig, M.A., Khan, H. K., & Ahmed, T. (2019). A qualitative analysis of context-aware ubiquitous learning environments using Bluetooth beacons. *Technology, Pedagogy, Education*, 28,1, 53-71.
- [10] Hinostraza, J. E., Armstrong-Gallegos, S., & Villafaena, M. (2024). Roles of digital technologies in the implementation of inquiry-based learning (IBL): A systematic literature review. *Social Sciences & Humanities Open*, 9, 100874.
- [11] Lu, K. Pang, F. & Shadiev, R. (2022). Understanding the mediating effect of learning approach between learning factors and high order thinking skills in collaborative inquiry based learning. *Educational Technology Research and Development*, 69, 2475-2492.
- [12] Sepp, S., Wong, M., Hoogerheide, V., & Castro-Alonso, J. C. (2022). Shifting online: 12 tips for online teaching derived from contemporary educational psychology research. *Journal of Computer Assisted Learning*, 38(5), 1304-1320.
- [13] Katanosaka, T., Khan, M., & Sakamura, K. (2024). PhyGame: An Interactive and Gamified Learning Support System for Secondary Physics Education. *International Journal of Advanced Computer Science & Applications*, 15(6).
- [14] Noh, S. N. A., Zin, N. A. M., & Mohamed, H. (2020). Serious games requirements for higher-order thinking skills in science education. *International Journal of Advanced Computer Science and Applications*, 11(6).
- [15] Green, N. C., Edwards, H., Wolodko, B., Stewart, C., Brooks, M., & Littledyke, R. (2010). Reconceptualising higher education pedagogy in online learning. *Distance Education*, 31(3), 257-273.
- [16] Chan, J. W., & Pow, J. W. (2020). The role of social annotation in facilitating collaborative inquiry-based learning. *Computers & Education*, 147, 103787..
- [17] Battou, A. (2017). Designing an adaptive learning system based on a balanced combination of agile learner design and learner centered approach. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, 37(1), 178-186. (Matthies, 2018)
- [18] Mishra, N., & Aithal, P. S. (2023). Effect of Extracurricular and Co-Curricular Activities on Students' Development in Higher Education. *International Journal of Management, Technology, and Social Sciences (IJMITS)*, 8(3), 83-88.
- [19] Alaidaros, H., Omar, M., & Romli, R. (2021). The state of the art of agile kanban method: challenges and opportunities. *Independent Journal of Management & Production*, 12(8), 2535-2550.
- [20] Salykova, L., Sembinova, M., & Ibadildin, N. (2024). Application Of Project Management Methodologies In Modern Higher Education Institutions: A Systematic Review. *'Gosudarstvennyi Audit (State Audit)'*, 62(1), 21-31.
- [21] Fatima, A., Shaheen, A., Ahmed, S., Fazal, B., Ahmad, F., Liew, T. W., & Ahmed, Z. (2024). Exploring the use of gamification in human-centered agile-based requirements engineering. *Frontiers in Computer Science*, 6, 1442081.
- [22] Thongkoo, K., Daungcharone, K., Thanyaphongphat, J., & Panjaburee, P. (2023, November). Ubiquitous Learning Management Using Collaborative Inquiry-based Approach for Programming Course: A Case Study of 3 Universities. In *Proceedings of the 2023 7th International Conference on Education and E-Learning* (pp. 116-123).
- [23] Adhami, N., & Taghizadeh, M. (2024). Integrating inquiry-based learning and computer supported collaborative learning into flipped classroom: Effects on academic writing performance and perceptions of students of railway engineering. *Computer Assisted Language Learning*, 37(3), 521-557.
- [24] Naik, N., & Jenkins, P. (2019, August). Relax, it's a game: Utilising gamification in learning agile scrum software development. In *2019 IEEE Conference on Games (CoG)* (pp. 1-4). IEEE.
- [25] Vogelzang, J., Admiraal, W. F., & van Driel, J. H. (2019). Scrum Methodology as an Effective Scaffold to Promote Students' Learning and Motivation in Context-Based Secondary Chemistry Education. *Eurasia journal of mathematics, science and technology education*, 12(12).
- [26] Yildiz Durak, H., & Atman Uslu, N. (2023). Group regulation guidance through agile learning strategies: empowering co-regulation, transactive memory, group cohesion, atmosphere, and participation. *Educational technology research and development*, 71(4), 1653-1685.
- [27] Milićević, J. M., Filipović, F., Jezdović, I., Naumović, T., & Radenković, M. (2019). Scrum agile framework in e-business project management: an approach to teaching scrum. *European Project Management Journal*, 9(1), 52-60.
- [28] Parsons, D., Thorn, R., Inkila, M., & MacCallum, K. (2018, December). Using Trello to support agile and lean learning with Scrum and Kanban in teacher professional development. In *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)* (pp. 720-724). IEEE.

- [29] Aragonés-Jericó, C., & Canales-Ronda, P. (2022). Agile learning in marketing: Scrum in higher education. *Journal of Management and Business Education*, 5(4), 345-360.
- [30] Song, Y., & Wen, Y. (2018). Integrating various apps on BYOD (Bring Your Own Device) into seamless inquiry-based learning to enhance primary students' science learning. *Journal of Science Education and Technology*, 27, 165-176.
- [31] Parsons, D., MacCallum, K., & Sparks, H. (2021). Integrating Scrum With Other Design Approaches to Support Student Innovation Projects. In *Agile Scrum Implementation and Its Long-Term Impact on Organizations* (pp. 190-208). IGI Global.
- [32] Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help?. *Educational Psychologist*, 51(2), 247-2
- [33] McKeown, J., Hmelo-Silver, C. E., Jeong, H., Hartley, K., Faulkner, R., & Emmanuel, N. (2017). A meta-synthesis of CSCL literature in STEM education. Philadelphia, PA: International Society of the Learning Sciences.
- [34] Wang, C., Wang, J., Shi, Z., & Wu, F. (2021). Comparison of the effects of 1:1 and 1:m CSCL environments with virtual manipulatives for scientific inquiry-based learning: a counterbalanced quasi-experimental study. *Interactive Learning Environments*, 31(6), 3982-3999. <https://doi.org/10.1080/10494820.2021.1948431>.
- [35] Fabian, M. K. (2018). *Maths and Mobile Technologies: Effects on Students' Attitudes, Engagement and Achievement* (Doctoral dissertation, University of Dundee).
- [36] Vavoula, G. & Sharples, M. (2009). Meeting the Challenges in Evaluating Mobile Learning: A 3-Level Evaluation Framework. *International Journal of Mobile and Blended Learning (IJMBL)*, 1(2), 54-75. <https://doi.org/10.4018/jmbll.2009040104>.
- [37] Ahmed, S., Javaid, S., Niazi, M. F., Alam, A., Ahmad, A., Baig, M. A., ... & Ahmed, T. (2019). A qualitative analysis of context-aware ubiquitous learning environments using Bluetooth beacons. *Technology, Pedagogy and Education*, 28(1), 53-71.

M-COVIDLex: The Construction of a Domain-Specific Mixed Code Sentiment Lexicon

Siti Noor Allia Noor Ariffin, Sabrina Tiun, Nazlia Omar

Center for Artificial Intelligence Technology-Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia
Bangi, Selangor, Malaysia

Abstract—Sentiment lexicons serve as essential components in lexicon-based sentiment analysis models. Research on sentiment analysis based on the Malay lexicon indicates that most existing sentiment lexicons for this language are developed from official text corpora, general domain social media text corpora, or domain-specific social media text corpora. Nonetheless, none of the current sentiment lexicons adequately complement the corpus utilized in this study. The rationale is that words in established sentiment lexicons may convey different sentiments compared to those in this paper’s corpus, as the strength and sentiment of words are context-dependent, influenced by varying terminology or jargon across domains, and words may not share the same sentiment across multiple domains. This paper proposes the construction of a domain-specific mixed-code sentiment lexicon, termed M-COVIDLex, through the integration of corpus-based and dictionary-based techniques, utilizing seven Malay part-of-speech tags, and enhancing Malay part-of-speech tagging for social media text by introducing a new tag: FOR-POS. The constructed M-COVIDLex is evaluated using two distinct domains of social media text corpus: the specific domain and the general domain. The performance indicates that M-COVIDLex is more appropriate as a sentiment lexicon for analyzing sentiment in a domain-specific social media text corpus, providing valuable insights to governments in assessing the sentiment level regarding the analyzed topic.

Keywords—Malay social media text; mixed-code sentiment lexicon; sentiment analysis; domain-specific; lexicon-based; informal Malay; Malay part-of-speech; public health emergencies; COVID-19 Malaysia

I. INTRODUCTION

Sentiment analysis is a critical domain within natural language processing [1-4], as sentiment constitutes a significant aspect of human communication [5,6], often articulated through ambiguous and creative language [1]. For example, text that incorporates slang is comprehensible only to particular groups [7], presenting significant challenges for analysis [1]. Analyzing sentiment is essential for understanding individuals’ beliefs regarding various issues and their decision-making processes [8-11].

Rising demand for a model that can analyze the sentiment of mixed code (also known as multilingual or cross-language language) text particularly text from social media arisen recently [12,13]. It occurred on account of (i) limited sentiment resources needed for sentiment analysis for languages other than English [14-18], such as sentiment lexicon, language tools, corpora, sentiment analysis algorithm, and sentiment classification algorithm, and (ii) the escalation of social media posts exercising

mixed code in multilingual low-resource societies [19-23]. Furthermore, mixed code sentiment analysis eases a better understanding of the sentiment enunciated in various languages [12]. Up to the present time, Malay is still considered as a low-resource language, specifically in this field [5,24] since it is less studied and resource-scarce [25,26].

Sentiment lexicons are a crucial tool in lexicon-based sentiment analysis model [27-31], for any language [32]. Lexicon-based approaches demand human involvement to build a dictionary that has positive and negative lexicons [33]. This approach is divided into dictionary-based and corpus-based [20]. The former relies on sentiment words existing in digital linguistic dictionaries such as WordNet [34] and the latter relies on sentiment words that exist in the study corpus [35]. The success of lexicon-based sentiment analysis depends entirely on a sentiment lexicon that can be constructed manually or semi-automatically [36]. Furthermore, sentiment lexicons, which are also known as lexical dictionaries, are linguistic resources that have lexicons of opinions [37,38] that are labelled as positive or negative according to their semantic orientation [37-40]. For instance, words such as *baik* (good), *bagus* (excellent), and *hebat* (great) are often labelled as words with positive sentiment, while words such as *kejam* (cruel), *jahat* (evil), dan *hodoh* (ugly) are usually labelled as word with negative sentiment. The main purpose of sentiment lexicons is to assign each word in a text a corresponding sentimental weight [41]. Sentiment lexicons are divided into two types: general domain (or all-purpose) sentiment lexicon and domain-specific sentiment lexicon. A general domain sentiment lexicon is a list of words that carry either positive or negative meanings in casual conversation [42], while a domain-specific sentiment lexicon is a list of words that carry either positive or negative meanings in a discussion about a specific domain [43,44].

Recent literature review discloses that most existing sentiment lexicons developed for the field of Malay sentiment analysis are produced using either official text corpora [45], general domain social media text corpora [46], or domain-specific social media text corpora, such as affordable housing projects [47], crisis management [48], and telecommunications [49-51]. It is possible that words in the existing sentiment lexicon do not have any sentiment or carry different sentiments than words in this paper’s corpus [52,53] considering the strength and sentiment of words depend on the context of their use and the terminology or jargon differs between domains [36]. Furthermore, it is impossible for a word to have a single score or sentiment in several domains [53-58]. For example, in Malay, the term *kacau* in the food domain carries a positive sentiment

with a score of +1, if the context of its use is to stir or mix something, such as food and the term can also carry a negative sentiment with a score of -1 in the public health emergency domain, if the context of its use is to cause chaos or anxiety. Therefore, the sentiment lexicon built from this paper's corpus is more practical since the words in the sentiment lexicon reflect the sentiment in the context of the domain being studied [55].

Moreover, sentiment lexicons can be generated using two seed words expansion methods: manual and automatic based on words in a dictionary or corpus [56]. Recent literature review reveals that these seeds words can be produced by employing part-of-speech (POS) tags extraction [59], such as adjectives [60-62], verbs [50,51], adverbs [46], and nouns [63]. POS tagging is a feature in the sentiment analysis model that groups words based on their category or role in a sentence [64]. Originally in Malay, words can be classified using four POS tags: nouns, verbs, adjectives, and task words [65]. These POS tags are known to be more suitable for tagging standard Malay text than Malay social media text [64] due to the difference in the writing structure in both texts. However, the latest enhancements on Malay POS tags by [64] have made tagging social media text much effortless, since they created new POS tags especially for mixed code word. For example, FOR-NEG tag for tagging foreign words with negative meaning and NEG tag for tagging Malay words with negative meaning. Though the newly developed Malay POS tags by [64] can tag negative meanings, they lack a tag for tagging foreign words with positive meanings. Therefore, this paper will further improve the Malay POS tags by adding a new POS tag, namely FOR-POS, for tagging foreign words with positive meanings. This enhancement is made to aid in speeding sentiment analysis process by instantly distinguishing words that carry sentiment in the corpus.

This paper presented a new polyglot sentiment lexicon that specific to the latest public health emergency issue in Malaysia which is Coronavirus 2019 (COVID-19). This sentiment lexicon is known as M-COVIDLex (Malay Coronavirus Lexicon) and consist of lexicon of two main languages in Malaysia: Malay and English. The construction of M-COVIDLex utilized a combination of corpus-based and dictionary-based techniques and seven Malay POS tags: adjectives (KA), verbs (KK), adverbs (KAD), nouns (KN), FOR-NEG, FOR-POS, and NEG. The contributions of this paper are listed as follows:

- A new domain-specific social media text corpus focusing on the topic of "the impact of the implementation of government efforts to address public health emergencies in the daily routines of Malaysians". Until the data's copyright is enforced, this corpus will be inaccessible to the public or future research.
- Enriching existing Malay Normalizer by [66] through adding four new rule elements to its existent database.
- Normalizing the generated corpus using an improved Malay Normalizer.
- Enhancing existing Malay POS tags by [64] through adding one new POS tag for tagging foreign words with positive meanings: FOR-POS.

- Tagging all words in the generated corpus with the recently enhanced Malay POS tags.
- Annotated each post in the generated corpus with its proper sentiment polarity either positive, negative, or neutral.
- M-COVIDLex: a domain-specific mixed code sentiment lexicon where each lexicon has been classified as either positive or negative sentiment word.
- The performance evaluation proves that the sentiment lexicon built from this paper's corpus is more practical in analyzing sentiment from the same domain corpus than the general domain corpus.

This paper is structured as follows. The following section (Section II) reviews the related work. In Section III, the proposed method is introduced in detail. Section IV presents experiment and obtains results. Section V discussed the experiments and obtained results and finally, Section VI concludes how this paper can be expanded to further contribute to the Malay social media text sentiment analysis fields.

II. RELATED WORKS

A. Types of Sentiment Lexicon

Sentiment lexicons are divided into two types: general domain and domain specific. General domain sentiment lexicon is a list of words that carry either positive or negative meanings in casual conversation [42], while domain-specific sentiment lexicon is a list of words that carry either positive or negative meanings in conversation about a specific domain [43,44].

A general domain sentiment lexicon is suitable for development and implementation in a model that analyzes sentiment text that is not based on any domain for the list of words in the general domain sentiment lexicon is limited to words commonly used in daily conversations and its polarity score is not sensitive to any domain [67-70]. In addition, the study by [36], [47], [71], and [72] stated that domain-specific sentiment analysis models that implement the studied domain-specific sentiment lexicon have the potential to provide better sentiment analysis results and more accurate classification results compared to the general domain sentiment lexicon. The reason for this is that the sentiment polarity scores of words in the sentiment lexicon are obtained based on their meanings in the domain studied [73,74] and sentiment analysis research is sensitive to the domain analyzed. Opinion expressions that carry sentiments, whether positive or negative, differ between domains since each domain has its own language, terminology, or jargon [53-58]. Furthermore, it is impossible for an opinion expression to have the same sentiment polarity score in all domains because words and their sentiment polarity scores are closely related depending on the domain context in which they are used [36].

In conclusion, models developed to analyze sentiment in a specific domain are encouraged to use a sentiment lexicon specific to that domain rather than a general domain sentiment lexicon. The reason for this is that the use of a correct and proper sentiment lexicon can produce good analytical accuracy.

B. Techniques for Constructing Sentiment Lexicon

Sentiment lexicons can be generated using two seed word expansion methods: manual and automatic based on words in a dictionary or corpus [56]. However, generating a comprehensive sentiment lexicon using manual expansion techniques is complicated, time-consuming, and prone to various errors [56]. Therefore, researchers prefer to use existing sentiment lexicons, such as General Inquirer [75], WordNet [34], Opinion Lexicon [76], Subjectivity Lexicon [77], SentiWordNet [78], SentiStrength [79], SenticNet [80], AFINN [81], Semantic Orientation CALculator (SO-CAL) [63], National Research Council Canada (NRC) Emotion Lexicon [82], Valence Aware Dictionary and sEntiment Reasoner (VADER) [83], LIWC [84], TextBlob [85], and Bing [86].

Dictionary-based expansion techniques require researchers to manually collect initial list of seed words and their polarities before expanding them by extracting synonyms and antonyms of each seed word from existing dictionaries [20]. Corpus-based expansion techniques also use the initial list of seed words to identify words that carry sentiment values and their polarities in the research corpus [56]. Sentiment lexicon generation using corpus-based expansion techniques is more suitable for domain-specific sentiment analysis research than dictionary-based and manual expansion techniques [9]. This is due to the technique is extremely sensitive to the domain, capable of generating sentiment lexicons specific to the domain [40], and capable of handling informal texts well, for instance social media texts [9]. However, sentiment lexicon construction using this corpus-based expansion technique is a lengthy process [87], inefficient in labelling texts with formal terms [9] and has limitations in distinguishing all opinion words compared to dictionary-based expansion techniques [9]. The reason for this is that this technique requires a large corpus to include all opinion words in the study language and large-scale corpus collection is strenuous to be prepared [9]. However, regardless of the sentiment lexicon production technique chosen by the researcher to develop or use, the overall quality of this sentiment lexicon is hard to measure [9]. A summary of sentiment lexicon construction techniques implemented by past research that analyzed Malay sentiment can be seen in Table I.

Based on Table I, the corpus-based sentiment lexicon construction technique can be implemented using the feature extraction method, where this method is done by extracting words that belong to certain POS tag such as adjectives, verbs, and adverbs. Additionally, this paper discovered that there is other POS tag that can be extracted together as they also have sentiment values, namely nouns [63]. The dictionary-based sentiment lexicon construction technique can be performed via the translation method, where the lexicon in the English dictionary is interpreted into Malay. Meanwhile, the manual sentiment lexicon production technique is executed by compounding both techniques, namely corpus-based and dictionary-based.

In conclusion, sentiment lexicons can be developed through various techniques such as corpus-based, dictionary-based, or manually. Choosing the right technique is the key to certify that the sentiment lexicon produced is of excellent quality and capable of providing good sentiment analysis.

TABLE I. SUMMARY OF SENTIMENT LEXICON CONSTRUCTION TECHNIQUES BY PREVIOUS STUDY ON MALAY SENTIMENT ANALYSIS

Techniques	Methods	Method Description	References
Corpus-based	Feature extraction	POS tag: adjective	[46,50,51,61]
		POS tag: verb	[46,50,51,61]
		POS tag: adverbs	[46]
		POS tag: negation	[46]
Dictionary-based	Translation	AFINN to Malay	[88]
Manual	Combination of corpus-based & dictionary-based techniques	Corpus: Malay Sabah lexicons Dictionary: multilingual lexicons	[89]
		Corpus: emoticon, neologism Dictionary: Malay lexicons (MySentiDic), English lexicons (MySentiDic translation)	[45]
		Corpus: feature extraction (POS tag: adjective) Dictionary: WordNet Bahasa & WordNet translation to Malay	[60]
		Corpus: feature extraction (POS tag: adjective) Dictionary: lexicons by Alexander & Omar (2017)	[62]

III. METHODOLOGY

As previously mentioned, this paper aims to construct domain-specific mixed code sentiment lexicons otherwise known as M-COVIDLex through a combination of corpus-based and dictionary-based techniques along with seven Malay POS tags. The proposed M-COVIDLex construction method entails five key phases: (i) data gathering, (ii) data preprocessing, (iii) construction of M-COVIDLex, (iv) sentiment analysis, and (v) sentiment classification. Each phase is enlightened in further detail below and it is important to emphasize that the data gathering and analysis presented in this paper adhere to the terms and conditions of social media platform, X.

A. Data Gathering

This paper gathered data manually from the social media platform, X (formerly Twitter), where the data must be composed of the combination of two languages, Malay and English, concerning “the impact of the implementation of government efforts to address public health emergencies in the daily routines of Malaysians”. This sort of data is recognized as mixed code or code-switching or multilingual. To achieve this purpose, this paper employed keyword-driven data-gathering techniques [64,66,90]. The search was performed on X’s advanced search functions [64,66,89] using a predefined list of thirty-three keywords of four affected sectors during COVID-19 in Malaysia: education, safety, health, and economy (see Table II). The keywords were obtained from data issued by [91] and [92]. Nevertheless, the keywords used are restricted to the Malay

language apart from acronyms such as SOP (Standard Operational Procedure) as well as the English loanword like “moratorium” and “internet”. X’s advanced search function permits users to refine search results based on several criteria including keywords, publication date, language type, and account type, which authorizes researchers to oversee the kind of posts suitable to be extracted. For a post to be included in the search, it must have at least one keyword and was posted between March 2020 and September 2021. The sectors and keywords were selected in such a way to boost the number of posts about the selected topic and limit any extraneous posts [90]. As a result, this paper achieved in gathering 16,898 related posts which were saved in textfiles (txt). The data gathering stage is deemed complete and adequate when all posts resulting from the keyword search have been extracted.

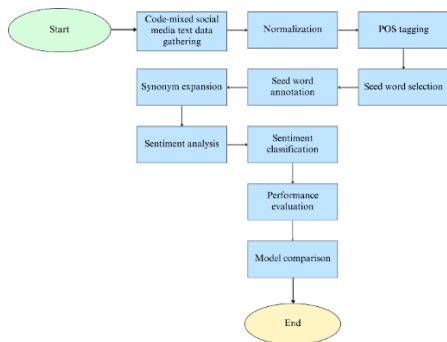


Fig. 1. M-COVIDLex flowchart.

TABLE II. DATA GATHERING KEYWORDS USED IN THIS PAPER

Keywords	Sectors	Overall Keywords	Total Post
<i>harga barang</i>	Economy	11	4,838
<i>tiket pengangkutan</i>			
<i>bantuan kerajaan</i>			
<i>pemulihan ekonomi</i>			
<i>bayaran pinjaman</i>			
<i>inisiatif kerajaan</i>			
<i>pelancongan</i>			
<i>BDR</i>			
<i>moratorium</i>			
<i>wang simpanan</i>			
<i>golongan rakyat</i>	Health	6	3,773
<i>vaksin</i>			
<i>pencegahan</i>			
<i>kontak</i>			
<i>kesihatan</i>			
<i>kuarantin</i>	Safety	5	3,131
<i>kawalan pergerakan</i>			
<i>kawalan</i>			
<i>kebenaran</i>			
<i>rentas</i>			
<i>SOP</i>	Education	11	5,156
<i>sekatan</i>			
<i>bayaran belia</i>			
<i>bayaran pendidikan</i>			
<i>pengurangan yuran</i>			
<i>bantuan makanan</i>			
<i>bantuan pelajar</i>			
<i>internet</i>			
<i>kediaman pelajar</i>			
<i>pergerakan pelajar</i>			
<i>pelajar institusi</i>			
<i>pelajar sekolah</i>			
<i>PDPR</i>			

B. Data Preprocessing

Data preprocessing plays a fundamental role in cleansing text data, specifically social media text, where typographical errors, slang terms, acronyms, and polyglot (code-mixed, code-switching, and multilingual) lexes are frequent [93]. According to [88], the existence of noise in the Malay text is not being managed at this phase [13] affirmed that various preprocessing procedures are important to alleviate noise, typographical divergences, and morphological intricacy, especially in a low-resources language, such as Malay. In this phase, the gathered data is overseen in two steps: normalization and POS tagging. Normalization is essential to diminish the presence of noise from the data that has the potential to interfere with the results of the sentiment analysis and to make the data more manageable [89]. POS tagging is needed to annotate each word in the data with a suitable POS tag.

1) *Normalization*: A recent study by [66] proposed a rule-based normalization application exclusively built to refine Malay social media text, where it achieves high accuracy in their analysis (97 percent). Therefore, in this paper, this Malay Normalizer was employed to normalize all lexicons in the gathered data as it incorporates diverse essential preprocessing procedure for Malay social media text. For instance, (i) removing noise, (ii) normalizing Malay, English, and Romanized Arabic words to their standard form, (iii) expanding abbreviations, contractions, and acronyms, and (iv) normalizing slang term, colloquialism, and dialects. Nonetheless, it is expected that the application will be inept at normalizing most of the words that exist in this paper’s gathered data seeing that it was built using a smaller corpus size. Hence, it needs to be enriched with words from this paper’s corpus. Albeit [94] exclaimed that the size of a corpus does not imply its quality and could have more noise, this enrichment is compulsory to ensure all lexicons in the corpus are normalized to their standard form. The enrichment entails adding four new rule elements to its existent database: (i) normalization of novel words, for example, slang term, dialects, and domain-specific terms, (ii) normalization of emoticons, (iii) exclusion of Malay and English question words, and (iv) elimination of English and selected Malay stop words. This supplementation effectively reduces the number of X posts from 16,898 posts to 16,600 posts, where 298 posts were removed from this paper’s corpus due to noise.

2) *POS tagging*: The normalized gathered data initially annotated using Malay POS tags by [64]. These Malay POS tags was chosen given that (i) it achieved high accuracy in their analysis (95 percent) against Malay social media text, (ii) it has been thoroughly upgraded from the previous Malay POS tags by [95], (iii) the POS tags were tailored to be able to annotate each word in the Malay social media text, for example, any foreign language exist in the corpus will be tag with FOR tag, slang term will be tag with SL tag, and dialect terms will be tag with LD tag, and (iv) standard Malay POS tags absence proper tags to label all words in this paper’s corpus since social media text are usually written using mixed code otherwise known as multilingual or code-switching [66]. Although [64] has

improved these Malay POS tags to accommodate words in Malay social media text, this paper discovered that it has yet to set up a proper tag for tagging foreign words with positive sentiments. Therefore, it needs to be improved by adding a new POS tag exclusively for tagging positive sentiment foreign words. This paper ruled out to name the newly created POS tag as FOR-POS (foreign positive). Words for the FOR-POS tag are found and extracted from the word that conveys positive meanings in the FOR-tag word list. This paper performed this improvement to aid in speeding the sentiment analysis phase by instantaneously distinguishing words that carries sentiment in the corpus.

C. Construction of M-COVIDLex

The proposed M-COVIDLex construction method entails three steps: (i) seed word selection, (ii) seed word annotation, and (iii) synonym expansion. In this phase, the proposed method will be enlightened in detail.

1) *Seed word selection*: A seed word is a lexicon that carries either positive or negative sentiment polarity. Based on the literature review in Section II, this paper discovered that these seed words can be generated using four types of POS tags: KA, KK, KAD, and KN. Therefore, in this step, the lexicon tagged with these four POS tags will be extracted as M-COVIDLex seed words. However, since this paper's corpus is made of mixed code social media text and each lexicon is tagged using the improvised Malay POS tags, this paper decided to add three more POS tags to produce M-COVIDLex seed words. The three additional POS tags are (i) NEG, a POS tag designed specifically for negative particles in the Malay language, (ii) FOR-NEG, a POS tag for foreign language words that carry negative sentiments, and (iii) FOR-POS, a POS tag for foreign language words that carry positive sentiments. The lexicons from these seven POS tags were extracted from the corpus using the POS tagging extraction technique. The technique used specific patterns to extract all related lexicons in the corpus, and its implementation produced a list of seed words for each POS tag, where the list functions as a lexical dictionary and is needed when constructing M-COVIDLex. Algorithm 1 shows how this seed word extraction was performed for the lexicon with the KA POS tag. Fig 2 presents the result of executing Algorithm 1 and Fig 3 summarizes the total number of seed words for each POS tag in the M-COVIDLex.

Algorithm 1: M-COVIDLex Seed Word Selection

Input: tagged_corpus K_G , post S , lexicon L , part_of_speech G
Output: seed_word $M\text{-COVIDLex} = (L, G)$

```
Start
  for all  $S$  in  $K_G$  do
    for all  $L$  in  $S$  do
      find  $L == G$  (KA)
      if  $L == G$  (KA)
        add  $L$  in  $M\text{-COVIDLex}$ 
      end if
    end for
  end for
End
```

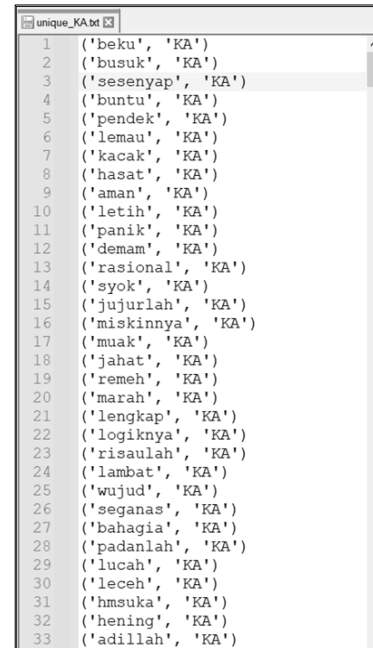


Fig. 2. Seed word selection result for lexicon with the KA POS tag.

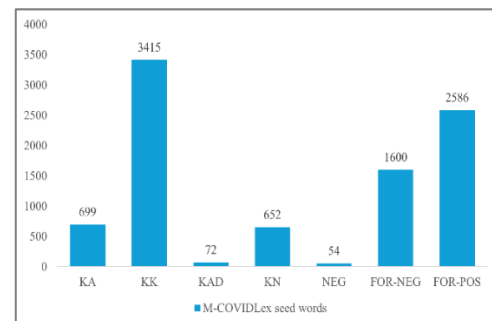


Fig. 3. Summary of M-COVIDLex.

2) *Seed word annotation*: In this step, seed words will be annotated with either positive or negative sentiment polarity. However, only seed words with KA, KK, KAD, and KN POS tags engage in this annotation process. Seed words with NEG, FOR-NEG, and FOR-POS POS tags are exempted since their sentiment polarity is clear based on the POS tag type. The annotation process in this paper was done by the hired annotators. According to [96], an annotator is an individual appointed to annotate data according to guidelines and time limits given. The following are the guidelines criteria for selecting annotator listed by [96] and followed in this paper:

- Researchers need to figure out the type of language the annotator needs to know or be fluent in before doing the annotation task.
- The researcher needs to set up the specific knowledge that the annotator needs to know about the annotation task.
- Researchers need to make practical considerations about several things, for instance, funds, time, data size, etc.

These guidelines are crucial for researchers to be able to find and appoint annotators who meet the qualifications and criteria [47]. Table III presents the criteria and qualifications of the appointed annotator in this paper along with its references.

TABLE III. THE NEWLY REVISED CRITERIA AND QUALIFICATION OF AN ANNOTATOR USED IN THIS PAPER

Criteria & Qualification of an Annotator	References
Language: Malay native speakers and fluent in English.	[47,96]
Knowledge: Have basic knowledge of the topic and field of study.	[47,96]
Practical Consideration: <ol style="list-style-type: none"> Current academic graduate (highest): <ul style="list-style-type: none"> Malaysian Certificate of Education or Malaysian Higher Certificate of Education or Matriculation or Diploma or Degree or Master Aware of the meaning and use of current social media language. Able to give full commitment to annotating data within the specified period. 	[96]

The appointed annotator was assigned to manually annotate the lexicons based on the search results on the *Pusat Rujukan Persuratan Melayu*, or PRPM for short [89]. PRPM is an online dictionary specifically for the Malay language developed by DBP. It can be reached at the following link: <https://prpm.dbp.gov.my/>. The task of the annotator is to identify and label the sentiment polarity of words based on the definition issued by PRPM and the context of its use. Fig 4 presents some of the seed words for KA POS tag in the M-COVIDLex that have been annotated with their proper sentiment polarity.

3) *Synonym expansion*: In this step, all M-COVIDLex seed words will be expanded. This expansion aims to expand the coverage of M-COVIDLex seed word variations based on the

search results in PRPM for Malay and WordNet [34] for English.

a) *Malay lexicon synonyms*: The expansion of the Malay language lexicon synonyms was conducted based on the lexicon search results in PRPM [89] and the technique was manual [97]. This paper aims to standardize the expansion of the Malay lexicon by limiting it to level one, given the variability in the number of synonyms across different Malay lexicons. Alternative approaches to this process include expanding the lexicon by accepting all possible synonyms and antonyms.

The first step in the expansion of the Malay lexicon synonyms was to conduct a lexicon search on the PRPM homepage. Fig 5 shows the search conducted on the PRPM homepage for the lexicon *lupa* and Fig 6 shows the search results for the lexicon, where there is word information consisting of its definition and thesaurus. The second step is to extract lexical synonyms up to level one only. Lexical synonyms can be obtained in the thesaurus section. Fig 7 shows the thesaurus for *lupa*, where the lexicon has synonyms up to level three: not remembering (level one), not aware (level two), and not arising in memory (level three). The third step is to add the level one synonym to the M-COVIDLex seed words. Table IV presents examples of synonym expansion for the *lupa*, *layak*, and *usaha* lexicon.

	A	B
1	Leksikon	Polariti
2	adil	Positive
3	adillah	Positive
4	adilnya	Positive
5	agam	Positive
6	aib	Negative
7	ajaib	Negative
8	aktif	Positive
9	alang	Positive
10	alpa	Negative
11	aman	Positive
12	aneh	Negative
13	angkuh	Negative
14	asing	Negative
15	asyik	Positive
16	atasi	Positive

Fig. 4. Annotation results from several seed words in the M-COVIDLex.

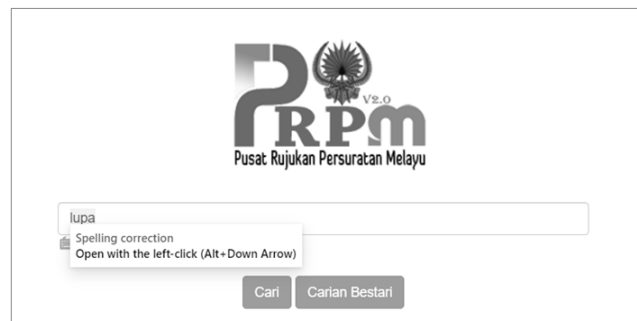


Fig. 5. Search for *lupa* on the PRPM homepage.



Fig. 6. Search results for *lupa* on the PRPM website.

TABLE IV. EXAMPLE OF SYNONYMS EXPANSION FOR MALAY LEXICON

M-COVIDLex Seed Words				
Malay Lexicon	Malay Synonym	POS Tag	Polarity	Polarity Score
<i>lupa</i>	<i>tidak ingat</i>	KK	Negative	-1
<i>layak</i>	<i>padan</i>	KA	Positive	+1
<i>usaha</i>	<i>daya upaya</i>	KN	Positive	+1

b) *English lexicon synonyms*: The expansion of the English lexicon synonyms is conducted using a lexical dictionary-based technique, namely WordNet [34]. However, this method only involves lexicons belonging to the FOR-NEG and FOR-POS POS tags. The reason for this is that only these two POS tags have the English lexicon in the M-COVIDLex seed word list. Algorithm 2 below presents how the English synonym expansion method is conducted using the WordNet application.

Algorithm 2: M-COVIDLex English Synonym Expansion

Input: seed_word *M-COVIDLex*, lexicon *L*, wordNet *W*, english_synonym *S_i*

Output: expansion *M-COVIDLex* = (*L*: *P*);

L: lexicon, *P*: lexicon_polarity

Start

```

for all L in M-COVIDLex do
    if exist L in W then
        add Si in M-COVIDLex
    end if
end for
    
```

End

D. Sentiment Analysis

Sentiment analysis is the process of deciding sentiment polarity score, where the technique that will be implemented in this paper is based on (i) grammatical rules of four Malay POS tags: KNF (negation), KP (intensifier), KB (auxiliary), and KH (conjunction), and (ii) word frequency calculations. In this paper, the order of importance of these four Malay POS tags are presumed as follows: KH > KB > KNF > KP (see Algorithm 3). This order is established by recognizing the significance of each POS tag within grammatical rules for determining sentiment polarity scores. For instance, KP POS tag is of the highest importance because of its aptitude to increase or decrease the strength of the polarity of post sentiment, hence found at the end

of the order. Fig 8 shows how the sentiment analysis phase happens, and Table V presents the criteria of sentiment polarity used in this paper.



Fig. 7. *Lupa* thesaurus on the PRPM website.

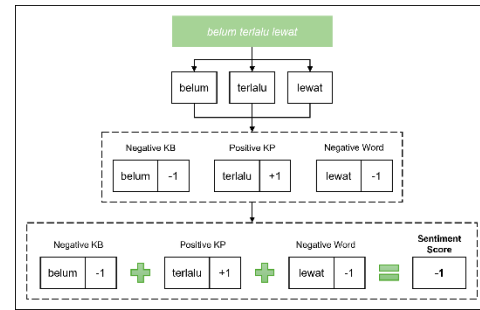


Fig. 8. Demonstration of how M-COVIDLex sentiment analysis phase analyzes the sentiment of a text.

Algorithm 3: M-COVIDLex Sentiment Analysis

Input: Post *S*, Lexicon *L*

Output: Post *S*, Lexicon *L*, Polarity *P*

Start

```

if there exists L == positive in S, then
    | score L == 1;
else exists L == negative in S, then
    | score L == -1
end if
if there exists L == conjunction in S, then
    | classify S using conjunction rules;
else if exists L == auxiliary in S, then
    | classify S using auxiliary rules;
else if exists L == negation in S, then
    | classify S using negation rules;
else if exists L == intensifier in S, then
    | classify S using intensifier rules;
else
    | classify S using word count
end if
    
```

End

TABLE V. SENTIMENT POLARITY CRITERIA USED IN THIS PAPER

Sentiment Score Criteria	Sentiment Polarity
Sentiment score value > 0	Positive
Sentiment score value = 0	Neutral
Sentiment score value < 0	Negative

E. Sentiment Classification

Sentiment classification is the process of classifying posts according to their corresponding polarity, where in this paper, the post is classified into either positive (sentiment polarity score equal to 1), negative (sentiment polarity score equal to -1), or neutral (sentiment polarity score equal to 0) based on the results

gained from sentiment analysis phase. The sentiment classification technique used in this paper is a simple classification otherwise known as lexicon-based classification [89,98]. Algorithm 4 enlightens how this sentiment classification process is executed, and Fig 9 reveals the implementation results of this phase.

Fig. 9. M-COVIDLex sentiment classification results.

Algorithm 4: M-COVIDLex Sentiment Classification

Input: Post S , Lexicon L , Positive_Word P_W , Negative_Word N_W , Conjunction_Rules R_{KH} , Auxiliary_Rules R_{KB} , Negation_Rules R_{KNF} , Intensifier_Rules R_{KP}

Output: Post S , Conjunction_Rules R_{KH} , Auxiliary_Rules R_{KB} , Negation_Rules R_{KNF} , Intensifier_Rules R_{KP} , Lexicon_Frequency K_L , DataFrame RD , Classification SC , Polarity_Score P , Sentiment $LexClass$

Start

```

count total  $P_W$  in  $S$ 
count total  $N_W$  in  $S$ 
for every  $L$  in  $S$ 
    count frequency  $L == R_{KH}$ 
    add frequency  $L$  in  $R_{KH}$  [ $RD$ ]
    print  $P$  frequency  $L$  in  $R_{KH}$  [ $RD$ ]
    count frequency  $L == R_{KB}$ 
    add frequency  $L$  in  $R_{KB}$  [ $RD$ ]
    print  $P$  frequency  $L$  in  $R_{KB}$  [ $RD$ ]
    count frequency  $L == R_{KNF}$ 
    add frequency  $L$  in  $R_{KNF}$  [ $RD$ ]
    print  $P$  frequency  $L$  in  $R_{KNF}$  [ $RD$ ]
    count frequency  $L == R_{KP}$ 
    add frequency  $L$  in  $R_{KP}$  [ $RD$ ]
    print  $P$  frequency  $L$  in  $R_{KP}$  [ $RD$ ]
end for
for every  $P$  in [ $RD$ ]
    count  $P$  in  $R_{KH}$  [ $RD$ ] +  $R_{KB}$  [ $RD$ ] +  $R_{KNF}$  [ $RD$ ] +  $R_{KP}$  [ $RD$ ]
    add  $P$  in  $SC$ 
end for
if  $P$  in  $SC == 0$ ;
    print  $P$  in  $LexClass == 0$ 
else if  $P$  in  $SC > 0$ ;
    print  $P$  in  $LexClass == 1$ 
else if  $P$  in  $SC < 0$ 
    print  $P$  in  $LexClass == -1$ 
end if

```

End

IV. EXPERIMENT AND RESULTS

The evaluation of M-COVIDLex will involve an analysis of sentiments from two distinct social media text corpora: a domain-specific corpus (the one presented in this paper) and a general domain corpus as referenced in [64].

A. M-COVIDLex

The third phase has successfully resulted in the development of M-COVIDLex, a domain-specific mixed code sentiment lexicon. The lexicon in M-COVIDLex is restricted to two sentiment polarities: positive sentiment, assigned a score of +1, and negative sentiment, assigned a score of -1. Neutral sentiment lexicons with a score of 0 are excluded from the seed words of M-COVIDLex, as they lack sentiment value. M-COVIDLex contains 6,698 lexicons associated with positive sentiment and 3,813 lexicons associated with negative sentiment.

B. Performance Evaluation

The performance of M-COVIDLex is initially assessed on a domain-specific social media text corpus, concentrating on the effects of government initiatives aimed at addressing public health emergencies on the daily routines of Malaysians. The experimental dataset corpus was derived from the execution of the second phase. This study utilized a subset of 800 posts from a total of 16,600 normalized posts. The selection of posts was conducted randomly, based on sentiment polarity, resulting in 400 posts with positive sentiment and 400 posts with negative sentiment. The chosen experimental datasets are accompanied by confusion matrix tables to derive their values. Fig 10 illustrates a portion of the data annotation results, while Table VI displays the corresponding values derived from these results.

teks	polariti	penilaian
aaron kwok years old withdraw kumpulan wang simpanan pekerja already	1	True Positive
abah bantuan sara hidup bantuan prihatin rakyat citra ilestari isinar lepas ringgit malaysia ewallet hear shit one time going lose ringgit malaysia bloody ewallets given promote cashless transactions bloody financial aid	1	False Positive
abah guru kaunseling sekolah rendah pagi abah pergi sekolah jam pagi sebentar langit masih bergelap untuk menyambut anak anak muridnya kalau sahajalah murid darjah yang menangis tak mahu sekolah abah pujuk masuk	1	True Positive
abah kau tak keluarkan duit kumpulan wang simpanan pekerja akaun bukan nak dapat okay sekali sudah okay nak harap bantuan prihatin nasional yang dapat biarlah yang perlukan duit mengeluarkan duit divaktu terdesak	1	True Positive
abah kawan yang tinggal program perumahan rakyat sudah golongan bottom forty cakap beli beras yang harganya kampung ringgit malaysia boleh makan minggu tetapi nak bayar ansuran atau sewa rumah kereta dan data internet untuk pengajaran dan pembelajaran rumah	1	False Positive

Fig. 10. M-COVIDLex data annotation results.

TABLE VI. CONFUSION MATRIX TABLE VALUES FOR M-COVIDLEX PERFORMANCE EVALUATION

Confusion Matrix	Experimental Dataset
True Positive (TP)	274 posts
False Positive (FP)	126 posts
True Negative (TN)	305 posts
False Negative (FN)	95 posts

The values presented in Table VI serve to evaluate the effectiveness and quality of M-COVIDLex in sentiment analysis

of the experimental dataset. This paper employs the following evaluation metrics to assess the performance of the proposed sentiment lexicon: (i) error rate, (ii) accuracy, (iii) sensitivity, (iv) specificity, (v) precision, and (vi) F1-score. Fig 11 illustrates the performance outcomes of M-COVIDLex, utilising the confusion matrix alongside six evaluation metrics.

Error rate. This evaluation measure was selected to assess the effectiveness of M-RuleScore in analysing the sentiment of the dataset, indicating that a lower error rate value corresponds to improved performance of the proposed M-RuleScore.

$$Error\ rate = \frac{(FP + FN)}{(TP + FP + FN + TN)} \quad (1)$$

Accuracy. This evaluation measure was selected to assess the proportion of posts accurately predicted from the entire dataset, with a higher accuracy value indicating superior performance of the proposed M-RuleScore.

$$Accuracy = \frac{(TP + TN)}{(TP + FP + FN + TN)} \quad (2)$$

Sensitivity. This evaluation measure was selected to assess the classification error of the dataset, indicating that a higher sensitivity value correlates with improved performance of the proposed M-RuleScore.

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (3)$$

Specificity. This evaluation measure was selected to assess the suitability of M-RuleScore for analysing the dataset, indicating that a higher specificity value reflects improved performance of M-RuleScore in analysing negative sentiments.

$$Specificity = \frac{TN}{(TN + FP)} \quad (4)$$

Precision. This evaluation measure was selected to assess the effectiveness and robustness of the proposed M-RuleScore in analysing the sentiment of the dataset.

$$Precision = \frac{TP}{(TP + FP)} \quad (5)$$

F1-score. This evaluation measure was selected to assess the mean values of sensitivity and accuracy for the proposed M-RuleScore in analysing the dataset's sentiment.

$$F1\ -\ score = \frac{(2 \times precision \times sensitivity)}{(precision + sensitivity)} \quad (6)$$

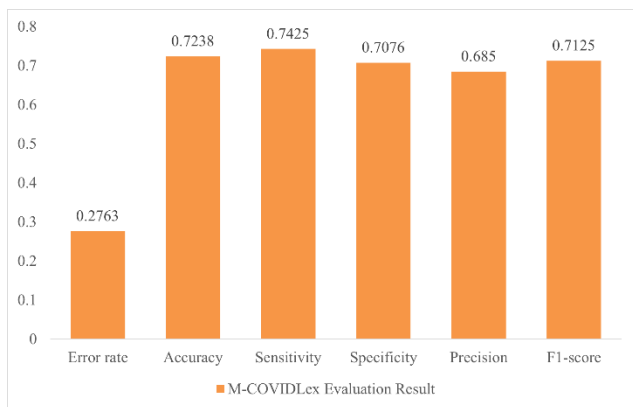


Fig. 11. M-COVIDLex performance evaluation result.

The performance evaluation results of M-COVIDLex presented in Fig 11 yield several conclusions.

- M-COVIDLex achieved an accuracy of 72.38% in the analysis of the experimental dataset. The analysis indicates that, from a total of 800 posts by Malaysians, M-COVIDLex successfully evaluated 274 posts expressing positive sentiments and 305 posts expressing negative sentiments regarding the government's response to the COVID-19 crisis.
- M-COVIDLex recorded an error rate of 27.63% in the analysis of the experimental dataset. Of the 800 posts analysed concerning the government's efforts in addressing the COVID-19 crisis and their impact on the daily lives of Malaysians, 126 posts were incorrectly classified as expressing positive sentiments, while 95 posts were misclassified as expressing negative sentiments.
- M-COVIDLex achieved a precision of 0.6850 in the analysis of the experimental dataset. The precision value indicates that, among the 400 posts predicted to express positive sentiments in the experimental dataset, only 68.50% accurately reflected positive sentiments and aligned with the government's efforts in addressing the COVID-19 crisis.
- M-COVIDLex demonstrated a sensitivity of 0.7425 in the analysis of the experimental dataset. The sensitivity value indicates that, among 369 genuinely positive posts in the experimental dataset, only 74.25% expressed positive sentiments regarding the government's efforts to address the COVID-19 crisis.
- M-COVIDLex demonstrated a specificity of 0.7076 when analysing the experimental dataset. The specificity value indicates that, among the 400 posts predicted to express negative sentiments in the experimental dataset, 70.76% accurately reflect negative sentiments opposing the government's efforts in addressing the COVID-19 crisis.

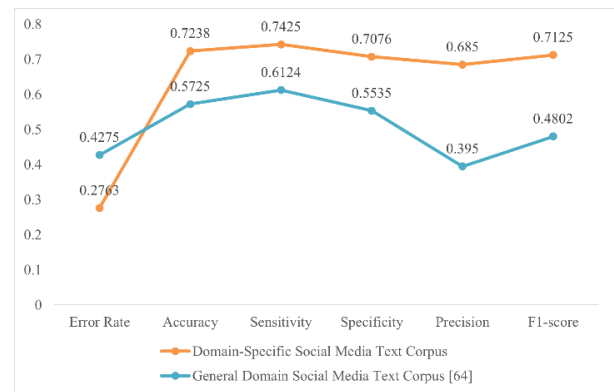


Fig. 12. Comparison of evaluation metrics between domain-specific and general domain social media text corpus.

To ensure comparability in the number of posts used for model comparison with other datasets and for evaluating the performance of M-COVIDLex, the analysis was limited to 800 posts. The performance of M-COVIDLex may be impacted by

this limitation. Evaluating performance across all 16,600 posts is likely to produce more favourable outcomes for M-COVIDLex. M-COVIDLex effectively assessed Malaysian sentiment concerning the government’s response to the COVID-19 crisis and its effects on daily life. Despite M-COVIDLex achieving an accuracy of 72.38 percent and an error rate of 27.63 percent, which are lower than many leading models in sentiment analysis for comparable domains, the performance evaluation indicates that a significant majority of Malaysians approve of and support the government’s crisis management efforts.

C. Model Comparison

This section evaluates the performance of M-COVIDLex on the general domain social media text corpus as referenced in [64]. The corpus selected by [64] was generated from Malay social media text, focussing exclusively on the contextual use of Malay social media terminology without emphasising any specific domain. The purpose of this assessment is to evaluate the suitability and effectiveness of M-COVIDLex as a sentiment lexicon for analysing sentiment within this domain. This study utilised a dataset comprising 800 posts selected from a total corpus of 1,791 posts, with 400 posts representing positive sentiment and 400 posts representing negative sentiment. The selection of these posts was conducted randomly, determined by the polarity of their sentiment. Table VII presents a comparison of the confusion matrix values for domain-specific social media text versus general domain social media text. Fig 12 presents a comparison of the evaluation metrics for the social media texts analysed.

TABLE VII. COMPARISON OF CONFUSION MATRIX VALUES BETWEEN DOMAIN-SPECIFIC AND GENERAL DOMAIN SOCIAL MEDIA TEXT CORPUS

Performance Evaluation	Domain-Specific Social Media Text Corpus	General Domain Social Media Text Corpus [64]
Corpus size	800 posts	
Positive sentiment posts	400 posts	
Negative sentiment posts	400 posts	
True Positive (TP)	274 posts	158 posts
False Positive (FP)	126 posts	242 posts
True Negative (TN)	305 posts	306 posts
False Negative (FN)	95 posts	94 posts

The comparison of M-COVIDLex performance evaluation results between the domain-specific social media text corpus dataset and the general domain social media text corpus dataset [64] in Fig 12 allows for several conclusions to be drawn.

- M-COVIDLex demonstrated superior error rates and accuracy on the domain-specific social media text corpus compared to the general domain social media text corpus. The observed result aligns with expectations, as M-COVIDLex was developed utilising lexicons that contain sentiment values within a domain-specific social media text corpus.
- The quantity of posts exhibiting positive sentiment in both datasets is identical, totalling 400 posts. Nevertheless, the sensitivity outcomes of M-COVIDLex on the domain-specific social media text corpus dataset surpass those of the general domain social media text corpus dataset. M-COVIDLex accurately predicted 274 out of 400 posts with positive sentiment in the domain-

specific social media text corpus, compared to 158 in the general domain corpus. Additionally, the classification error for M-COVIDLex in the domain-specific dataset is marginally lower than that in the general dataset. The sensitivity results indicate that M-COVIDLex effectively identified true positive posts and significantly minimised false negative posts within the domain-specific social media text corpus dataset.

- The quantity of posts exhibiting negative sentiment in both datasets is identical, totalling 400 posts. Nevertheless, the specificity outcomes of M-COVIDLex are superior in the domain-specific social media text corpus compared to the general domain social media text corpus. Although the number of negative posts indicating negative sentiment in the domain-specific social media text corpus dataset is slightly higher than in the general domain dataset, the number of negative posts misclassified as positive in the domain-specific dataset is significantly lower than in the general dataset. The findings demonstrate that M-COVIDLex effectively identifies true negative posts and minimises false positive posts within the domain-specific social media text corpus dataset.
- The precision results of M-COVIDLex are superior on the domain-specific social media text corpus dataset compared to the general domain social media text corpus dataset, exhibiting a difference of 29 percent. The findings indicate that the sentiment analysis model demonstrates improved accuracy in predicting true positive sentiment posts within the domain-specific social media text corpus dataset, thereby minimising the classification error associated with false positive posts. This result was anticipated, as the sentiment analysis is dependent on the lexicon list in M-COVIDLex.
- The F1-score results of M-COVIDLex are superior on the domain-specific social media text corpus dataset compared to the general domain social media text corpus dataset. The results demonstrate that M-COVIDLex effectively balances the analysis of sentiment within the domain-specific social media text corpus dataset.

The sentiment analysis results derived from the general domain social media text corpus dataset are more significant than those from the domain-specific social media text corpus dataset. This result demonstrates that M-COVIDLex, developed from a domain-specific social media text corpus, is exclusively appropriate for sentiment analysis within the same domain corpus. The application of this method on a general domain-specific social media text corpus is not advisable, as the resulting analysis yields significantly lower outcomes and fails to offer any valuable insights. Despite M-COVIDLex demonstrating superior performance on a domain-specific social media text corpus, its application to analyse sentiments in other specific domains, such as banking, entertainment, and food, yields a low probability of accurate sentiment analysis results for those domains. The primary purpose of constructing M-COVIDLex is to address the public health emergency related to COVID-19 in Malaysia. Given that most domains possess distinct sub-languages, the implementation of M-COVIDLex on other public

health emergency domain-specific corpora, such as Influenza, is anticipated to yield superior sentiment analysis performance compared to general domain corpora. The lexicon in M-COVIDLex includes general terms related to public health emergencies, including vaccines, protection, monitoring, and prevention. The enumeration of these general terms demonstrates that M-COVIDLex's contribution extends beyond COVID-19 to encompass other public health emergencies.

V. DISCUSSION

The experiments conducted in the previous section highlight the significance of analysing sentiment within a domain-specific social media text corpus dataset, utilizing a sentiment lexicon derived from the same corpus. The performance evaluation of M-COVIDLex on the social media text corpus dataset indicates that, despite the inclusion of 10,511 sentiment lexicons, the accuracy results remained below 80 percent. The limitations on the expansion of Malay synonyms may stem from the selection of only level one synonym words as seed words. Additionally, given that M-COVIDLex is derived from a domain-specific social media text corpus, it is logical that the accuracy results of this dataset surpass those of a general domain social media text corpus dataset. This analysis confirms the assertion from [55] that sentiment lexicons developed from domain-specific corpora are capable of providing high classification accuracy and thorough insights exclusively for that domain. The efficiency of the proposed method is significantly affected by the quality of the developed lexical dictionaries for all seven Malay POS tags [8]. The process necessitates the engagement of human resources to manually annotate the data [94], a task that is both time-consuming and costly, as indicated by multiple studies [12,29,63,99]. This proposed method may establish a baseline for future research on sentiment analysis of multilingual, code-mixed, or code-switching social media texts. Future initiatives in safety sectors during public health emergencies require the creation of an application that can analyse code-mixed sentiment on social media. A tool of this nature would facilitate prompt notifications to government entities and pertinent organisations concerning individuals who breach movement control orders, rather than relying exclusively on physical enforcement strategies such as roadblocks. The performance evaluation results of the proposed method will assist industrial researchers and relevant agencies in comprehending the Malaysian sentiment regarding government-structured relief aids during the health crisis. The performance result can be enhanced by (i) utilising a larger corpus, ideally the entire dataset of 16,600 X posts, rather than a subset, and (ii) broadening the lexical dictionaries to include a greater variety of words, including antonyms and third-level synonyms.

VI. CONCLUSION

This paper outlines a method for constructing a domain-specific mixed code sentiment lexicon, referred to as M-COVIDLex, through the integration of corpus-based and dictionary-based techniques, utilising seven Malay POS tags: KA, KK, KAD, KN, FOR-NEG, FOR-POS, and NEG. This mixed code sentiment lexicon addresses the absence of a sentiment lexicon tailored for the public health emergency context, specifically regarding COVID-19 in Malaysia. The M-COVIDLex construction method consists of five fundamental

phases: (i) Acquiring the sentiment of Malaysians from social media platform X regarding the impact of government efforts in addressing the COVID-19 crisis on their daily lives; (ii) Processing the acquired data with an enhanced Malay Normaliser and an improved set of 46 Malay POS tags; (iii) Constructing a mixed code sentiment lexicon through seed word selection, annotation, and synonym expansion; (iv) Analysing the sentiments of the acquired data using grammatical rules of four Malay POS tags: KNF, KP, KB, and KH, along with word frequency calculations; and (v) Classifying the sentiments using a lexicon-based classification technique. The evaluation of sentiment classification is conducted through a confusion matrix and six metrics: error rate, accuracy, sensitivity, specificity, precision, and F1-score. The performance evaluation of the proposed M-COVIDLex on the domain-specific social media text corpus dataset exceeds its performance on the general domain social media text corpus dataset.

ACKNOWLEDGMENT

The first author has rendered M-COVIDLex available for public use and future research endeavours. The list is available for download at the following link: <https://doi.org/10.6084/m9.figshare.26826250.v1>.

REFERENCES

- [1] Rajkumar Buyya, Calheiros, R. N., & Amir Vahid Dastjerdi. (2016). Big data : principles and paradigms. Elsevier/Morgan Kaufmann.
- [2] Bakar, M. F. R. A., Idris, N., Shuib, L., & Khamis, N. (2020). Sentiment Analysis of Noisy Malay Text: State of Art, Challenges and Future Work. *IEEE Access*, 8, 24687-24696.
- [3] Zunic, A., Corcoran, P., & Spasic, I. (2020). Sentiment analysis in health and well-being: systematic review. *JMIR medical informatics*, 8(1), e16023.
- [4] Nandwani, P., & Verma, R. (2021). A review on sentiment analysis and emotion detection from text. *Social network analysis and mining*, 11(1), 81.
- [5] Enjop, V., Adnan, R., Jamil, N., Ahmad, S., Zainol, Z., & Ahmad, S. A. (2022). Does Google Translate Affect Lexicon-Based Sentiment Analysis of Malay Social Media Text?. *Malaysian Journal of Computing*, 7(2), 1236-1249.
- [6] Hartmann, J., Heitmann, M., Siebert, C., & Schamp, C. (2023). More than a feeling: Accuracy and application of sentiment analysis. *International Journal of Research in Marketing*, 40(1), 75-87.
- [7] Cambridge Dictionary. (2019, December 4). SLANG | meaning in the Cambridge English Dictionary. Cambridge.org. <https://dictionary.cambridge.org/dictionary/english/slang>
- [8] Drus, Z., & Khalid, H. (2019). Sentiment analysis in social media and its application: Systematic literature review. *Procedia Computer Science*, 161, 707-714.
- [9] Darwich, M., Mohd, S. A., Omar, N., & Osman, N. A. (2019). Corpus-Based Techniques for Sentiment Lexicon Generation: A Review. *J. Digit. Inf. Manag.*, 17(5), 296.
- [10] Birjali, M., Kasri, M., & Beni-Hssane, A. (2021). A comprehensive survey on sentiment analysis: Approaches, challenges and trends. *Knowledge-Based Systems*, 226, 107134.
- [11] Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, 55(7), 5731-5780.
- [12] Tan, K. L., Lee, C. P., & Lim, K. M. (2023). A survey of sentiment analysis: Approaches, datasets, and future research. *Applied Sciences*, 13(7), 4550.
- [13] Yusuf, A., Sarlan, A., Danyaro, K. U., Rahman, A. S. B., & Abdullahi, M. (2024). Sentiment Analysis in Low-Resource Settings: A

- Comprehensive Review of Approaches, Languages, and Data Sources. IEEE Access.
- [14] Batanović, V., Cvetanović, M., & Nikolić, B. (2020). A versatile framework for resource-limited sentiment articulation, annotation, and analysis of short texts. *PLoS One*, 15(11), e0242050.
- [15] Meetei, L. S., Singh, T. D., Borgohain, S. K., & Bandyopadhyay, S. (2021). Low resource language specific pre-processing and features for sentiment analysis task. *Language Resources and Evaluation*, 55(4), 947-969.
- [16] Kumari, D., Ekbal, A., Haque, R., Bhattacharyya, P., & Way, A. (2021). Reinforced nmt for sentiment and content preservation in low-resource scenario. *Transactions on Asian and Low-Resource Language Information Processing*, 20(4), 1-27.
- [17] Marreddy, M., Oota, S. R., Vakada, L. S., Chinni, V. C., & Mamidi, R. (2022). Am I a resource-poor language? Data sets, embeddings, models and analysis for four different NLP tasks in telugu language. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(1), 1-34.
- [18] Ekbal, A., Bhattacharyya, P., Saha, T., Kumar, A., & Srivastava, S. (2022, June). HindiMD: A multi-domain corpora for low-resource sentiment analysis. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 7061-7070).
- [19] Zabha, N. I., Ayop, Z., Anawar, S., Hamid, E., & Abidin, Z. Z. (2019). Developing cross-lingual sentiment analysis of Malay Twitter data using lexicon-based approach. *International Journal of Advanced Computer Science and Applications*, 10(1).
- [20] Mahadzir, N. H. (2021). Sentiment Analysis of Code-Mixed Text: A Review. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(3), 2469-2478.
- [21] Konate, A., & Du, R. (2018). Sentiment analysis of code-mixed Bambara-French social media text using deep learning techniques. *Wuhan University Journal of Natural Sciences*, 23(3), 237-243.
- [22] Srinivasan, R., & Subalalitha, C. N. (2023). Sentimental analysis from imbalanced code-mixed data using machine learning approaches. *Distributed and Parallel Databases*, 41(1), 37-52.
- [23] Hidayatullah, A. F., Apong, R. A., Lai, D. T., & Qazi, A. (2023). Corpus creation and language identification for code-mixed Indonesian-Japanese-English Tweets. *PeerJ Computer Science*, 9, e1312.
- [24] Laumann, F. (2022, June 10). Low-resource language: what does it mean? *NeuralSpace*. <https://medium.com/neuralspace/low-resource-language-what-does-it-mean-d067ec85dea5>
- [25] Nasharuddin, N. A., Abdullah, M. T., Azman, A., & Kadir, R. A. (2017). English and Malay cross-lingual sentiment lexicon acquisition and analysis. In *Information Science and Applications 2017: ICISA 2017 8* (pp. 467-475). Springer Singapore.
- [26] Magueresse, A., Carles, V., & Heetderks, E. (2020). Low-resource languages: A review of past work and future challenges. *arXiv preprint arXiv:2006.07264*.
- [27] Deng, D., Jing, L., Yu, J., & Sun, S. (2019). Sparse self-attention LSTM for sentiment lexicon construction. *IEEE/ACM transactions on audio, speech, and language processing*, 27(11), 1777-1790.
- [28] Alsolamy, A. A., Siddiqui, M. A., & Khan, I. H. (2019). A corpus based approach to build arabic sentiment lexicon. *International Journal of Information Engineering and Electronic Business*, 11(6), 16-23.
- [29] Machová, K., Mikula, M., Gao, X., & Mach, M. (2020). Lexicon-based sentiment analysis using the particle swarm optimization. *Electronics*, 9(8), 1317.
- [30] Wang, Y., Yin, F., Liu, J., & Tosato, M. (2020). Automatic construction of domain sentiment lexicon for semantic disambiguation. *Multimedia Tools and Applications*, 79, 22355-22373.
- [31] Du, M., Li, X., & Luo, L. (2021). A Training-Optimization-Based Method for Constructing Domain-Specific Sentiment Lexicon. *Complexity*, 2021(1), 6152494.
- [32] Chaturanga, P. D. T., Lorensuhewa, S. A. S., & Kalyani, M. A. L. (2019, September). Sinhala sentiment analysis using corpus based sentiment lexicon. In *2019 19th international conference on advances in ICT for emerging regions (ICTer)* (Vol. 250, pp. 1-7). IEEE.
- [33] Tho, C., Heryadi, Y., Lukas, L., & Wibowo, A. (2021, April). Code-mixed sentiment analysis of Indonesian language and Javanese language using Lexicon based approach. In *Journal of Physics: Conference Series* (Vol. 1869, No. 1, p. 012084). IOP Publishing.
- [34] Miller, G. A. (1995). *WordNet: a lexical database for English*. *Communications of the ACM*, 38(11), 39-41.
- [35] Shayaa, S., Jaafar, N. I., Bahri, S., Sulaiman, A., Wai, P. S., Chung, Y. W., ... & Al-Garadi, M. A. (2018). Sentiment analysis of big data: methods, applications, and open challenges. *Ieee Access*, 6, 37807-37827.
- [36] Labille, K., Gauch, S., & Alfarhood, S. (2017, August). Creating domain-specific sentiment lexicons via text mining. In *Proc. Workshop Issues Sentiment Discovery Opinion Mining (WISDOM)* (pp. 1-8).
- [37] Sazzed, S. (2020, August). Development of sentiment lexicon in bengali utilizing corpus and cross-lingual resources. In *2020 IEEE 21st International conference on information reuse and integration for data science (IRI)* (pp. 237-244). IEEE.
- [38] Piryani, R., Piryani, B., Singh, V. K., & Pinto, D. (2020). Sentiment analysis in Nepali: exploring machine learning and lexicon-based approaches. *Journal of Intelligent & Fuzzy Systems*, 39(2), 2201-2212.
- [39] Sun, S., Luo, C., & Chen, J. (2017). A review of natural language processing techniques for opinion mining systems. *Information fusion*, 36, 10-25.
- [40] Bonta, V., Kumares, N., & Janardhan, N. (2019). A comprehensive study on lexicon based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, 8(S2), 1-6.
- [41] Yang, L., Li, Y., Wang, J., & Sherratt, R. S. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE access*, 8, 23522-23530.
- [42] Ofek, N., Caragea, C., Rokach, L., Biyani, P., Mitra, P., Yen, J., ... & Greer, G. (2013, May). Improving sentiment analysis in an online cancer survivor community using dynamic sentiment lexicon. In *2013 international conference on social intelligence and technology* (pp. 109-113). IEEE.
- [43] Han, H., Zhang, J., Yang, J., Shen, Y., & Zhang, Y. (2018). Generate domain-specific sentiment lexicon for review sentiment analysis. *Multimedia Tools and Applications*, 77, 21265-21280.
- [44] Liu, Y., Jiang, C., & Zhao, H. (2019). Assessing product competitive advantages from the perspective of customers by mining user-generated content on social media. *Decision Support Systems*, 123, 113079.
- [45] Chekima, K., Alfred, R., & Chin, K. O. (2017). Rule-based model for Malay text sentiment analysis. In *Computational Science and Technology: 4th ICCST 2017, Kuala Lumpur, Malaysia, 29-30 November, 2017* (pp. 172-185). Springer Singapore.
- [46] Shamsudin, N. F., Basiron, H., & Sa'aya, Z. (2016). Lexical based sentiment analysis-verb, adverb & negation. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, 8(2), 161-166.
- [47] Mahadzir, N. H., Omar, M. F., Nawi, M. N. M., Salameh, A. A., Hussin, K. C., & Sohail, A. (2022). Melex: The construction of malay-english sentiment lexicon. *Computers, Materials and Continua*.
- [48] Suhaimi, S. H., Bakar, N. A. A., & Azmi, N. F. M. (2021). Proposing Malay Sarcasm Detection on Social Media Services: A Machine Learning Approach. *Open International Journal of Informatics*, 9(Special Issue 2), 1-10.
- [49] Tan, Y. F., Lam, H. S., Azlan, A., & Soo, W. K. (2016, April). Sentiment Analysis for Telco Popularity on Twitter Big Data Using a Novel Malaysian Dictionary. In *ICADIWT* (pp. 112-125).
- [50] Anbananthen, K. S. M., Selvaraju, S., & Krishnan, J. K. (2017). The generation of malay lexicon. *Am. J. Applied Sci*, 14, 503-510.
- [51] Selvaraju, S., & Anbananthen, K.S. (2019). Opinion Extraction on Online Malay Text. *American Journal of Applied Sciences*, 16, 134-142.
- [52] Li, W., Zhu, L., Guo, K., Shi, Y., & Zheng, Y. (2018). Build a tourism-specific sentiment lexicon via word2vec. *Annals of Data Science*, 5, 1-7.
- [53] Muhammad, S. H., Brazdil, P., & Jorge, A. (2020). Incremental approach for automatic generation of domain-specific sentiment lexicon. In *Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14-17, 2020, Proceedings, Part II 42* (pp. 619-623). Springer International Publishing.

- [54] Kreutz, T., & Daelemans, W. (2018). Enhancing general sentiment lexicons for domain-specific use. In Proceedings of the 27th International Conference on Computational Linguistics, Santa Fe, New Mexico, USA, August 20-26, 2018 (pp. 1056-1064).
- [55] Feng, J., Gong, C., Li, X., & Lau, R. Y. (2018). Automatic approach of sentiment lexicon generation for mobile shopping reviews. *Wireless Communications and Mobile Computing*, 2018.
- [56] Almatarneh, S., & Gamallo, P. (2018). Automatic construction of domain-specific sentiment lexicons for polarity classification. In Trends in Cyber-Physical Multi-Agent Systems. The PAAMS Collection-15th International Conference, PAAMS 2017 15 (pp. 175-182). Springer International Publishing.
- [57] Bergsma, T., van Stegeren, J., & Theune, M. (2020, May). Creating a sentiment lexicon with game-specific words for analyzing NPC dialogue in the elder scrolls V: Skyrim. In Workshop on Games and Natural Language Processing (pp. 1-9).
- [58] Shaukat, K., Hameed, I. A., Luo, S., Javed, I., Iqbal, F., Faisal, A., ... & Adeem, G. (2020). Domain Specific Lexicon Generation through Sentiment Analysis. *International Journal of Emerging Technologies in Learning (iJET)*, 15(9), 190-204.
- [59] Singh, V., Singh, G., Rastogi, P., & Deswal, D. (2018, December). Sentiment analysis using lexicon based approach. In 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC) (pp. 13-18). IEEE.
- [60] Alexander, N. S., & Omar, N. (2017). Generating a Malay Sentiment Lexicon Based on WordNet. *Asia-Pacific Journal of Information Technology and Multimedia*, 6(1).
- [61] bin Rodzman, S. B., Rashid, M. H., Ismail, N. K., Abd Rahman, N., Aljunid, S. A., & Abd Rahman, H. (2019, April). Experiment with lexicon based techniques on domain-specific Malay document sentiment analysis. In 2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE) (pp. 330-334). IEEE.
- [62] Sukawai, E. Z. U. A. N. A., & Omar, N. A. Z. L. I. A. (2020). Corpus Development for Malay Sentiment Analysis Using Semi Supervised Approach. *Asia-Pacific Journal of Information Technology and Multimedia*, 9(01), 94-109.
- [63] Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), 267-307.
- [64] Ariffin, S. N. A. N., & Tiun, S. (2022). Improved POS Tagging Model for Malay Twitter Data based on Machine Learning Algorithm. *International Journal of Advanced Computer Science and Applications*, 13(7).
- [65] Safiah, N., Onn, F. M., Musa, H. H., & Mahmood, A. H. (2010). *Tatabahasa Dewan Edisi Ketiga*. Kuala Lumpur: Dewan Bahasa dan Pustaka.
- [66] Ariffin, S. N. A. N., & Tiun, S. (2020). Rule-based text normalization for Malay social media texts. *International Journal of Advanced Computer Science and Applications*, 11(10).
- [67] Salas-Zárate, M. D. P., Medina-Moreira, J., Lagos-Ortiz, K., Luna-Aveiga, H., Rodríguez-García, M. A., & Valencia-García, R. (2017). Sentiment analysis on tweets about diabetes: An aspect-level approach. *Computational and mathematical methods in medicine*, 2017(1), 5140631.
- [68] Ikoro, V., Sharmina, M., Malik, K., & Batista-Navarro, R. (2018, October). Analyzing sentiments expressed on Twitter by UK energy company consumers. In 2018 Fifth international conference on social networks analysis, management and security (SNAMS) (pp. 95-98). IEEE.
- [69] Jiang, K., & Li, Y. (2020, December). Mining customer requirement from online reviews based on multi-aspected sentiment analysis and Kano model. In 2020 16th Dahe Fortune China Forum and Chinese High-educational Management Annual Academic Conference (DFHMC) (pp. 150-156). IEEE.
- [70] Yuan, H., Tang, Y., Xu, W., & Lau, R. Y. K. (2021). Exploring the influence of multimodal social media data on stock performance: an empirical perspective and analysis. *Internet Research*, 31(3), 871-891.
- [71] Zhi, S., Li, X., Zhang, J., Fan, X., Du, L., & Li, Z. (2017, August). Aspects opinion mining based on word embedding and dependency parsing. In Proceedings of the International Conference on Advances in Image Processing (pp. 210-215).
- [72] Chen, Y., & Ji, W. (2021). Public demand urgency for equitable infrastructure restoration planning. *International Journal of Disaster Risk Reduction*, 64, 102510.
- [73] Aqlan, A. A. Q., Manjula, B., & Naik, R. L. (2019). A Study of Sentiment Analysis: Concepts, Techniques, and Challenges. In Proceedings of International Conference on Computational Intelligence and Data Engineering (pp. 147-162). Springer, Singapore.
- [74] Sanagar, S., & Gupta, D. (2020). Unsupervised genre-based multidomain sentiment lexicon learning using corpus-generated polarity seed words. *IEEE Access*, 8, 118050-118071.
- [75] Stone, P. J., Bales, R. F., Namenwirth, J. Z., & Ogilvie, D. M. (1962). The general inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. *Behavioral Science*, 7(4), 484.
- [76] Hu, M., & Liu, B. (2004, July). Mining opinion features in customer reviews. In AAAI (Vol. 4, No. 4, pp. 755-760).
- [77] Wilson, T., Wiebe, J., & Hoffmann, P. (2005). Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. In Proceedings of HLT/EMNLP.
- [78] Sebastiani, F., & Esuli, A. (2006, May). Sentiwordnet: A publicly available lexical resource for opinion mining. In Proceedings of the 5th international conference on language resources and evaluation (pp. 417-422). European Language Resources Association (ELRA) Genoa, Italy.
- [79] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., & Kappas, A. (2010). Sentiment strength detection in short informal text. *Journal of the American society for information science and technology*, 61(12), 2544-2558.
- [80] Cambria, E., Speer, R., Havasi, C., & Hussain, A. (2010, November). Senticnet: A publicly available semantic resource for opinion mining. In 2010 AAAI fall symposium series.
- [81] Nielsen, F. Å. (2011). A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. arXiv preprint arXiv:1103.2903.
- [82] Mohammad, S. M., & Turney, P. D. (2013). Nrc emotion lexicon. *National Research Council, Canada*, 2, 234.
- [83] Hutto, C., & Gilbert, E. (2014, May). Vader: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the international AAAI conference on web and social media (Vol. 8, No. 1, pp. 216-225).
- [84] Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015.
- [85] Loria, S. (2018). textblob Documentation. Release 0.15, 2(8), 269.
- [86] Liu, B. (2022). Sentiment analysis and opinion mining. Springer Nature.
- [87] Handayani, D., Bakar, N. S. A. A., Yaacob, H., & Abuzaraida, M. A. (2018, July). Sentiment analysis for Malay language: systematic literature review. In 2018 International Conference on Information and Communication Technology for the Muslim World (ICT4M) (pp. 305-310). IEEE.
- [88] Bakar, N. S. A. A., Rahmat, R. A., & Othman, U. F. (2019). Polarity classification tool for sentiment analysis in Malay language. *IAES International Journal of Artificial Intelligence*, 8(3), 259.
- [89] Hijazi, M. H. A., Libin, L., Alfred, R., & Coenen, F. (2016, October). Bias aware lexicon-based Sentiment Analysis of Malay dialect on social media data: A study on the Sabah Language. In 2016 2nd International Conference on Science in Information Technology (ICSITech) (pp. 356-361). IEEE.
- [90] Talbot, J., Charron, V., & Konkle, A. T. (2021). Feeling the void: lack of support for isolation and sleep difficulties in pregnant women during the COVID-19 pandemic revealed by Twitter data analysis. *International Journal of Environmental Research and Public Health*, 18(2), 393.
- [91] Google Trends. (n.d.). Google's Year in Search. Retrieved June 22, 2021, from <https://trends.google.com/trends/yis/2020/MY/>
- [92] Shakeel, S., Ahmed Hassali, M. A. & Abbas Naqvi, A. 2020. Health and economic impact of covid-19: Mapping the consequences of a pandemic in malaysia. *Malaysian Journal of Medical Sciences* 27(2): 159-164. doi:10.21315/mjms2020.27.2.16

- [93] Abdullah, N. A. S., & Rusli, N. I. A. (2021). Multilingual Sentiment Analysis: A Systematic Literature Review. *Pertanika Journal of Science & Technology*, 29(1).
- [94] Sadia, A., Khan, F., & Bashir, F. (2018, February). An overview of lexicon-based approach for sentiment analysis. In 2018 3rd International Electrical Engineering Conference (IEEC 2018) (pp. 1-6).
- [95] Ariffin, S. N. A. N., & Tiun, S. (2018). Part-of-Speech Tagger for Malay Social Media Texts. *GEMA Online® Journal of Language Studies*, 18(4).
- [96] Pustejovsky, J., & Stubbs, A. (2012). *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications.* " O'Reilly Media, Inc."
- [97] Shamsudin, N. F., Basiron, H., Saaya, Z., Rahman, A. F. N. A., Zakaria, M. H., & Hassim, N. (2015). Sentiment classification of unstructured data using lexical based techniques. *Jurnal Teknologi*, 77(18).
- [98] Iqbal, M., Karim, A., & Kamiran, F. (2015, April). Bias-aware lexicon-based sentiment analysis. In *Proceedings of the 30th Annual ACM Symposium on Applied Computing* (pp. 845-850).
- [99] Hota, H. S., Sharma, D. K., & Verma, N. (2021). Lexicon-based sentiment analysis using Twitter data: a case of COVID-19 outbreak in India and abroad. In *Data science for COVID-19* (pp. 275-295). Academic Press.

Optimized Hybrid Deep Learning for Enhanced Spam Review Detection in E-Commerce Platforms

Abdulrahman Alghaligah, Ahmed Alotaibi, Qaisar Abbas, and Sarah Alhumoud*

College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU),
Riyadh 11432, Saudi Arabia

Abstract—Spam reviews represent a real danger to e-commerce platforms, steering consumers wrong and trashing the reputations of products. Conventional Machine learning (ML) methods are not capable of handling the complexity and scale of modern data. This study proposes the novel use of hybrid deep learning (DL) models for spam review detection and experiments with both CNN-LSTM and CNN-GRU architectures on the Amazon Product Review Dataset comprising 26.7 million reviews. One important finding is that 200k words vocabulary, with very little preprocessing improves the models a lot. Compared with other models, the CNN-LSTM model achieves the best performance with an accuracy of 92%, precision of 92.22%, recall of 91.73% and F1-score of 91.98%. This outcome emphasizes the effectiveness of using convolutional layers to extract local patterns and LSTM layers to capture long-term dependencies. The results also address how high constraints and hyperparameter search, as well as general-purpose represents such as BERT. Such advancements will help in creating more reliable and reliable spam detection systems to maintain consumer trust on e-commerce platforms.

Keywords—Spam review detection; CNN-LSTM; CNN-RNN; CNN-GRU; big data; deep learning; amazon product review dataset

I. INTRODUCTION

E-commerce platforms are becoming the primary marketplaces for almost every good, replacing traditional stores in many fields. Both the seller and the customer widely accept them because they reduce costs for the seller and allow the customer to access the goods faster. Platforms like Amazon, Alibaba, and Noon dominate the global retail landscape. However, consumers face a difficult challenge when shopping online. They are unable to assess products before purchase. They tend to check online reviews and base their buying decision on them Najada and Zhu [1]. This has given rise to the issue of deceptive content, known as spam reviews. It aims to misguide consumers for specific gains. According to a report from the Department of Business & Trade from the UK government, “At least 10% of all product reviews on third-party e-commerce platforms are likely to be fake”[2]. This underlined the importance of checking the trustworthiness of online reviews. Since online reviews are important in consumers' buying decisions, spam review detection is a priority. When traditional spam detection techniques began, the focus was on areas like email. Nowadays, the focus has shifted to review-based spam detection Li et al. [3]. Spam detection can be categorized into two techniques: content-based approach and user-behavior-based approach. Content-based approach analyzes the text content only and extracts semantic relations

between words. On the other hand, the user-behavior approach focuses on the patterns of reviewer activities Li et al. [4].

Detecting spam reviews has several challenges. Spam reviews may look like genuine ones in content, making it difficult to distinguish spam from non-spam reviews. Also, spammers are always improving their techniques; they tend to use auto-generated tools to evade detection Bhuvaneshwari et al. [5]. ML started a revolution to protect consumers from scams. Traditional ML classifiers, such as Naive Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT), have good results in detecting spam reviews Rizali et al. [6]. Additionally, with the increasing complexity of the reviews, ML classifiers were not enough to handle this matter. Hence, DL comes to replace ML by showing a better performance in detecting complex patterns. However, current studies have not used DL on large datasets such as the Amazon Product Reviews dataset Hussain et al. [7] which contains 26.7 million labeled reviews. This opens a research gap to leverage hybrid deep learning models' capabilities to process large datasets, achieving superior performance. This study aims to enhance the ability to detect spam reviews by using large datasets leveraging the latest DL technologies. CNN-LSTM, CNN-RNN, and CNN-GRU were applied to the Amazon Products Reviews dataset Hussain et al. [7], which offers a rich and diverse set of labeled data. To optimize the detection process, various text preprocessing techniques are evaluated, including tokenization, lowercasing, lemmatization, stop word removal, punctuation removal, and embedding. The proposed framework provides an efficient, and accurate spam reviews detection, providing valuable insights into the impact of preprocessing techniques on detection outcomes. To verify the effectiveness of the proposed classifiers and preprocessing steps, accuracy, precision, recall, and F1Score were applied. The contributions of the paper are as follows:

1) A novel and accurate spam review detection is presented in this paper in which the accuracy is 92% through CNN-LSTM model which is quite better than results acquired from traditional ML methods leading to new detection accuracy benchmark.

2) Using the Amazon Product Review Dataset (26.7 million reviews) for demonstration, the research provides evidence for how hybrid models can be utilized to efficiently process vast-scale, diverse data, which detects an essential scalability barrier.

3) In the same work, the authors state that less preprocessing obtains better performance, since it retains

important information from the text, and that the common tendency to make more preprocessing does not favor the use of the model; these findings can provide important guidelines for the design of future models.

4) This study found that a large vocabulary size (200,000 max words) allows the model to better represent complex relationships between words, showing that careful vocabulary selection can improve spam detection effectiveness significantly.

5) This paper suggested a scalable and efficient spam detecting framework to better maintain consumer trust in the marketplace by excluding fake reviews and pretending user occurrence.

The rest of the paper is structured as follows: Section I and Section II discusses the background of spam detection and previous works in the field. Section III presents the proposed methodology. Section IV shows the result of the proposed work and discusses the findings in Section V. Finally, Section VI presents a summary of the study and future work.

II. LITERATURE REVIEW

Spam refers to deceptive content intentionally made to mislead or manipulate users for a specific gain, usually a commercial gain Jakupov et al. [8]. In online reviews, spam is a false or misleading review designed to promote or demote a particular product, store, book, or goods and services. It can influence customers purchasing decisions and damage the reputation of products on e-commerce platforms like Amazon Fei et al. [9]. Most users refer to reviews to decide whether they buy a product, as they need physical access to assess it. Studies indicate that nearly 30% of online reviews are spam Farooq [10], highlighting the issue of reviews on famous platforms such as Amazon. Spam detection in online reviews has become crucial as most e-commerce platforms rely heavily on user input to guide consumer choices. In the past decade, spam detection focused more on traditional applications like spam Short Message Service (SMS). However, in the era of online stores, especially in advanced countries like China, attention has moved towards detecting deceptive reviews Li et al. [3]. Spam detection in online reviews can be classified into content-based or user behavior-based techniques. Content-based analyzes textual features of reviews, linguistic patterns, and sentiment analysis. User behavior-based focuses on patterns like reviewing users' behavior, metadata, and social connections between reviewers Li et al. [3], Ennaouri and Zellou [11]. Usually, these techniques are combined to improve detection accuracy. Identifying spam reviews is a sophisticated task due to several challenges. One of the primary challenges is distinguishing between genuine and fake reviews. While counterfeit reviews may imitate the style and content of genuine ones, specific patterns such as overuse of promotional language or usually high volumes of reviews in short periods can provide clues Li et al. [3]. Language variability also helps in detection, as spam reviews may appear in multiple languages or use specific accents Ennaouri and Zellou [11]. Moreover, spammers continuously improve their techniques, making relying on static detection methods difficult. They may use techniques such as duplicating legitimate reviews or using an

automated system to generate spam, which can evade traditional methods Bhuvaneshwari et al. [5]. The large number of online reviews presents a scalability challenge as manual moderation becomes unfeasible for platforms like Amazon, which hosts millions of products with billions of reviews. ML and DL are powerful technologies for overcoming the challenges in spam detection. Traditional ML models, such as NB and SVM, have been used widely to classify reviews based on a content-based technique Saumya and Singh [12]. However, these models often struggle with the complexity of spam patterns in large datasets Kalaivani et al. [13], and this will be discussed in the next subsection based on related work. DL models like Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and their hybrid variants perform superiorly in detecting spam reviews Ghourabi et al. [14], Deshai and Rao [15], Shahariar et al. [16]. These models can learn complex patterns in text and capture long relations between words, making them more effective at detecting spam. Additionally, attention mechanisms, such as the one used in self-attention-based models, have successfully identified key features of spam reviews by focusing on specific text parts Bhuvaneshwari et al. [5]. The combination of DL and traditional ML technologies offers promising solutions to detect spam in online reviews. The following section will examine existing research on spam reviews using both ML and DL technologies.

A. Spam Review Detection

Many studies have discussed spam detection methods. The following subsections will discuss these studies in detail, starting with traditional ML. After that, DL studies showed promising improvement in solving this problem. "Table I" summarizes all the ten related works, from both traditional ML and DL subsections.

1) *Traditional machine learning approach for spam review detection:* Traditional ML approaches were used to solve the problem of detecting Spam reviews using classifiers such as NB, SVM, and Random Forest (RF) Ahsan et al. [17], Tripathy et al. [18]. These classifiers are usually used alongside feature selection techniques to detect fake reviews. In Kalaivani et al. [13], two traditional ML models were used to detect spam reviews. The first algorithm was SVM, while the second was NB. The dataset that was used is from Kaggle, with 20k reviews. By preprocessing the data before training the above-mentioned models, SVM achieved 76%, while NB achieved 84%. In Etaiwi and Naymat [19], the author tried a bunch of traditional ML algorithms, which are Gradient Boosted Trees (GBT), NB, RF, DT, and SVM. All these algorithms were used to train models on the Hotel Reviews dataset of 1600 reviews named Deceptive Opinion Spam Corpus (DOSC). Among all these models, SVM achieved the best accuracy with 85.5%. In Saeed et al. [20], many spam review detection approaches were evaluated. The paper proposed four detection approaches: a rule-based classifier, machine learning classifiers, a majority voting classifier, and a stacking ensemble classifier. All these approaches were trained and tested on two datasets, DOSC and Hotel Arabic Reviews Dataset (HADR). The stacking ensemble approach clearly outperformed the other approaches with

95.25% accuracy on the DOSC dataset and 99.98% on the HARD dataset by combining a rule-based classifier with a k-means classifier.

In Ibrahim et al. [21] the authors investigated the use of ensemble learning techniques to enhance spam review detection accuracy. They explored three classifiers: NB, SVM, and Logistic Regression (LR). They combined these algorithms to form an ensemble classifier. They used the Amazon dataset and got the best accuracy of 88.09%. Etaiwi and Awajan [22] explored how different feature selection methods impact the performance of spam review detection. They applied four ML algorithms: NB, SVM, DT, and RF. They used the DOSC dataset and got the best accuracy results of 87.31% with NB.

2) *Deep learning approaches for spam review detection:* Deep learning techniques have succeeded in enhancing the accuracy of spam detection. They are more effective than traditional ML approaches, which rely more on manually engineered features. DL models can automatically learn from

complex patterns, which makes them suitable for distinguishing between truthful and deceptive reviews. CNN, RNN, and hybrid approaches like CNN-RNN are widely used in the field Zhao et al. [23]. They have the potential to discover long relations between words. This section presents studies that have proposed DL techniques to detect spam reviews.

Shahariar et al. [16] presented a multi-layer perceptron (MLP) framework to detect spam reviews in the YELP and DOSC datasets. They compared the DL model with traditional ML, like NB and SVM. Their findings showed that Long Short-Term Memory (LSTM) outperformed all other algorithms with 96.75% accuracy. Deshai and Rao [15] Proposed two hybrid models integrating CNN and LSTM for fake reviews. Also, they presented LSTM-RNN for fake ratings detection. Their results showed that CNN-LSTM and LSTM-RNN methods are the most efficient, with 93.09% accuracy, using a subset of the Amazon Product Reviews dataset.

TABLE I. STATE-OF-THE-ART ALGORITHMS FOR SPAM REVIEW DETECTION

Paper	Year	Algorithm	Dataset	Best Accuracy
[13]	2023	NB, SVM	Review dataset from kaggle	84% NB
[19]	2017	GBT, NB, RF, DT, SVM	DOSC	85.5% SVM
[20]	2019	rule-based classifier, machine learning classifiers, majority voting classifier, stacking ensemble classifier	DOSC, HARD	99.98% rule-based + k-means
[16]	2019	MLP, CNN, LSTM	YELP., DOSC	96.75% LSTM
[15]	2023	CNN-LSTM, LSTM-RNN hybrid	Amazon Dataset	93.09% LSTM-RNN
[14]	2020	CNN-LSTM hybrid	UCI dataset	98.37% CNN-LSTM
[25]	2023	NB, KNN, SVM, CNN, LSTM	YELP DOSC	94.88% LSTM
[21]	2017	NB, SVM, LR	Amazon Dataset	88.09% ensemble
[26]	2020	LSTM Autoencoder	YouTube	-
[22]	2017	NB, SVM, DT, RF	DOSC	87.31% NB

Ghourabi et al. [14] proposed a hybrid CNN-LSTM model for detecting mixed text messages written in Arabic and English. They designed a model to let CNN capture n-gram features while LSTM is used to retain long-term information. They achieved an accuracy of 98.37% using a dataset from UCI repository Almeida et al. [24]. Singh et al. [25] Explored the effectiveness of DL models, especially for CNN and LSTM, the authors benchmarked ML models with DL. They emphasized the superior performance of deep learning models over traditional approaches for handling textual data. The datasets used in their study are YELP and DOSC. They got the best results using LSTM model with an accuracy of 94.88%. Saumya and Singh [26] Presented an unsupervised model for detecting spam reviews without requiring labeled data. The authors used a combination of LSTM networks and autoencoders to learn patterns of true reviews, allowing the models to distinguish reviews anomalies. They used a YouTube dataset, which includes reviews of popular videos. The authors did not use accuracy metrics in their study. The studies showed that DL approaches perform superiorly in spam review

detection, especially hybrid ones. In the next section, the gaps in existing studies will be discussed.

B. Gaps in Existing Research

Spam review detection is a hot topic that is widely covered. However, several advancements in the field, particularly in the DL field, introduce gaps and challenges that open the door to resolving this problem by applying new technologies. This section aims to discuss these challenges. These gaps need to be addressed and apply new technologies to resolve them. One primary challenge is the need for a labeled dataset Hussain et al. [27].

Hussain et al. [7] addressed this issue and solved it with Spam Review Detection using Behavioral Method (SRD-BM). However, new ML and DL technologies are needed to help in labeling the dataset and help researchers work on it in the future. Another significant challenge is the use of hybrid DL architecture on large datasets. Ghourabi et al. [14], Deshai and Rao [15], Shahariar et al. [16], Wayal and Bhandari [28] applied a hybrid method in small datasets. Applying this kind of architecture on large datasets requires vast computational

power. Fortunately, hybrid DL architecture has been applied to the Amazon Product Review Dataset, "Table II" shows the detailed distribution of the dataset. This study will discuss the proposed work in detail in the methodology section.

III. METHODOLOGY

This study demonstrated the application of three hybrid DL models for spam review detection. Each model was selected based on its ability to recognize complex patterns in large datasets and understand long-range word relations. Also, the effects of applying extensive text preprocessing will be discussed. In the following subsections, start by introducing the development environment. Then, the dataset and data preprocessing steps that convert human sentences to a readable format for the machine will be presented. After that, the feature selection process, the three models' architecture, and how each model is trained on the dataset will be described respectively.

C. Development Environment

In this study, all the preprocessing, training, and evaluation processes are conducted using Google Colab Pro. It is a solution on the cloud provided by Google to access high computational resources, utilizing the NVIDIA A100 graphics processing unit (GPU), which helped accelerate the training process across the hybrid DL models. Colab provides a high Random-Access Memory (RAM) up to 83GB and Virtual Random-Access Memory (VRAM) 40GB. It will allow the ability to load a large volume of data and preprocess it before feeding it into the neural network. Additionally, Python is the programming language used to implement the entire pipeline. Many libraries have been utilized to support this process, including Pandas and NumPy for data manipulation and preprocessing. Natural Language Toolkit (NLTK) is used to prepare text data. Scikit-learn for splitting train and test data, as well as for evaluation metrics. TensorFlow is used to train neural networks. Finally, Matplotlib and Seaborn were used to visualize the performance analysis.

D. Dataset

The dataset used for this study was acquired from Amazon Product Reviews Hussain et al. [7], which contains 26.7 million reviews written by 15.4 million reviewers on 3.1 million products. The dataset has six categories, shown in "Table II". The reviews cover many product categories, such as electronics, home and kitchen, and more, to ensure that the spam detection models are exposed to various review

patterns. The dataset's original source was unlabeled. However, Hussain et al. [7] did excellent work by labeling it using SRD-BM technique. The SRD-BM utilizes rich of behavioral features in the dataset to identify the spam and non-spam reviews, then labeling the data. In this study, the labeled dataset by the SRD-BM method was used, then preprocessing steps explained in the next section are applied.

E. Data Preprocessing

Preprocessing is an essential step to minimize the noise of the data and transform the raw text into a format that can be fed into the neural networks. However, applying the extensive preprocessing steps may upgrade or degrade the model performance HaCohen-Kerner et al. [29]. In this study, two different preprocessing steps were applied to hybrid DL models, one with extensive preprocessing steps that change the original text, another with a few steps that retain the original text. "Fig. 1." shows the preprocessing steps used in this study. For the extensive preprocessing steps, the text is converted to lowercase to standardize the input and reduce the complexity caused by case differences. This step will convert words like "Product" to "product", so they are treated as the same token. Additionally, all the punctuation marks are removed to avoid unnecessary symbols, which may add noise to the data. Another common concept in Natural Language Processing(NLP) was applied, which is Tokenization that converts text into numerical sequences to retain the frequent words in the dataset T. Limisiewicz et al. [30]. This will help reduce the noise in data and improve model efficiency. After that, Stop Words such as "is", "the", and "a" were removed. Also, it is essential to reduce every word to its root. Words like "playing" and "Played" will be reduced to "play" using a technique called Lemmatization. Also, to uniform the sequence length, Padding was used with configuration of 200 tokens in each sequence. Finally, an Embedding layer was added. It simply converts the integer sequences into dense vectors. This will allow the DL models to learn the relationship between words during the training Tegene et al. [31]. It is important to note that all these steps were applied to all the hybrid DL techniques to avoid biases in the benchmark. For the fewer preprocessing steps, only three steps were applied which are Tokenization, Padding, and Embedding. Also, for the max words parameter. Two configurations were applied, one with 10,000 and the other with 200,000 max words, this choice is due to the majority of words inside the dataset appearing very infrequently.

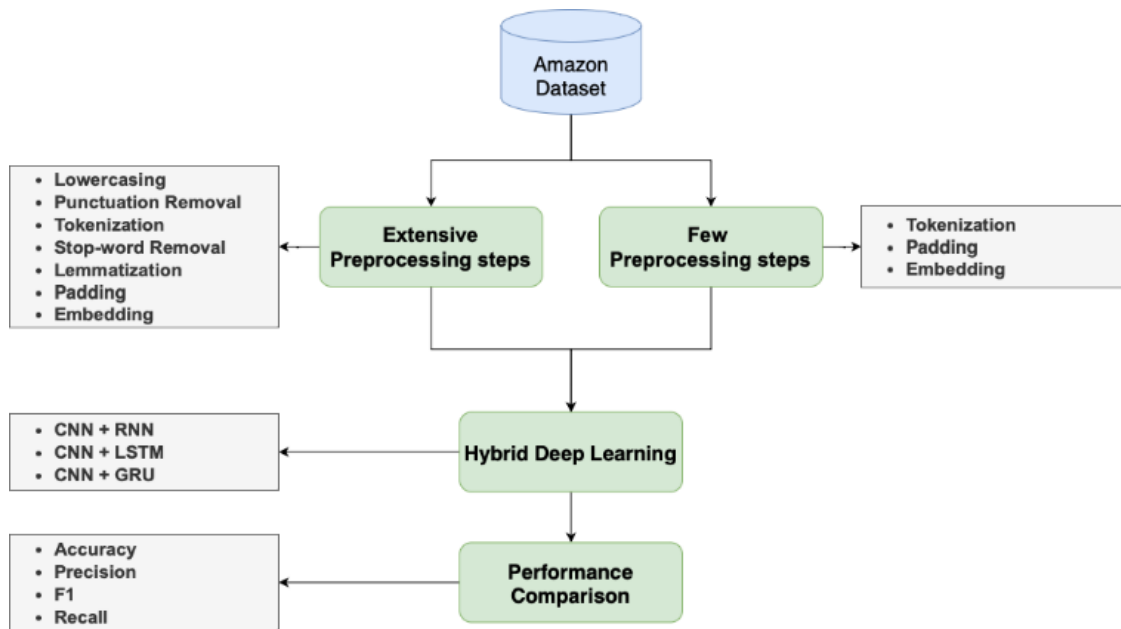


Fig. 1. Proposed system architecture.

F. Spam Review Detection

The methodology for detecting spam reviews involves designing and evaluating hybrid deep learning models. These models aim to extract both local features and long-term dependencies from textual data, leveraging the strengths of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), specifically Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU).

Since many people rely on reviews before purchasing products, detecting spam reviews can improve customer experience by protecting them from being scammed. Hence, this study intends to compare multiple hybrid DL approaches to reach the best classifier in this matter which are: CNN-LSTM, CNN-RNN, and CNN-GRU. Combining CNN with LSTM, RNN, and Gated Recurrent Unit (GRU) will give the ability to extract features by learning the local patterns and n-grams using the CNN layer Zhou et al. [32]. For CNN-based hybrids, the CNN component identifies key local features. Combining CNN-LSTM will make the classifier capture long-term patterns and model the temporal dependencies using the LSTM component Greff et al. [33]. Similarly, GRU can also be combined with CNN (CNN-GRU) to use its efficient gating mechanisms that will help avoid vanishing gradient problems Cho et al. [34]. Another component that can be combined with CNN is RNN (CNN-RNN), which captures context and temporal relationships in the text to help detect subtle patterns. “Table III” presents hyperparameters configuration of the implemented models.

The dataset used in this study is the Amazon Product Review Dataset, containing 26.7 million reviews across six categories. Each review x_i is labeled as spam ($y_i = 1$) or non-spam ($y_i = 0$). The raw text reviews are preprocessed to convert them into numerical representations. Two preprocessing strategies were compared: minimal and extensive preprocessing. Minimal preprocessing retained most

of the text structure, while extensive preprocessing involved steps like lowercasing, stop word removal, punctuation removal, and lemmatization.

TABLE II. AMAZON PRODUCT REVIEW DATASET

Category	Total Reviews	Total Reviewers	Total Products
Cell Phones and Accessories	3,446,396	2,260,636	319,652
Clothing, Shoes, and Jewellery	5,748,260	3,116,944	1,135,948
Electronics	7,820,765	4,200,520	475,910
Home and Kitchen	4,252,723	2,511,106	410,221
Sports and Outdoor	3,267,538	1,989,985	478,846
Toys and Games	2,251,775	1,342,419	327,653
Total	26,787,457	15,421,610	3,148,230

The input dataset can be represented as:

$$D = \{(x_i, y_i) | x_i \in R^d, y_i \in \{0,1\}, i = 1,2,\dots, n\} \quad (1)$$

Where d is the sequence length, and n is the total number of reviews. Each review x_i is tokenized and padded to ensure uniform length. The tokens are then passed through an embedding layer, which maps them into dense vector representations:

$$e_i = embedding(x_i), e_i \in R^{d \times n} \quad (2)$$

Where, the parameter m is the embedding dimension, and $d=200$ is the maximum sequence length.

The CNN layer is applied to extract local patterns and n-grams from the embedding vectors. The convolution operation is defined as:

$$c_{i,j} = ReLU(W_{conv} \times e_{i,j} + b_{conv}) \quad (3)$$

Where, the parameter W_{conv} and b_{conv} are the convolutional filter weights and biases. In addition, the operator (\times) represents the convolution operation. ReLU introduces non-linearity. The result, ci is a feature map containing extracted local patterns. To capture long-term dependencies in the text, the feature map ci is fed into an LSTM or GRU layer.

$$h_t = LSTM(C_t, h_{t-1}, C_{t-1}) \quad (4)$$

And for GRU

$$h_t = GRU(C_t, h_{t-1}) \quad (5)$$

Here, the ht parameter represents the hidden state at time t , which encodes sequential dependencies. The final hidden state h from the recurrent layer is passed through a fully connected layer to predict the probability of a review being spam:

$$p(y_i = 1 | x_i) = \sigma(W_{fc} \cdot h_i + b_{fc}) \quad (6)$$

Where, the parameter σ is the sigmoid activation function. Also, the parameter W_{fc} and b_{fc} are the weights and biases of the fully connected layer. The binary cross-entropy loss function is used to optimize the model:

$$Loss = -\frac{1}{n} \sum_{i=1}^n [y_i \log(p(y_i = 1 | x_i)) + (1 - y_i) \log p(1 - y_i = 1 | x_i)] \quad (7)$$

G. Evaluation Metrics

In our study, four evaluation metrics have been used to evaluate the models. These evaluation metrics depend on, True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN), defined as follows: TP is the number of correctly identified spam. TN is the number of reviews correctly identified as non-spam. FP is the number of non-spam reviews incorrectly identified as spam. FN is the number of spam reviews incorrectly identified as non-spam. All proposed models were evaluated using several key metrics, including Accuracy, Precision, Recall, and F1 score. Accuracy provides an overall measure of the model's performance in both positive and negative Sokolova and Lapalme [35].

$$Accuracy(ACC) = (TP + TN)/(TP + TN + FN + FP) \quad (8)$$

Precision is the evaluated proportion of correctly predicted as positive and shows model capability to avoid FP J. Davis and Goadrich [36], T. Saito and Rehmsmeier [37].

$$Precision(PR) = TP/(TP + FP) \quad (9)$$

Recall or, in other words, sensitivity, is the ratio of correctly predicted positives to all the actual positives J. Davis and Goadrich [36], T. Saito and Rehmsmeier [37].

$$Recall(RE) = TP/(TP + FN) \quad (10)$$

F1 Score is the harmonic mean of precision and Recall; it balances the two and is widely used in scenarios where precision and Recall are important.

$$F1 - score = 2 \times (PR \times RE)/(PR + RE) \quad (11)$$

These metrics are crucial to understanding the model's behavior across different situations. Therefore, they will all be

used in the next section to benchmark different Hybrid DL models.

IV. RESULTS

In this section, results of the DL models with and without the text preprocessing step are discussed. The comparison is based on the four metrics mentioned earlier, accuracy, precision, recall and F1 score. Then, the section will present a discussion of the result.

A. Results Analysis

The performance of hybrid DL models will be evaluated with and without some text preprocessing steps which are lowercasing, Stop Words removal, punctuation removal, and lemmatization. Also, two vocabulary sizes 10,000 and 200,000 max words will be used. "Table IV" shows all the results. Also, "Fig. 2." shows the models comparisons.

TABLE III. MODELS CONFIGURATIONS

Hyperparameters	CNN-LSTM	CNN-RNN	CNN-GRU
Batch size	128	128	128
Dropout	0.5 in each layer	0.5 in each layer	0.5 in each layer
Nodes	128 in LSTM layer	128 in RNN layer	128 in GRU layer
Training split	0.6	0.6	0.6
Testing split	0.2	0.2	0.2
Validation split	0.2	0.2	0.2
Epoch	10	10	10
Optimizer	Adam	Adam	Adam
Loss function	Binary cross-entropy	Binary cross-entropy	Binary cross-entropy
Vector size	128	128	128

Regarding the models trained with all the preprocessing steps, the CNN-LSTM achieved the highest performance with an accuracy of 88.88%, a precision of 88.39%, a recall of 89.52%, and an F1Score of 88.95%. The CNN-GRU model has a nearby result with an accuracy of 89%, a precision of 88.39%, a recall of 89.12%, and an F1Score of 88.75%. The CNN-RNN has the lowest performance with an accuracy of 86.64%, a precision of 87.07%, a recall of 86.05%, and an F1Score of 86.56%. All these results show that using a max word of 200,000 gives a better result than 10,000. When the models were trained by eliminating some preprocessing steps which are lowercasing, Stop Words removal, punctuation removal, and lemmatization, the CNN-LSTM also gave the best performance with an accuracy of 92%, a precision of 92.22%, a recall of 91.73%, and an F1Score of 91.98%. The CNN-GRU similarly performed well, accuracy of 92.08%, a precision of 92.22%, a recall of 91.19%, and an F1-score of 92.07%. While the CNN-RNN has the worst performance with an accuracy of 87.93%, a precision of 88.43%, a recall of 87.28%, and an F1Score of 87.85%.

The results showed that CNN-LSTM and CNN-GRU architectures give better performance without applying

extensive preprocessing steps. Among all the models, CNN-LSTM with 200,000 max words has the best overall performance compared to all the other architectures. This shows that the combination of CNN and LSTM layers with a large vocabulary is very effective in detecting spam reviews, while keeping the original text. This explains the ability of CNN-LSTM to capture both local patterns and long-term dependencies, which gives importance to understanding relationships across multiple words in a text. This result is aligned with recent studies on text classification using CNN-LSTM architecture Bhuvaneshwari et al. [5] Sagnika et al. [38].

Etaiwi and Naymat [19] showed that using many preprocessing steps affects the overall performance of spam review classification when using ML models. As in this study, the result of applying many preprocessing steps on hybrid DL models will affect the performance. When comparing the models based on the vocabulary size, 200,000 max words consistently performed better. This indicates giving the hybrid DL models a larger vocabulary helps capture more complicated relationships between words, which leads to better performance. Although the study showed a promising result, some limitations were faced. One major constraint was the use of Google Colab Pro. It is a paid service that allows you to use high computational resources, such as A100 GPUs and a high amount of RAM, with a certain number of compute units. This restricted the ability to empiric many models and try different hyperparameters. For that reason, it is recommended that future studies explore additional hyperparameter tuning to further improve model performance, as well as experimenting with pre-trained models like BERT to improve the ability to capture both local and contextual word representation. It can also help in transferring knowledge from one domain to another.

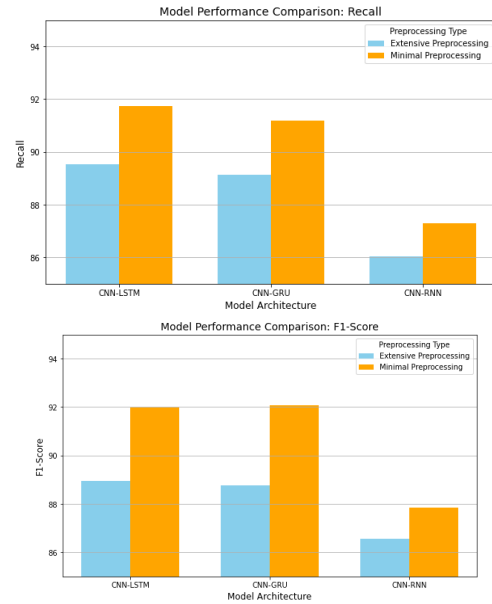


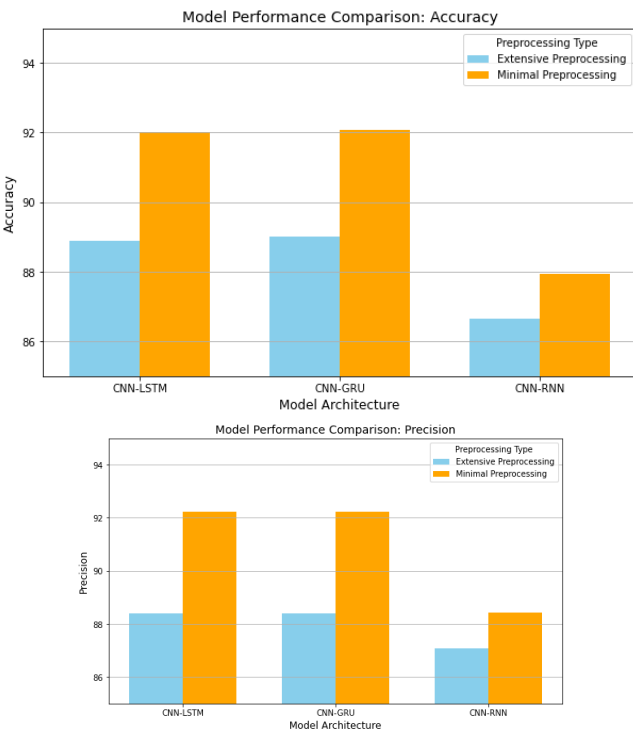
Fig. 2. Models comparisons.

TABLE IV. MODELS RESULTS

Model	Preprocessing	Max Words	Accuracy	Precision	F1	Recall
CNN+LSTM	Few	10,000	92.13%	92.77%	92.07%	91.38%
CNN+RNN	Few	10,000	88.44%	87.45%	88.59%	89.76%
CNN+GRU	Few	10,000	91.75%	92.56%	91.67%	90.80%
CNN+GRU	Few	200,000	92.08%	92.22%	92.07%	91.19%
CNN+LSTM	Few	200,000	92%	92.22%	91.98%	91.73%
CNN+RNN	Few	200,000	87.93%	88.43%	87.85%	87.28%
CNN+GRU	Extensive	10,000	88.67%	89.63%	88.53%	87.46%
CNN+LSTM	Extensive	10,000	88.72%	88.84%	88.76%	89.07%
CNN+RNN	Extensive	10,000	86.36%	86.73%	86.29%	85.85%
CNN+GRU	Extensive	200,000	89%	88.39%	88.75%	89.12%
CNN+LSTM	Extensive	200,000	88.88%	88.39%	88.95%	89.52%
CNN+RNN	Extensive	200,000	86.64%	87.07%	86.56%	86.05%

V. DISCUSSIONS

The experimental results revealed that hybrid deep learning models are effective in spam review detection, specifically for CNN-LSTM and CNN-GRU architectures. A simple CNN-LSTM model with very little preprocessing, and vocabulary of 200,000 outperformed others on a consistent basis. This shows that the model is capable of capturing rival patterns locally and the long-term dependence as observed by Bhuvaneshwari et al. Sequential models like LSTM: LSTMs have proven to be the backbone of text classification in various tasks Bhuvaneshwari et al. [5]. CNN-GRU model also provided competitive



performance, which further confirms that the combined structures are proven efficient in these applications.

Our hybrid deep learning models have performed significantly higher than the traditional machine learning approaches (Support Vector Machines (SVM) and Naive Bayes (NB)). Studies, for example Etaiwi and Naymat [19], highlighted SVM performance with smaller datasets like DOSC, achieving high accuracies (even up to 85.5%). On the much larger dataset of the Amazon Product Review, our CNN-LSTM model reached 92% accuracy, while ML models underperformed dramatically. This large increase demonstrates that the merits of deep learning to effectively learn complex patterns in large-scale data.

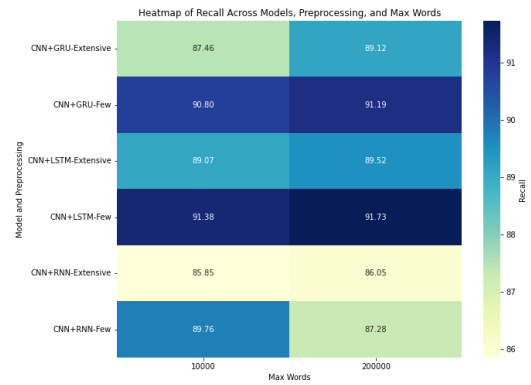
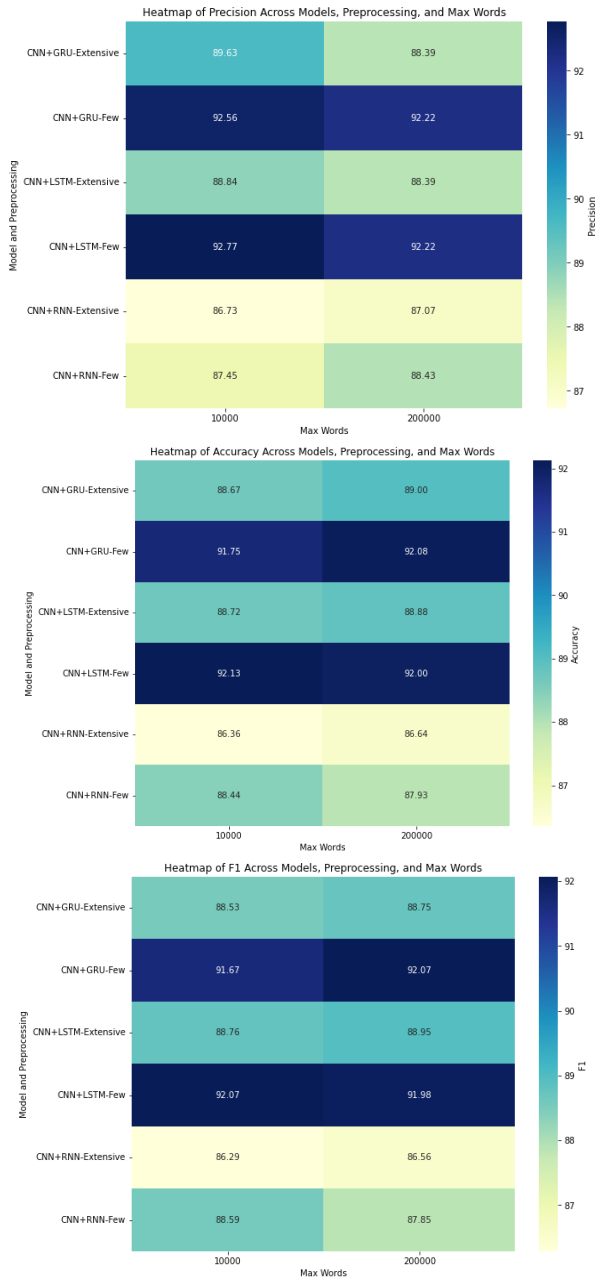


Fig. 3. Models Comparisons in terms of heatmap by different parameter sizes.

Fig. 3 represented the heatmaps for the performance of different models (CNN+LSTM, CNN+GRU, CNN+RNN) across key metrics: Accuracy, Precision, F1, and Recall. The heatmaps provide a clear comparison for the proposed hybrid system based on preprocessing type and vocabulary size, highlighting the effectiveness of minimal preprocessing and large vocabulary sizes.

Also, it is in accordance with the results of Ghourabi et al. [14] CNN text classification which has shown that CNN-LSTM hybrid could successfully learn n-gram features as well as long-term dependencies. However, they performed their study from a smaller dataset with a restricted size of vocabulary. This study used larger vocabularies than used in the previous study to train the models, and our results appear to expand on these findings, suggesting that larger vocabularies produce even better models, at least on some data sets, this further boost in size compares to a level of detail in relationships examinable within the text.

Fig. 4 bar chart highlights that CNN+LSTM with minimal preprocessing and a large vocabulary size (200k) achieves the highest accuracy, followed closely by CNN+GRU under similar conditions. Models with extensive preprocessing generally show lower accuracy. An interesting takeaway is the effect of preprocessing on model performance. As also mentioned by HaCohen-Kerner et al. [29] among others, aggressive preprocessing like lemmatization and stop word removal limited the models' ability to pick up any useful signals. On the other hand, less preprocessing helps the models keep the richness of the original text, which performed better. This is an important finding because it counters the widespread assumption that more preprocessing is always a good thing for model performance.

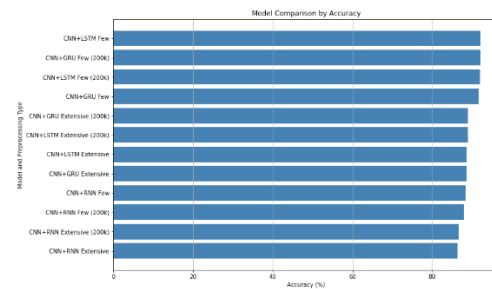


Fig. 4. A bar chart comparing the accuracy of different models and preprocessing configurations.

CNN-LSTM model has shown the best performance with less preprocessing steps and using large vocabulary size thus providing a very practical way for spam review detection. The benefit of preserving its original text is that this model can utilize its full potential to find hard to detect spam. This method is especially beneficial when the datasets are large, and context retention is essential for efficient classification. Moreover, this study shows that the proposed model is deployable efficiently under limited resources on cloud platforms such as Google Colab Pro.

Despite these optimistic results, there were several limitations to the study. However, the need for labeled datasets presents a critical issue because manual labeling is often a laborious, inconsistent process. Semi-supervised or Active Learning based techniques may also be a direction for future research to automate the labeling process as well as further explore the state-of-the-art discussed in the previous section for a better initial core for semi-supervised learning. For even better performance, pretrained models like BERT can be utilized, as they are able to learn much more complex relationships, as shown recently in the field of NLP.

Therefore, this study offers a solid foundation for hybrid deep learning models to detect spam reviews. The solution proposed not just increases detection accuracy but also provides a scalable approach with the use of dataset used in e-commerce enabling a better legitimate and deterring e-commerce platforms.

VI. CONCLUSION AND FUTURE WORK

This research has underlined the importance of combining many deep learning architectures to achieve optimal results in detecting spam reviews. It showed the capabilities of CNN and the sequential learning strength of LSTM and GRU. This contribution will help e-commerce platforms to build consumer trust by detecting spam reviews effectively. Another superior contribution of this paper is the impact of preprocessing steps on hybrid DL models' performance. Interestingly, when eliminating some of preprocessing steps the models performed better than those trained with all preprocessing steps. The combination of large datasets with hybrid DL models showed promising results in spam detection. However, the study identified a key limitation, the need for new labeled datasets for online spam reviews. As spamming techniques evolve, addressing this limitation in future work will encourage researchers to keep datasets updated for recent spam behaviors. Also, exploring recent ML techniques to automate the task of labeling the datasets is important. Methods such as semi-supervised learning or active learning could be implemented to get accurate datasets, reducing the dependence on manually labeled datasets. Furthermore, empiric hyperparameters and optimizers may further improve the performance of the models. Finally, the findings of this study indicate that the CNN-LSTM model using 200,000 max words outperformed other Hybrid DL models with an accuracy of 92%, a precision of 92.22%, a recall of 91.73%, and an F1Score of 91.98%.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University for funding and supporting this work through Graduate Students Research Support Program.

REFERENCES

- [1] H. A. Najada and X. Zhu, "iSRD: Spam review detection with imbalanced data distributions," in Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014), Redwood City, CA, USA: IEEE, Aug. 2014, pp. 553–560. doi: 10.1109/IRI.2014.7051938.
- [2] Department for Business and Trade (DBT), "FAKE ONLINE REVIEWS RESEARCH," UK Government, London, UK, 2023. Accessed: Oct. 12, 2024. [Online]. Available: <https://assets.publishing.service.gov.uk/media/6447c00c529eda000c3b03c5/fake-online-reviews-research.pdf>
- [3] Y. Li, Y. Liu, and C. Liu, "Research on Spam Review Detection: A Survey," in 2023 19th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), Harbin, China: IEEE, Jul. 2023, pp. 1–6. doi: 10.1109/ICNC-FSKD59587.2023.10281054.
- [4] Q. Li, Q. Wu, C. Zhu, J. Zhang, and W. Zhao, "Unsupervised User Behavior Representation for Fraud Review Detection with Cold-Start Problem," in Advances in Knowledge Discovery and Data Mining, vol. 11439, Q. Yang, Z.-H. Zhou, Z. Gong, M.-L. Zhang, and S.-J. Huang, Eds., in Lecture Notes in Computer Science, vol. 11439, Cham: Springer International Publishing, 2019, pp. 222–236. doi: 10.1007/978-3-030-16148-4_18.
- [5] P. Bhuvaneswari, A. N. Rao, and Y. H. Robinson, "Spam review detection using self attention based CNN and bi-directional LSTM," *Multimed. Tools Appl.*, vol. 80, no. 12, pp. 18107–18124, May 2021, doi: 10.1007/s11042-021-10602-y.
- [6] M. N. Rizali, M. M. Rosli, and N. A. S. Abdullah, "Spam Review Detection in E-Commerce Using Machine Learning," in 2024 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET), Kota Kinabalu, Malaysia: IEEE, Aug. 2024, pp. 189–193. doi: 10.1109/IICAIET62352.2024.10730100.
- [7] N. Hussain, H. Turab Mirza, I. Hussain, F. Iqbal, and I. Memon, "Spam Review Detection Using the Linguistic and Spammer Behavioral Methods," *IEEE Access*, vol. 8, pp. 53801–53816, 2020, doi: 10.1109/ACCESS.2020.2979226.
- [8] A. Jakupov, J. Longhi, and B. Zeddini, "The Language of Deception: Applying Findings on Opinion Spam to Legal and Forensic Discourses," *Languages*, vol. 9, no. 1, p. 10, Dec. 2023, doi: 10.3390/languages9010010.
- [9] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting Burstiness in Reviews for Review Spammer Detection," *Proc. Int. AAAI Conf. Web Soc. Media*, vol. 7, no. 1, pp. 175–184, Aug. 2021, doi: 10.1609/icwsm.v7i1.14400.
- [10] M. S. Farooq, "Spam Review Detection: A Systematic Literature Review," Sep. 17, 2020. doi: 10.36227/techrxiv.12951077.v1.
- [11] M. Ennaouri and A. Zellou, "Machine Learning Approaches for Fake Reviews Detection: A Systematic Literature Review," *J. Web Eng.*, Dec. 2023, doi: 10.13052/jwe1540-9589.2254.
- [12] S. Saumya and J. P. Singh, "Detection of spam reviews: a sentiment analysis approach," *CSI Trans. ICT*, vol. 6, no. 2, pp. 137–148, Jun. 2018, doi: 10.1007/s40012-018-0193-0.
- [13] P. Kalaivani, V. D. Raj, R. Madhavan, and A. P. Naveen Kumar, "Fake Review Detection using Naive Bayesian Classifier," in 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS), Coimbatore, India: IEEE, Jun. 2023, pp. 705–709. doi: 10.1109/ICSCSS57650.2023.10169838.
- [14] A. Ghourabi, M. A. Mahmood, and Q. M. Alzubi, "A Hybrid CNN-LSTM Model for SMS Spam Detection in Arabic and English Messages," *Future Internet*, vol. 12, no. 9, p. 156, Sep. 2020, doi: 10.3390/fi12090156.

- [15] "Deep Learning Hybrid Approaches to Detect Fake Reviews and Ratings," *J. Sci. Ind. Res.*, vol. 82, no. 01, Jan. 2023, doi: 10.56042/jsir.v82i1.69937.
- [16] G. M. Shahariar, S. Biswas, F. Omar, F. M. Shah, and S. B. Hassan, "Spam Review Detection Using Deep Learning," in 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Oct. 2019, pp. 0027–0033. doi: 10.1109/IEMCON.2019.8936148.
- [17] M. N. I. Ahsan, T. Nahian, A. A. Kafi, Md. I. Hossain, and F. M. Shah, "An ensemble approach to detect review spam using hybrid machine learning technique," in 2016 19th International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh: IEEE, Dec. 2016, pp. 388–394. doi: 10.1109/ICCITECHN.2016.7860229.
- [18] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of sentiment reviews using n-gram machine learning approach," *Expert Syst. Appl.*, vol. 57, pp. 117–126, Sep. 2016, doi: 10.1016/j.eswa.2016.03.028.
- [19] W. Etaiwi and G. Naymat, "The Impact of applying Different Preprocessing Steps on Review Spam Detection," *Procedia Comput. Sci.*, vol. 113, pp. 273–279, 2017, doi: 10.1016/j.procs.2017.08.368.
- [20] R. M. K. Saeed, S. Rady, and T. F. Gharib, "An ensemble approach for spam detection in Arabic opinion texts," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 1, pp. 1407–1416, Jan. 2022, doi: 10.1016/j.jksuci.2019.10.002.
- [21] A. J. Ibrahim, M. M. Siraj, and M. M. Din, "Ensemble classifiers for spam review detection," in 2017 IEEE Conference on Application, Information and Network Security (AINS), Miri: IEEE, Nov. 2017, pp. 130–134. doi: 10.1109/AINS.2017.8270437.
- [22] W. Etaiwi and A. Awajan, "The Effects of Features Selection Methods on Spam Review Detection Performance," in 2017 International Conference on New Trends in Computing Sciences (ICTCS), Amman: IEEE, Oct. 2017, pp. 116–120. doi: 10.1109/ICTCS.2017.50.
- [23] S. Zhao, Z. Xu, L. Liu, and M. Guo, "Towards Accurate Deceptive Opinion Spam Detection based on Word Order-preserving CNN," *Mar. 19, 2018*, arXiv: arXiv:1711.09181. Accessed: Oct. 19, 2024. [Online]. Available: <http://arxiv.org/abs/1711.09181>
- [24] T. A. Almeida, T. P. Silva, I. Santos, and J. M. Gómez Hidalgo, "Text normalization and semantic indexing to enhance Instant Messaging and SMS spam filtering," *Knowl.-Based Syst.*, vol. 108, pp. 25–32, Sep. 2016, doi: 10.1016/j.knosys.2016.05.001.
- [25] D. Singh, M. Memoria, and R. Kumar, "Deep Learning Based Model for Fake Review Detection," in 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT), Gharuan, India: IEEE, May 2023, pp. 92–95. doi: 10.1109/InCACCT57535.2023.10141826.
- [26] S. Saumya and J. P. Singh, "Spam review detection using LSTM autoencoder: an unsupervised approach," *Electron. Commer. Res.*, vol. 22, no. 1, pp. 113–133, Mar. 2022, doi: 10.1007/s10660-020-09413-4.
- [27] N. Hussain, H. Turab Mirza, G. Rasool, I. Hussain, and M. Kaleem, "Spam Review Detection Techniques: A Systematic Literature Review," *Appl. Sci.*, vol. 9, no. 5, p. 987, Mar. 2019, doi: 10.3390/app9050987.
- [28] G. Wayal and V. Bhandari, "Enhancing Review Spam Detection with a Hybrid Approach Integrating Association Rule Mining and Convolutional Neural Networks," in 2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET), Indore, India: IEEE, Sep. 2024, pp. 1–8. doi: 10.1109/ACROSET62108.2024.10743414.
- [29] Y. HaCohen-Kerner, D. Miller, and Y. Yigal, "The influence of preprocessing on text classification using a bag-of-words representation," *PLOS ONE*, vol. 15, no. 5, p. e0232525, May 2020, doi: 10.1371/journal.pone.0232525.
- [30] T. Limisiewicz, J. Balhar, and D. Mareček, "Tokenization Impacts Multilingual Language Modeling: Assessing Vocabulary Allocation and Overlap Across Languages".
- [31] A. Tegene, Q. Liu, Y. Gan, T. Dai, H. Leka, and M. Ayenew, "Deep Learning and Embedding Based Latent Factor Model for Collaborative Recommender Systems," *Appl. Sci.*, vol. 13, no. 2, p. 726, Jan. 2023, doi: 10.3390/app13020726.
- [32] C. Zhou, C. Sun, Z. Liu, and F. C. M. Lau, "A C-LSTM Neural Network for Text Classification," 2015, arXiv. doi: 10.48550/ARXIV.1511.08630.
- [33] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," 2015, doi: 10.48550/ARXIV.1503.04069.
- [34] K. Cho et al., "Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation," 2014, arXiv. doi: 10.48550/ARXIV.1406.1078.
- [35] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Inf. Process. Manag.*, vol. 45, no. 4, pp. 427–437, Jul. 2009, doi: 10.1016/j.ipm.2009.03.002.
- [36] J. Davis and M. Goadrich, "The relationship between Precision-Recall and ROC curves," in Proceedings of the 23rd international conference on Machine learning - ICML '06, Pittsburgh, Pennsylvania: ACM Press, 2006, pp. 233–240. doi: 10.1145/1143844.1143874.
- [37] T. Saito and M. Rehmsmeier, "The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets," *PLOS ONE*, vol. 10, no. 3, p. e0118432, Mar. 2015, doi: 10.1371/journal.pone.0118432.
- [38] S. Sagnika, B. S. P. Mishra, and S. K. Meher, "An attention-based CNN-LSTM model for subjectivity detection in opinion-mining," *Neural Comput. Appl.*, vol. 33, no. 24, pp. 17425–17438, Dec. 2021, doi: 10.1007/s00521-021-06328-5.

Optimization of Fourth Party Logistics Routing Considering Infection Risk and Delay Risk

Guihua Bo¹, Sijia Li^{2,*}, Mingqiang Yin³, Mingkun Chen⁴, and Xin Liu⁵

School of Information and Control Engineering, Liaoning Petrochemical University, Fushun, China^{1, 2, 3, 4, 5}

Abstract—In the context of the rapid development of e-commerce and the increasing demands for logistics services, particularly in the face of challenges posed by public health emergencies, this paper explores how to integrate supply chain resources and optimize delivery processes. It provides an in-depth analysis of the characteristics of the Fourth Party Logistics Routing Optimization Problem (4PLROP) in complex environments, specifically focusing on the impacts of infection risk and delay risk, and proposes a new risk measurement tool. By constructing a mathematical model aimed at minimizing Conditional Value-at-Risk (CVaR) and improved Q-learning algorithm, the study addresses the 4PLROP while considering cost and risk constraints. This approach enhances the efficiency and service quality of the logistics industry, offers effective strategies for 4PL companies in the face of uncertainty, and provides customers with safer and more reliable logistics solutions, contributing to sustainable development.

Keywords—Logistics services; public health emergencies; logistics routing optimization; improved Q-learning algorithm; CVaR; infection risk

I. INTRODUCTION

As the limitations of Third Party Logistics (3PL) capabilities become increasingly apparent, the traditional route planning problem is transitioning into the more complex Fourth Party Logistics Routing Optimization Problem (4PLROP). 4PL, known as the integrator of supply chains, consolidates its own resources, capabilities, and technologies, as well as those of other 3PL providers, to offer comprehensive supply chain solutions to clients. The concept of 4PL has garnered widespread attention from both the industry and academia since its inception. Presently, 4PL enterprises or platforms such as UPS, Cainiao, and Ningbo Fourth Party Logistics Market have established long-term cooperative relationships with manufacturing enterprises like Haier, providing them with specialized logistics and transportation services. In the academic sphere, numerous scholars have conducted research on various aspects of 4PL, including route optimization [1], network design [2], risk management [3], combinatorial auctions [4]-[5], supply chain integration [6], and information technology application [7]. Route optimization in 4PL, as one of the core issues at the tactical layer of 4PL operation and management, has been receiving considerable scholarly attention in recent years. It integrates 3PL and route selection decisions to enhance the efficiency of logistics transportation [8].

In the advent of public health emergencies, the logistics industry has played a pivotal role in ensuring resource supply, yet such occurrences also introduce new challenges and risks to

route planning. During the 2020 pandemic, international and Hong Kong, Macau, and Taiwan air passenger volumes plummeted by over 15%, particularly in key logistics hubs like Wuhan, where containment measures significantly impacted logistics routes. To circumvent high-risk pandemic areas, logistics enterprises were compelled to rechart transportation routes, which not only heightened the complexity of route planning but also augmented the time required. These shifts demand that logistics enterprises factor in the risks posed by pandemics during the planning process. 4PL must consider the risks brought about by pandemics and other public health emergencies, employing agile supply chain management and efficient resource allocation to mitigate the impact of these risks on the logistics network. In 4PL, such risk management is especially crucial. As the coordinator and integrator within the supply chain, 4PL is tasked with managing the demands and risks of multiple clients while ensuring service quality.

This article delves into the 4PLROP, which takes into account the risks of infection and delays, against the backdrop of the pandemic. Cities are categorized based on risk levels, and the infection risks of various cities are assessed using quantitative methods. Leveraging the extensive application of Conditional Value-at-Risk (CVaR) in the optimization domain, a mathematical model is established with the objective of minimizing CVaR, and constraints are set for delivery costs and infection risks. An improved Q-learning algorithm is employed to solve this model, aiming to identify the route that satisfies the conditions of the lowest CVaR for both infection risks and delivery costs. In this manner, 4PL can better adapt to the ever-changing market environment, ensuring the stability and efficiency of the supply chain.

The principal contributions of this article are as follows:

- In the context of public health emergencies, a novel 4PL route optimization problem that concurrently considers the risks of infection and delays has been investigated;
- The CVaR metric is employed to characterize risks, and a nonlinear programming model is established with constraints on delivery costs and infection risks, aiming to minimize the risk as the objective;
- An improved Q-learning algorithm is proposed to solve the model presented. Through this approach, 4PL can better adapt to the ever-changing market environment, ensuring the stability and efficiency of the supply chain.

The remainder of this article is structured as follows:

The remainder of this article is structured as follows: Section II provides a review of literature pertaining to 4PL route-related studies. Section III delineates the problem and elucidates the model and notation employed. Section IV applies the Q-learning algorithm to the 4PL route optimization problem, where the optimal path planning is achieved through the establishment of action-state pairs, construction of a reward function, enhancement of exploration strategies, and model training. Section V validates the efficacy of the improved Q-learning algorithm in the context of 4PL route optimization through experimental analysis, demonstrating the algorithm's high solution speed and stability across various scales of test cases, and its ability to provide customers with delivery routes that minimize risk at a specific confidence level. Section VI presents our conclusions and prospective directions for future research.

II. LITERATURE REVIEW

In the context of public health emergencies, a plethora of theoretical foundations and practical case studies has been provided by existing research to address the 4PLROP. Within this section, an exhaustive review of the pertinent literature on 4PLROP has been conducted. The existing research delineates the complexities and challenges posed by the emergence of public health crises on 4PLROP, highlighting the need for innovative approaches to mitigate the associated risks and delays.

In the field of 4PLROP, a relatively early study dates back to 1998. The concept of 4PL was initially introduced by Andersen Consulting [9]. Since then, based on this concept, a plethora of research on 4PLROP has been conducted: Huang et al. [10] conducted research on the 4PLROP with uncertain delivery time in emergencies. Huang et al. [11] proposed an improved genetic algorithm based on simple graphs and the Dijkstra algorithm to preclude the emergence of infeasible solutions in 4PLROP. Ren et al. [12] designed a genetic algorithm embedded with the Dijkstra algorithm to solve the 4PLROP problem, thereby laying the foundation for the research on 4PLROP. The existing studies can generally be categorized into two types. The first type is the 4PL route planning problem in a deterministic environment, and the second type is the 4PLROP in an uncertain environment.

In the realm of deterministic environment, an early scholar established a directed graph model to optimize the selection of routes, transportation modes, and third-party logistics providers [13]. Subsequently, a 4PL optimization model was constructed by another scholar to streamline the corresponding 4PLROP [14]. Thereafter, a 4PLROP approach based on the immune algorithm was proposed by some scholars, enhancing the algorithm's capability to address 4PLROP [15]. Building on this foundation, a mathematical model for point-to-point multi-task 4PLROP without edge repetition was established by other scholars, considering the cost and time attributes of each node and edge, and an ant colony optimization algorithm was designed to solve the path optimization problem [16]. Recently, Zhou et al. [17] addressed the 4PLROP problem considering cost discounts by minimizing operational costs, taking into account customer delivery deadlines and transportation capacity constraints. Cai et al. [15] minimized the linear

combination of transportation and time costs by considering certain customer preference factors.

In the realm of uncertainty, Huang et al. [18] proposed an uncertain programming model for the 4PLROP in emergency situations. The effectiveness of this model was verified through comparison with the stochastic programming model and numerical experiments. Huang et al. [19] transformed the uncertainty theory into a deterministic model and designed an improved genetic algorithm for solving to address the 4PLROP in an uncertain environment. Lu et al. [20] solved the uncertain delivery time control model of 4PLROP through the genetic algorithm. Lu et al. [21] dealt with the 4PLROP problem under the conditions of uncertainty in 3PL transportation time, transportation cost, node transfer time and transfer cost, and designed a solution model using the grey wolf optimization algorithm improved by the ant colony system. Lu et al. [22] considered the uncertainties in transportation time and cost caused by seasonal and human factors, constructed a multi-objective chance-constrained programming model aiming to minimize transportation time and cost, and proposed a hybrid beetle swarm optimization algorithm combined with the Dijkstra algorithm to solve the problem. Ren et al. [23] established a 4PLROP chance model with time windows and random transportation time under the constraint of total transportation cost, aiming to maximize the chance that the total transportation time meets the time windows. And the ant colony algorithm was used to solve the deterministic model. Gao et al. [24] applied the uncertain stochastic programming model to solve the 4PLROP with random demand and uncertain transportation and transshipment times, aiming to minimize the total transportation cost under various constraints. Recently, Ren et al. [25] aimed to examine the impact of decision-makers' risk preferences on the 4PLROP, contributing to the analysis of logistics behavior and route integration optimization in uncertain environments.

In addition to the aforementioned studies, recent years have witnessed a growing focus on the risk factors in 4PLROP. Deng et al. [26] utilized the ant-colony algorithm to address the mathematical model of 4PLROP, where the Value at Risk (VaR) was employed to represent the delay risk in an uncertain environment. Bo et al. [27] carried out research on the 4PLROP with tardiness risk by introducing VaR to measure the time-related risk. Wang et al. [28] established a mathematical model considering customers' risk-averse behavior and studied the 4PLROP in the context of customers' risk-avoidance behavior. Recently, Liu et al. [29] introduced the risk value VaR to measure the risk of delays, which has been a significant advancement in the risk assessment of 4PLROP.

The probability that the delay quantity is less than a certain value, as denoted by VaR, is required to be greater than or equal to the confidence level prescribed by the client. However, it merely takes into account the likelihood of the occurrence of delay risks, without considering the mean of such risks when they materialize under extreme conditions. The conditional mean of delay risks exceeding VaR can be determined by the CVaR model. By integrating the risk level of the distribution plan and the anticipated delay risks, rational decisions concerning distribution services can be formulated.

In summary, while the existing literature encompasses a multitude of aspects of 4PLROP, there is still a deficiency in addressing the infection risks and delays associated with public health emergencies. This article provides a novel perspective and practical methodologies for this field of research by constructing corresponding mathematical models based on CVaR and employing an improved Q-learning algorithm. The aim is to assist 4PL systems in better adapting to the ever-changing market environment, ensuring the stability and efficiency of the supply chain.

III. PROBLEM DESCRIPTION AND MODEL ASSUMPTIONS

A. The Path Optimization Problem in the Context of the Pandemic

In environments where logistics warehouses and transfer node cities are densely populated with personnel and abundant goods, the potential risk of virus transmission during the pandemic cannot be overlooked. Especially in the context of ongoing pandemic prevention and control, 4PL service providers must ensure that only goods, and not viruses, are transported by vehicles while maintaining smooth logistics operations. Consequently, stringent monitoring of infection risks in logistics transportation has become a crucial component of epidemic prevention efforts.

During the special period of the pandemic, to avoid potential infections in high-risk areas, this paper has developed a specific planning strategy for delivery routes, incorporating infection risk constraints. This issue can be specifically described as follows: as graphical illustration of the process in Fig. 1, a multi-layer graph "G=(V,E)" is used to represent the 4PLROP, where " $|V|=n$ " is the set of node cities and E is the set of edges. The node city s represents the supply city, node city t represents the destination city, and other node cities representing transfer node cities. The number of node cities " $|V|=n$ " indicates the total number of node cities, each of which has attributes such as time, cost, carrying capacity, and reputation. Since there may be multiple 3PL suppliers offering services between any two node cities, there are multiple edges between any two node cities in the graph (each edge represents a different 3PL, identified by a unique number). Consequently, each edge has different attributes related to time, cost, capacity, and reputation, meaning that each 3PL has its corresponding properties. Therefore, when selecting transfer node cities and 3PL suppliers, it is necessary to comprehensively weigh various factors to ensure that the chosen path is not only cost-effective and time-efficient but also minimizes infection risks to the greatest extent possible.

The goal is to provide customers with a delivery plan that meets cost budget and infection risk control requirements while minimizing CVaR, ensuring that goods are delivered safely and on time. Therefore, the following assumptions are proposed to address the aforementioned research issues:

- (1): It is assumed that infection risks only exist at node cities where 3PL suppliers are changed and where handling and unloading occur, while the transportation between node cities is considered risk-free.

- (2): It is assumed that the level of infection risk is directly related to the cumulative number of locally confirmed cases in the city over the past 14 days, and that high-risk node cities are strictly avoided. The specific risk assessment method will be detailed in Section III (C).

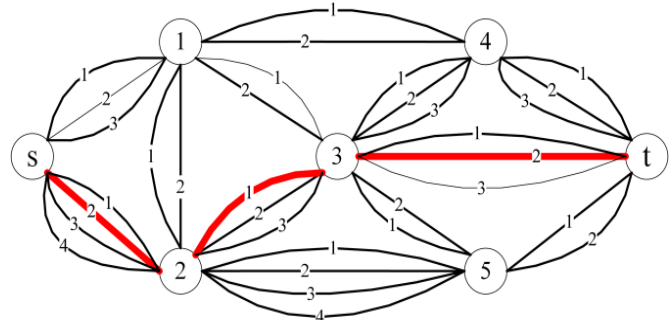


Fig. 1. 7-node problem description.

B. Parameters and Variables

By defining the following parameters and variables to establish a mathematical model, as shown in Table I.

TABLE I. DEFINITIONS OF PARAMETERS AND VARIABLES

Symbol	Definition description
r_{ij}	The number of 3PL providers that can offer delivery services between the node city i and node city j (namely, the number of edges between the two node cities).
C_{ijk}	The transportation cost required by the k -th 3PL provider for delivery services between the node city i and node city j .
T_{ijk}	The random transportation time required by the k -th 3PL provider for services between the node city i and node city j .
C'_j	The transshipment cost required when passing through node city j .
T'_j	The random transshipment time required when passing through node city j .
R	The set of node cities and edges contained in the path is, namely, $R=\{v_s, \dots, v_i, k, v_j, \dots, v_t\}$. As shown in Fig. 1, the red path can be represented by $R=\{v_s, 2, v_2, 1, v_3, 2, v_t\}$.
$x_{ijk}(R)$	Decision variable. When the 3PL provider represented by the k -th edge between city i and node city j provides the distribution task, it takes 1; otherwise, it takes 0. As shown in (1).
$y_j(R)$	Decision variable. If the city represented by node city j provides the transshipment task, it takes the value of 1; otherwise, it takes 0. As shown in (2).
X_j	The cumulative number of local confirmed cases within 14 days in node city j .
f	The unit person-time infection risk of the cumulative number of local confirmed cases within 14 days.
F_0	The maximum acceptable infection risk for customers.

$$x_{ijk} = \begin{cases} 1, & \text{The } k\text{-th edge between } i \text{ and } j \\ & \text{belong to path } R \\ 0, & \text{else} \end{cases} \quad (1)$$

$$y_j(R) = \begin{cases} 1, & \text{node city } j \text{ belong to path } R \\ 0, & \text{else} \end{cases} \quad (2)$$

C. Quantification of the Infection Risk

This paper studies the 4PLROP during the early stages of the pandemic. Therefore, it draws on the classification of cities into low, medium, and high-risk areas established in the early phase of the pandemic to assign infection risk values to the node cities, as shown in Table II.

TABLE II. RISK CLASSIFICATION CRITERIA

Risk rating	Classification criterion	Response policies
Low-risk area	Within 14 days, without any new or existing confirmed cases.	Strengthen external prevention and control, fully restart production and daily life, and lift road traffic restrictions
Medium-risk area	Within 14 days, if the number of new confirmed cases is ≤ 50 or there are no cluster outbreaks, even if the cumulative confirmed cases exceed 50.	Implement a dual prevention and control strategy, steadily and orderly restore the normal state of production and daily life
High-risk area	Cumulative cases exceed 50, and there have been cluster outbreaks in the past 14 days.	Implement strict management with dual-direction prevention and control, ensuring that the pandemic does not spread or overflow

According to the defined standards, when assessing the COVID-19 infection risk in a certain area, the number of locally confirmed cases over a continuous fourteen-day period is considered. If an area has no locally confirmed cases or has no new cases for fourteen consecutive days, it is regarded as low-risk. If there are new cases within fourteen days but the cumulative confirmed cases do not exceed 50, it is classified as a medium-risk area. When the cumulative confirmed cases exceed 50, the area is considered high-risk. Given that the incubation period of the COVID-19 virus is fourteen days and that travel codes also reference the travel history within the last fourteen days, this paper uses the number of locally confirmed cases in a node city over the past fourteen days as the basis. The infection risk f is quantified in terms of the number of confirmed cases per person. For example, if the cumulative confirmed cases in a node city within fourteen days amount to 25, then the infection risk is $25f$. Moreover, infection risk occurs only during the transfer at the node city.

D. Mathematical Model

Under the confidence level β , when minimizing CVaR, calculate the distribution path with the minimum average overdue risk. Add the infection risk constraint and establish the following mathematical model:

$$\begin{aligned} & \min(\sum_{i=1}^n \sum_{j=1}^n \sum_k^{r_{ij}} \mu_{ijk} x_{ijk}(R) + \sum_{j=1}^n \mu_j y_j - T_0) + c_1(\beta) \\ & \times \sqrt{\sum_{i=1}^n \sum_{j=1}^n \sum_k^{r_{ij}} \delta_{ijk}^2 x_{ijk}^2(R) + \sum_{j=1}^n \delta_j^2 y_j^2(R)} \quad (3) \\ \text{s.t.} \quad & \sum_{j=1}^n X_j f y_j \leq F_0 \quad (4) \end{aligned}$$

$$\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^{r_{ij}} C_{ijk} x_{ijk}(R) + \sum_{j=1}^n C'_j y_j(R) \leq C_0 \quad (5)$$

$$\Delta T = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^{r_{ij}} T_{ijk} x_{ijk}(R) + \sum_{j=1}^n T'_j y_j(R) - T_0 \quad (6)$$

$$R = \{v_s, \dots, v_i, k, v_j, \dots, v_k\} \in G \quad (7)$$

$$x_{ijk}(R), y_j(R) \in \{0, 1\} \quad (8)$$

$$X_j < 50 \quad (9)$$

Among them, Constraints (4) the capacity limit for the infection risk, F_0 denotes the maximum acceptable infection risk given by the customer. Constraints (5) the capacity limit for the delivery cost, where C_0 is the maximum cost acceptable to the customer. Constraints (6) is the expression of the overdue quantity ΔT , which is a random variable. Constraints (7) reflects the path to ensure that the path is a legal connected path from the initial node city to the destination city. Constraints (8) manifests $x_{ijk}(R)$ and $y_j(R)$ are decision variables. Constraints (9) conveys that the transportation path cannot pass through high-risk areas.

IV. ALGORITHM DESIGN

When the improved Q-learning algorithm is employed to address the 4PLROP, the primary procedures encompass the initialization of parameters, the establishment of action-state settings, the construction of the reward function, the formulation of exploration strategies and the training of the model.

A. Action-state Setting

This paper combines improved Q-learning with the 4PLROP, viewing the choice of actions as related to the selection of 3PL suppliers and treating node cities as different states. Let s represent the current state (the current node city). The corresponding 3PL suppliers for this node city can be represented by the action space $A = \{a_1, a_2, \dots, a_k, \dots, a_K\}, k=1, 2, \dots, K$. Each action-state pair corresponds to a Q-value.

Taking the 7-node for example, refer to Fig. 1 for the illustration, the initial node city can be regarded as the initial node city s . The node cities connected to the initial node city are node city 1 and node city 2. There are three selectable 3PL suppliers corresponding to the route from the initial node city to node city 1, labeled as 1, 2, and 3, which can be designated as actions α_1, α_2 and α_3 . Similarly, there are four selectable 3PL suppliers for the route from the initial node city to node city 2, labeled as 1, 2, 3, and 4, which can be designated as actions $\alpha_4, \alpha_5, \alpha_6, \alpha_7$. Thus, there are 7 actions available at the initial node city. When choosing actions α_1, α_2 and α_3 at the initial node city, the next state transitions to node city 1. Likewise, when choosing actions $\alpha_4, \alpha_5, \alpha_6, \alpha_7$ at the initial node city, the next state transitions to node city 2. The action sets for the other node cities can be defined in a similar manner. Using the 7-node as an example, the selected actions and their corresponding next states are shown in Table III.

TABLE III. 7-NODE PROBLEM ACTIONS AND CORRESPONDING NEXT STATES SETTINGS

The selected actions	Corresponding to the next state
A={a ₁ ,a ₂ ,a ₃ }	Transfer node city 1
A={a ₄ ,a ₅ ,a ₉ }	Transfer node city 2
A={a ₁₀ ,a ₁₁ ,a ₁₄ ,a ₁₅ ,a ₁₆ }	Transfer node city 3
A={a ₁₂ ,a ₁₃ ,a ₂₁ ,a ₂₂ ,a ₂₃ }	Transfer node city 4
A={a ₁₇ ,a ₁₈ ,a ₁₉ ,a ₂₀ ,a ₂₄ ,a ₂₅ }	Transfer node city 5
A={a ₂₆ ,a ₂₇ ,a ₃₃ }	Demand node city t

B. The Construction of the Reward Function

Since the Q-learning algorithm is based on the Markov Decision Process (MDP) model, a discrete reward and punishment function is adopted for computational convenience [30]. Given that transportation tasks cannot pass through high-risk areas, the number of infections at the transfer node cities along the path must not exceed 50. Therefore, when the number of infections in a certain city j is less than or equal to 50, the reward is 1. Conversely, when the number of infections in city j exceeds 50, the reward is -100, as depicted in Eq. (10).

$$r_0(s,a) = \begin{cases} 1 & \text{if } X_j \leq 50 \\ -100 & \text{if } X_j > 50 \end{cases} \quad (10)$$

When there is a connection between node city i and node city j and j is not the destination, the reward is 1. When there is no connection between node city i and node city j , the reward is -1. When there is a connection between node city i and node city j and j is the destination, the reward is 100, in (11) illustrates this concept.

$$f(x) = \begin{cases} 1, & i, j \text{ are connected}; j \text{ is not the end node city;} \\ -1, & i, j \text{ are not connected;} \\ 100, & i, j \text{ are connected}; j \text{ is the end node city} \end{cases} \quad (11)$$

Considering the magnitude of rewards is related to the mean and variance within the objective function, and also needs to meet certain constraints, the reward function for this issue can therefore be rewritten as shown in study (12). The parameter ω_1 is inversely proportional to the distribution cost corresponding to the selected 3PL supplier and transfer node city, that is, the smaller the distribution cost, the greater the reward value obtained, as illustrated in study (13). In a parallel manner, ω_2 is inversely proportional to the mean value of the distribution time corresponding to the selected 3PL supplier and transfer node city. When the mean value of the random time is smaller, the greater the reward value obtained, as evidenced in study (14). The parameter ω_3 is inversely related to the average distribution time associated with the selected 3PL supplier and transfer node city, When the variance is smaller, the corresponding reward value is greater, where k_1 and k_2 are the weighting coefficients of the reward function, as detailed in study (15). Additionally, ω_4 is correlated with the infection count at the node city, as elucidated in study (16).

$$r = \omega_1 r(s,a) + \omega_2 r(s,a) + \omega_3 r(s,a) + \omega_4 r_0(s,a) \quad (12)$$

$$\omega_1 = \frac{k_1}{C_{ijk} + C_j} \quad (13)$$

$$\omega_2 = \frac{k_2}{\mu_{ijk} + \mu_j} \quad (14)$$

$$\omega_3 = \frac{1 - k_1 - k_2}{\delta_{ijk}^2 + \delta_j^2} \quad (15)$$

$$\omega_4 = \frac{1}{X_j} \quad (16)$$

C. The Exploration Strategy of Improved Q-learning Algorithm

When the agent interacts with the environment to learn, it must choose known actions that maximize the reward while also ensuring that it can learn more experiences in an unknown environment, thereby laying the foundation for obtaining more cumulative rewards. Therefore, it is essential to establish an appropriate exploration strategy to achieve optimal training results. The traditional Q-learning algorithm typically employs the ϵ -greedy strategy as its exploration method.

The mathematical description of the ϵ -greedy strategy is as follows:

$$\pi(a,s) = \begin{cases} \arg \max Q(s,a) & 1 - \epsilon \\ q_{random} & \epsilon \end{cases} \quad (17)$$

Eq. (17), it can be understood as randomly selecting the selectable actions in the current state with a certain probability ϵ , and choosing the action corresponding to the maximum Q value among the current actions with a probability of $1 - \epsilon$.

When using the Q-learning algorithm to address the 4PLROP, the environment is relatively simple, and both states and actions are limited. Therefore, it is necessary to establish a corresponding reinforcement learning environment based on the characteristics of the problem. To better explore the environment, this paper adopts a random strategy more suitable for the problem to select actions. According to the reward matrix established by the reward function, it is observed that in the current state, when the reward value is -1, the two node cities are disconnected. Therefore, the exploration strategy is set to randomly select actions with reward values greater than -1 in the current state to reduce exploration time.

D. Model Training

By designing a Q-table to train the agent, each row in the Q-table represents all the states available to the agent, while each column represents the actions the agent can perform in the corresponding state. Each state in the multi-layer graph represents different node cities, and each action represents different 3PL suppliers in the multi-layer graph. Initially, all states in the Q-table are set to 0. The reward values obtained from executing different actions (selecting different suppliers) are then calculated based on the reward matrix established by the reward function, and the values of the elements in the Q-table are updated using Eq. (18). Each iteration is considered a training session for the agent. During each training session, the agent attempts to move from the initial node city to the destination node city, updating the elements in the Q-table after executing each action.

$$Q^*(s,a) \leftarrow Q(s,a) + \alpha [R(s,a,s) + \gamma \max_{a'} Q(s,a') - Q(s,a)] \quad (18)$$

E. The Flow Chart of Q-learning Algorithm

When using the improved Q-learning algorithm to solve the 4PLROP, firstly, based on the existing data, the elements in the matrix are initialized by using the reward function. Since there are multiple different 3PL suppliers between two node cities, that is, one state corresponds to multiple ones. Therefore, it is necessary to set the actions corresponding to each state, and then train and update the matrix Q through the setting of the matrix R and related parameters. Finally, the optimal path planning can be obtained based on the Q-table. The flowchart of the 4PLROP using the improved Q-learning algorithm is shown in Fig. 2. The specific steps are as follows:

Step 1: Load the known data information in MATLAB.

Step 2: Initialize the parameters γ , α and the Q-table, set the initial state and the final state, and at the same time, use the given data Eq. (12) to construct the reward matrix R.

Step 3: Set the initial state as the initial node city.

Step 4: Determine the action through the random selection strategy, that is, select a feasible 3PL supplier.

Step 5: Perform the action α (namely, select a 3PL supplier of the current node city), and then transfer to the new state s' (node city). Update the Q-table based on the reward matrix R and preset parameters.

Step 6: Determine whether s' is the final node city. If not, return to Step 4; if yes, proceed to Step 7.

Step 7: Determine whether the set number of training times has been completed. If not, return to Step 3 to continue training; if yes, proceed to Step 8.

Step 8: The training process is over and the final Q-table is output.

Step 9: Combine the Q-table to determine and output the best logistics distribution plan.

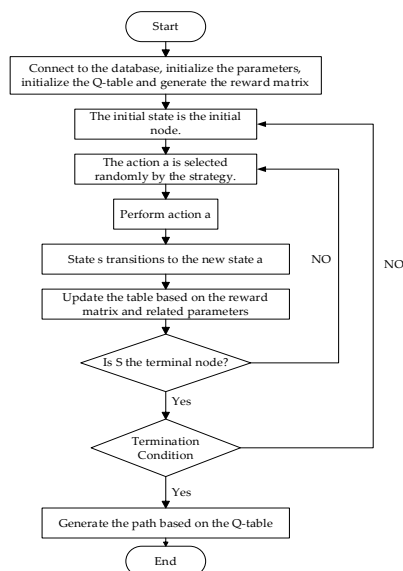


Fig. 2. Flow chart of improved Q-learning algorithm.

V. EXPERIMENTAL RESULTS AND ANALYSIS

Referring to Fig. 3, it is the epidemic data chart for some periods in 2022. Fig. 4 illustrates the cumulative number of confirmed cases in some cities across the country within 14 days. These are used as the data references for this section. The relevant data comes from the National Health Commission of China and the health commissions of various provinces and cities.

This section first takes the 7-node as an example to analyze the influence of the training times episode, discount factor γ , and learning rate α in the Q-learning algorithm on the calculation results, and obtains a set of optimal parameter combinations. Then, it solves the mathematical model with the constraint conditions of the delivery cost and the infection risk and the objective function of minimizing CVaR. Finally, the best distribution path obtained from the corresponding example is visualized. The software used by the algorithm is MATLAB 2023a, and the operating environment is Intel(R) Core(TM) i7-2600 @3.40GHz.

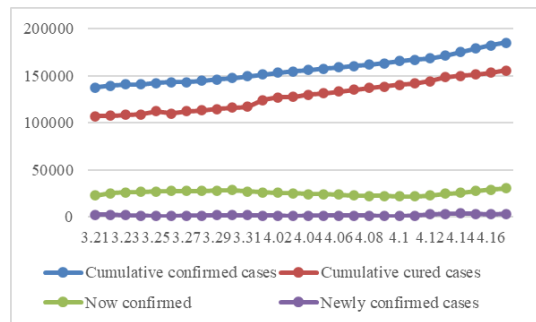


Fig. 3. Epidemic data chart for partial periods in 2022.

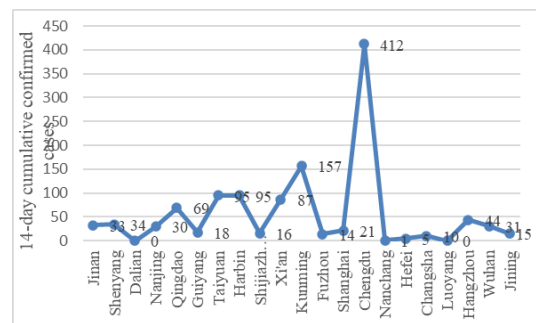


Fig. 4. Cumulative number of confirmed cases within 14 days in some cities.

A. Parameter Test

The parameters within the improved Q-learning algorithm are rigorously tested through extensive experimental simulations. This is achieved by keeping all other parameters constant and observing the impact of variations in a single parameter on the solution outcomes. The efficacy is denoted by the optimality rate, which represents the probability of obtaining the best solution during the execution of the algorithm.

Through repeated experiments on $k1$ and $k2$ in the reward function in different sized examples, when the values of $k1$ and $k2$ are the data in Table IV, the algorithm has the best solution effect.

TABLE IV. PARAMETER SETTINGS FOR DIFFERENT INSTANCES

Number of node	k1	k2	episode	CVaR	Best path	Time
7	0.1	0.8	100	27.789 2	$R=\{v_s, 2, v_2, 2, v_3, 1, v_t\}$	0.9s
15	0.1	0.2	200	12.158 9	$R=\{v_s, 3, v_3, 2, v_6, 3, v_{13}, 2, v_t\}$	1s

Fig. 5-7 offer a graphical representation of the data, the parameter test process of the improved Q-learning algorithm for solving the 7-node example is presented when the confidence level is 0.9, the infection risk constraint is 8.5×10^{-5} , and the cost constraint is 80. Wherein the "best rate" refers to the probability that the best solution is achieved during the algorithm's execution, with the total number of runs set to 100. The test results show that the best parameters of the improved Q-learning algorithm are $\gamma=0.8$, $\alpha=0.9$ and episode = 100, respectively.

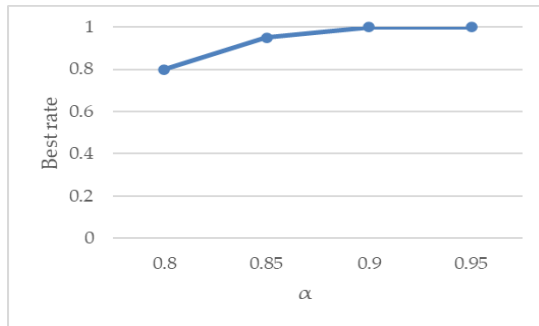


Fig. 5. Parameters α performance analysis.

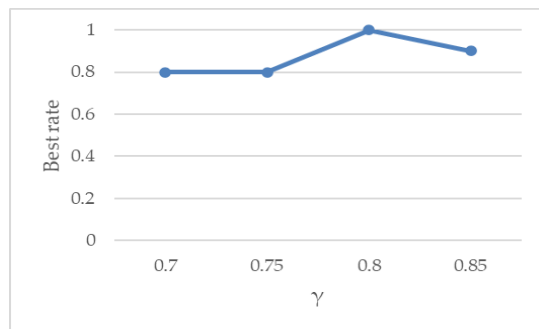


Fig. 6. Parameters γ performance analysis.

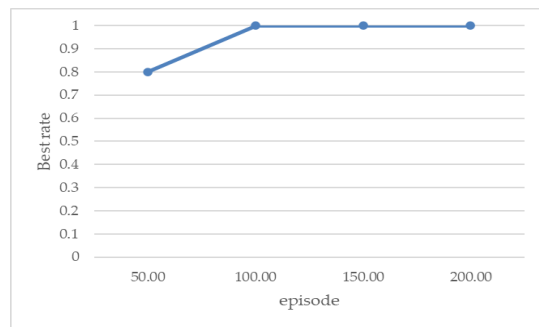


Fig. 7. Parameters episode performance analysis.

B. Case Analysis

To validate the effectiveness of the proposed model, we solved two instances of different scales, with 7-node and 15-node, and obtained the corresponding minimum CVaR values and the best routes. The information related to the 7-node and edges is shown in Table V and Table VI. Due to the large amount of information from the 3PL suppliers, only partial information is provided in Table VI.

TABLE V. 7-NODE CALCULATION EXAMPLE RELATED INFORMATION

Node	Cost	The mean of random time	The variance of random time	The number of infections
s	10	6	25	27
1	12	4	36	11
2	6	5	16	23
3	9	7	64	20
4	11	4	9	51
5	15	6	49	35
t	7	5	25	10

TABLE VI. 7-NODE CALCULATION EXAMPLE 3PL SUPPLIER RELATED INFORMATION

Initial	End	3PL Number	Transportation cost	The mean of random time	The variance of random time
s	1	1	20	12	169
s	1	2	18	15	196
s	1	3	24	10	64
s	2	1	18	16	225
s	2	2	17	17	196
s	2	3	19	14	169
s	2	4	15	20	329

As depicted in Fig. 8, it is the path diagram corresponding to the solution result of 7-node. Among them, the black pentagram s represents the initial node city, the black pentagram t is the destination node city, and the node city 4 marked by the red pentagram indicates the transfer node city that does not meet the constraint of the number of infections. That is, transfer node city 4 is in a high-risk area, so the path cannot pass through node city 4. The other transfer node cities that meet the constraint of the number of infections. The path marked by the blue thick line is the distribution path corresponding to the minimum CVaR that satisfies the constraints of cost and the infection risk, and the detailed data is shown in Table VI.

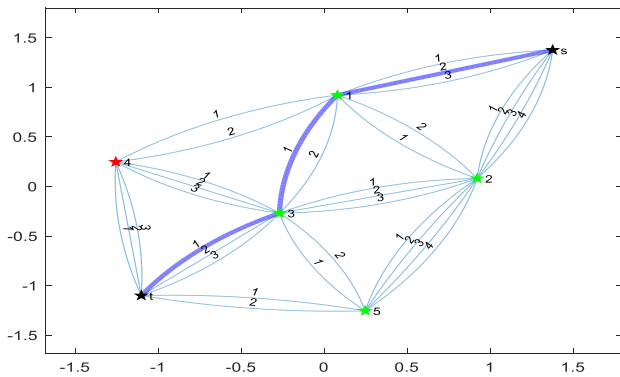


Fig. 8. 7-node solution path diagram.

The information is summarized in Table VII, the solution of the CVaR model of the 7-node problem are given under different confidence levels when the cost constraint $C_0=80$ and the maximum acceptable the infection risk given by the customer is 8.5×10^{-5} . Among them, " β " represents the confidence level, that is, the degree of risk aversion of the customer, " $CVaR$ " is the best solution obtained by the CVaR model (the evaluation criterion is the objective function), " $Best\ path$ " is the distribution path corresponding to the best solution obtained, " F " is the infection risk corresponding to the distribution path of the best solution obtained, " $Best\ rate$ " indicates the probability that the best solution obtained by the algorithm accounts for the total number of runs of the algorithm. At this time, the total number of runs is 100, and " $Time$ " is the running time of the algorithm for one run, in seconds.

TABLE VII. SOLUTION OF 7-NODE PROBLEMS AT $T_0=70$, $C_0=80$ AND $F_0=8.5 \times 10^{-5}$

β	episode	CVaR	Best path	F	Best rate	Time
0.9	100	27.7892	$R=\{v_s, 2, v_1, 1, v_3, 1, v_t\}$	7.352×10^{-5}	0.98	0.9s
0.95	100	35.1168	$R=\{v_s, 2, v_1, 1, v_3, 1, v_t\}$	7.352×10^{-5}	0.98	0.9s
0.99	100	49.4634	$R=\{v_s, 2, v_1, 1, v_3, 1, v_t\}$	7.352×10^{-5}	0.98	0.9s

According to the data in Table VII, when the confidence level is 0.9, the Delivery Cost is $C_0=80$, and the infection risk constraint is $F_0=8.5 \times 10^{-5}$, the obtained optimal CVaR value is 27.7892, indicating that the corresponding average delay risk of the distribution task is 27.7892, the corresponding distribution cost is 80, the infection risk is 7.352×10^{-5} , and the corresponding best distribution path is $R=\{v_s, 2, v_1, 1, v_3, 1, v_t\}$, indicating that when transporting from the source node city s to the destination node city t ,

the destination node city t , the selected transfer node cities are 1 and 3 respectively, and the 3PL supplier number selected between each two transfer node cities is 2, 1, 1; when the confidence level is 0.95, the cost constraint is $C_0=80$, and the infection risk constraint is $F_0=8.5 \times 10^{-5}$, the obtained minimum CVaR value is 35.1168, indicating that the corresponding average delay risk of the distribution task is 35.1168, the corresponding distribution cost is 80, the infection risk is 7.352×10^{-5} , and the corresponding best distribution path is $R=\{v_s, 2, v_1, 1, v_3, 1, v_t\}$, indicating that when transporting from the source node city s to the destination node city t , the selected transfer node cities are 1 and 3 respectively, and the 3PL supplier number selected between each two node cities is 2, 1, 1.

The relevant information of 15-node is shown in Table VIII. Since there are 91 rows of information corresponding to the 3PL supplier of 15-node, only a partial information is displayed in Table IX.

TABLE VIII. 15-NODE CALCULATION EXAMPLE RELATED INFORMATION

Node	Cost	The mean of random time	The variance of random time	The number of infections
s	10	6	4	34
1	12	7	4	33
2	8	4	1	95
3	6	5	1	16
4	14	8	4	87
5	9	6	4	30
6	8	5	1	15
7	12	6	4	69
8	10	5	1	31
9	11	6	4	44
10	9	6	1	18
11	14	7	4	21
12	8	5	1	10
13	15	6	4	14
t	7	5	1	40

TABLE IX. 15-NODE CALCULATION EXAMPLE 3PL SUPPLIER RELATED INFORMATION

Initial	End	3PL Number	Transportation cost	The mean of random time	The variance of random time
s	1	1	20	12	16
s	1	2	18	15	25
s	1	3	24	10	4
s	2	1	18	16	16
s	2	2	17	17	9
s	2	3	19	15	25
s	3	1	19	14	9
s	3	2	18	15	25
s	3	3	20	14	4
1	4	1	10	8	1
1	4	2	11	6	4
1	5	1	10	8	9
1	5	2	12	7	1

Fig. 9 depicts the detail, it is the path diagram corresponding to the solution result of 15-node. Among them, the black pentagram *s* represents the initial node city, and the black pentagram *t* is the destination node city. The node cities marked with red pentagrams 2, 4, and 7 are transfer node cities that do not meet the constraint of the number of infected people, that is, transfer node city 2, node city 4, and node city 7 are in high-risk areas, so the path cannot pass through node city 2, node city 4, and node city 7. The remaining transfer node cities that meet the constraint of the number of infected people. The path marked with the blue thick line is the distribution path corresponding to the minimum CVaR that satisfies the cost and the infection risk constraints obtained, and the detailed data is depicted in Table X.

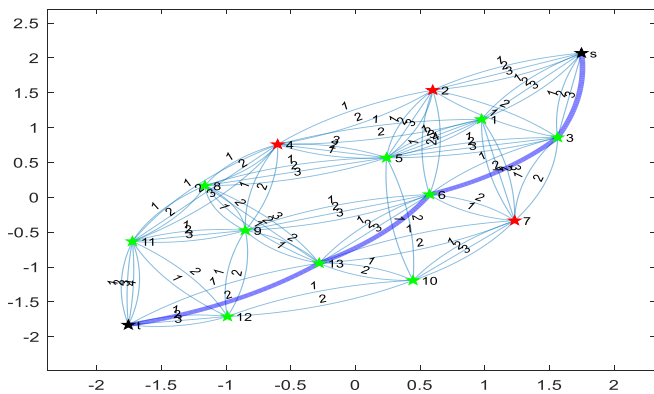


Fig. 9. 15-node solution path diagram.

The solution of the CVaR model of the 15-node problem are given by Table X under different confidence levels, when

the cost constraint is $C_0=115$ and the maximum acceptable the infection risk given by the customer is 1.6×10^{-4} .

TABLE X. SOLUTION OF 15-NODE PROBLEM WHEN $T_0=70$, $C_0=115$, $F_0=1.6 \times 10^{-4}$

β	episode	CVaR	Best path	F	Best rate	Time
0.9	200	12.1589	$R=\{v_s, 3, v_3, 2, v_6, 3, v_{13}, 2, v_t\}$	1.28×10^{-4}	0.98	0.9s
0.95	200	14.2909	$R=\{v_s, 3, v_3, 2, v_6, 3, v_{13}, 2, v_t\}$	1.28×10^{-4}	0.98	0.9s
0.99	200	18.4651	$R=\{v_s, 3, v_3, 2, v_6, 3, v_{13}, 2, v_t\}$	1.28×10^{-4}	0.98	0.9s

It can be known from the data in Table X that when the confidence level is 0.9, the cost constraint $C_0=115$, and the infection risk constraint $F_0=1.6 \times 10^{-4}$, the obtained optimal CVaR value is 12.1589, indicating that the corresponding average delay risk of the distribution task is 12.1589. The corresponding distribution cost is 115, and the infection risk is 1.28×10^{-4} . The corresponding best distribution path is $R=\{v_s, 3, v_3, 2, v_6, 3, v_{13}, 2, v_t\}$, indicating that when transporting from the source node city *s* to the destination node city *t*, the selected transfer node cities are 3, 6, and 13 respectively, and the number of the 3PL supplier selected between each two node cities is 3, 2, 3, 2. When the confidence level is 0.95, the cost constraint $C_0=115$, and the infection risk constraint $F_0=1.6 \times 10^{-4}$, the obtained optimal CVaR value is 14.2909, indicating that the corresponding average delay risk of the distribution task is 14.2909. The corresponding distribution cost is 115, and the infection risk is 1.28×10^{-4} . The corresponding best distribution path is $R=\{v_s, 3, v_3, 2, v_6, 3, v_{13}, 2, v_t\}$, indicating that when transporting from the source node city *s* to the destination node city *t*, the selected transfer node cities are 3, 6, and 13 respectively, and the number of the 3PL supplier selected between each two node cities is 3, 2. When the confidence level is 0.99, the cost constraint $C_0=115$, and the infection risk constraint $F_0=1.6 \times 10^{-4}$, the obtained optimal CVaR value is 18.4651, indicating that the corresponding average delay risk of the distribution task is 18.4651. The corresponding distribution cost is 115, and the infection risk is 1.28×10^{-4} . The corresponding best distribution path is $R=\{v_s, 3, v_3, 2, v_6, 3, v_{13}, 2, v_t\}$, indicating that when transporting from the source node city *s* to the destination node city *t*, the selected transfer node cities are 3, 6, and 13 respectively, and the number of the 3PL supplier selected between each two node cities is 3, 2.

The above data indicates that when the cost constraint and the infection risk constraint remain unchanged, as the confidence level increases, the corresponding optimal distribution path will not change, so the infection risk faced will not change either. However, the average delay risk faced by customers will be higher. Therefore, by using this model, 4PL suppliers can combine the customers' aversion to risk and

consider the impact of the infection risk on the distribution plan, so that the distribution path does not pass through high-risk areas, and at the same time, provide customers with the distribution path with the smallest average delay risk that meets the customer's infection risk and cost requirements under the given confidence level. It not only reduces the risk of virus transmission, but also provides customers with a green and efficient delivery solution with the lowest delay risk under a specific confidence level, helping the logistics industry move towards a safer and more sustainable future.

The confidence level selected by clients is significantly influenced by their risk preferences. A higher confidence level may be chosen by clients who are averse to delay risks in order to enhance security, whereas clients with a propensity for taking risks may opt for a lower confidence level to increase the diversity of viable routes. Consequently, our plan is meticulously aligned with the clients' risk preferences, enabling the formulation of bespoke control schemes. Utilizing this plan, 4PL providers can fully take into account the clients' aversion to delay risks, devising distribution routes that circumvent high-risk zones and achieve minimization of the mean delay risk under the specified confidence level. This approach not only mitigates the infection risk but also delivers a green and efficient distribution solution with the minimal delay risk at a particular confidence level, thereby aiding the logistics industry in forging a safer, more efficient, and sustainable distribution system.

C. Algorithm Comparison

In this paper, two distinct algorithms were utilized to tackle the 4PLROP: the Genetic algorithm embedded with the Dijkstra algorithm and the improved Q-learning algorithm. With the aim of assessing the efficacy of these algorithms across varying problem scales, three case studies of differing magnitudes were selected for analysis, encompassing 7 nodes, 15 nodes, and 30 nodes respectively. A comparative examination of the solution outcomes derived from these algorithms on the aforementioned case studies facilitates a profound comprehension of their divergent performances in terms of solution efficiency, solution quality, and stability. This, in turn, furnishes a more efficacious basis for algorithm selection in addressing real-world logistics routing optimization issues. Subsequently, in an endeavor to further scrutinize whether the improved Q-learning algorithm exhibits significant performance enhancements over the traditional Q-learning algorithm, a comparative study was undertaken between the improved Q-learning algorithm and the traditional Q-learning algorithm.

The Genetic algorithm embedded with the Dijkstra algorithm and the improved Q-learning algorithm are used to solve three examples of different scales. The comparison data are demonstrated in Table XI. It can be known from the data in Table XI that when solving the small scale problem of 7-node, both the improved Q-learning algorithm and the Genetic algorithm embedded with the Dijkstra algorithm can find the optimal solution, but the latter shows an inferior solving speed. With the increase of the solving scale, the improved Q-learning algorithm presents a higher solving speed and solving quality. Although the Genetic algorithm embedded with the Dijkstra algorithm can find the optimal solution, the number of

iterations and time increase significantly. The main reason is that in this algorithm, a simple graph is first generated, and the Dijkstra algorithm is used to find the optimal path on the simple graph, and then the Genetic algorithm is used for optimization. This leads to the possibility that different simple graphs may find the same path, thereby delaying the optimization convergence process and rapidly increasing the solving time.

TABLE XI. COMPARISON OF RESULTS OF DIFFERENT ALGORITHMS

Node	Algorithm	CVaR	Best path	Best rate	Time
7	Improved Q-learning	22.0456	$R=\{v_s, 2, v_2, 2, v_3, 1, v_t\}$	1	0.9s
	Embedded Dijkstra's Genetic Algorithm	22.0456	$R=\{v_s, 2, v_2, 2, v_3, 1, v_t\}$	0.95	19.4s
15	Improved Q-learning	3.7728	$R=\{v_s, 1, v_2, 2, v_6, 3, v_{13}, 2, v_t\}$	0.98	1s
	Embedded Dijkstra's Genetic Algorithm	3.7728	$R=\{v_s, 1, v_2, 2, v_6, 3, v_{13}, 2, v_t\}$	0.94	24.5s
30	Improved Q-learning	13.641	$R=\{v_s, 1, v_4, 2, v_8, 1, v_{12}, 1, v_{15}, 2, v_{18}, 4, v_{21}, 1, v_{25}, 1, v_t\}$	0.95	1.5s
	Embedded Dijkstra's Genetic Algorithm	13.641	$R=\{v_s, 1, v_4, 2, v_8, 1, v_{12}, 1, v_{15}, 2, v_{18}, 4, v_{21}, 1, v_{25}, 1, v_t\}$	0.9	28.5s

Fig. 10-12 provide a visual representation, they respectively represent the comparison curves of the traditional Q-learning algorithm and the improved Q-learning algorithm when solving 7-node, 15-node and 30-node. In the table, iQlearning refers to improved Q-learning algorithm.

Fig. 10-12 offer a graphical summary of the results, the red curve represents the solution curve of the traditional Q-learning algorithm, and the exploration strategy adopted is the ϵ -greedy strategy, where the value of ϵ is 0.5; the green curve represents the solution curve of the improved Q-learning algorithm described in this paper, and the strategy adopted is to randomly select actions with a reward value greater than -1 in the current state.

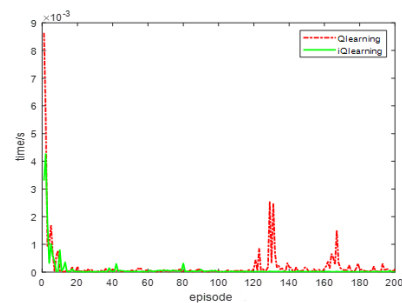


Fig. 10. Comparison of two algorithms at 7-node.

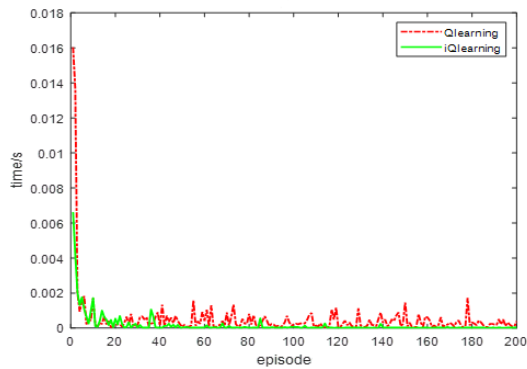


Fig. 11. Comparison of two algorithms at 15-node.

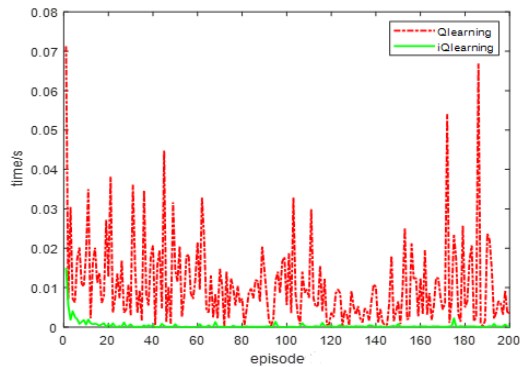


Fig. 12. Comparison of two algorithms at 30-node.

When solving the 7-node problem, both the improved Q-learning algorithm and the traditional Q-learning algorithm can converge relatively quickly, but the former has higher stability and a relatively faster convergence speed. As the problem scale increases, the improved Q-learning algorithm shows higher stability and solution speed. Through the performance of the traditional Q-learning algorithm and the improved Q-learning algorithm in different examples in the above text, it can be seen that the improved algorithm has a faster exploration speed, stronger stability, and a faster convergence speed, which has strong practical significance in solving the 4PLROP.

In summary, this paper has drawn the following conclusions through a comparative analysis of the solution performance of the Genetic algorithm embedded with the Dijkstra algorithm and the improved Q-learning algorithm on case studies of varying scales, as well as the performance disparities between the improved Q-learning algorithm and the traditional Q-learning algorithm: The improved Q-learning algorithm has demonstrated significant performance advantages in addressing the 4PLROP, whether it be in small-scale or large-scale issues, including a faster solution speed, a higher solution quality, and a stronger stability. This indicates that the improved Q-learning algorithm is an effective and practical solution method, capable of providing robust support for path optimization in actual logistics distribution. In the future, we will continue to conduct in-depth research and optimization of this algorithm to further enhance its adaptability and solution efficiency in complex logistics

environments, thereby offering a more comprehensive solution for the resolution of 4PLROP.

VI. CONCLUSIONS

In the context of a severe pandemic environment, the stability and efficiency of logistics services are crucial for ensuring the continuity of societal operations and the well-being of the populace. Particularly against the backdrop of the dual carbon goals (Carbon Peak and Carbon Neutrality), it is imperative not only to meet the demands of minimizing the risk of delivery delays for clients but also to give due consideration to the prevention and control of infection risks, as well as the imperatives of energy conservation and emission reduction. Together, these efforts weave a logistics network that is green, secure, and efficient, contributing to the sustainable development of the planet.

The present article introduces the CVaR measure and constructs a novel mathematical model. This model not only incorporates distribution costs and infection risks as significant constraints but also aims to minimize the CVaR as the optimization target. Through this model, it ensures that distribution plans can meet the requirements of cost effectiveness and risk control under the complex and variable epidemic environment. To satisfy the demands of this model, the reward and punishment mechanisms in the Q-learning algorithm have been redesigned to more accurately reflect the various risks and cost factors in the actual distribution process. By solving different cases, the lowest CVaR distribution routes that meet the requirements of cost and customer infection risks are obtained. Customers can obtain multiple schemes according to their risk preferences and take corresponding measures. This article provides scientific decision making basis and efficient and safe distribution plans for the 4PL, promoting the logistics industry to move towards a low-carbon, environmentally friendly, and sustainable direction.

Meanwhile, the probability distribution of the stochastic distribution time and transit time is known in this study. When these parameters are unknown, our research is intended to be extended to a robust 4PLROP considering infection risk and delay risk. We anticipate that the improved Q-learning algorithm and the Genetic algorithm embedded with the Dijkstra algorithm will continue to shine brightly. Their potential in 4PLROP is boundless. Moreover, we envision a future where they are seamlessly integrated into various other crucial aspects of the logistics and transportation ecosystem.

ACKNOWLEDGMENT

This research was funded by the Natural Science Foundation of Liaoning Province, grant number 2024-BS-227; the Foundation of the Educational Department of Liaoning Province, grant number No. JYTQN2023345; the Social Science Foundation of Liaoning Province, grant number L24CGL031.

REFERENCES

- [1] M. Huang, L. W. Dong, H. B. Kuang, Z. Z. Jiang, L. H. Lee, X. W. Wang, "Supply chain network design considering customer psychological behavior—a 4PL perspective," *Comput. Ind. Eng.*, pp. 159, 2021.

- [2] M. Q. Yin, M. Huang, X. H. Qian, D. Z. Wang, X. W. Wang, L. H. Lee, "Fourth-party logistics network design with service time constraint under stochastic demand," *J. Intell. Manuf.*, vol. 34, pp. 1203–1227, 2023.
- [3] D. G. Mogale, D. Xian, V. Sanchez Rodrigues, "Managing logistics risks in pharmaceutical supply chain: a 4PL perspective," *Prod. Plan. Control*, pp. 1–16, 2024.
- [4] K. Kang, R. Y. Zhong, S. X. Xu, "Auction-based cloud service allocation and sharing for logistics product service system," *J. Clean. Prod.*, vol. 278, 2021.
- [5] F. Q. Lu, H. L. Bi, W. J. Feng, Y. L. Hu, S. X. Wang, X. Zhang, "A Two-Stage Auction Mechanism for 3PL Supplier Selection under Risk Aversion," *Sustainability*, vol. 13, pp. 9745, 2021.
- [6] M. B. Çağlar Kalkan, K. Aydın, "The role of 4PL provider as a mediation and supply chain agility," *Mod. Supply Chain Res. & Appl.*, vol. 2, pp. 99–111, 2020.
- [7] D. Werf, "Information Technology and Data Use in 1PL-4PL Logistic Companies," 2021.
- [8] Y. Tao, E. P. Chew, L. H. Lee, Y. R. Shi, Y. Tao, E.P. Chew, L.H. Lee, Y.R. Shi, "A column generation approach for the route planning problem in fourth party logistics," *J. Oper. Res. Soc.*, vol. 68, pp. 165–181, 2017.
- [9] H. Pavlič Skender, P. A. Mirković, I. Prudky, "The role of the 4PL model in a contemporary supply chain," *Pomorstvo*, vol. 31, no. 2, pp. 96–101, 2017.
- [10] M. Huang, L. Ren, L. Hay. Lee, X. W. Wang, "4PL routing optimization under emergency conditions," *Knowl. - Based Syst.*, vol. 89, pp. 126–133, 2015.
- [11] M. Huang, L. Ren, L. H. Lee, X. W. Wang, H. B. Kuang, H. B. Shi, "Model and algorithm for 4PLRP with uncertain delivery time," *Inf. Sci.*, vol. 330, pp. 211–225, 2016.
- [12] R. R. Ren, Y. F. Zhao, F. Q. Lu, M. Feng, "Research on 4PL Routing Problem Considering Customer Risk Preference," *Math. Pract. Theory*, vol. 53, no. 1, pp. 163–174, 2023.
- [13] W. Hong, Z. I. Xu, W. Liu, L. H. Wu, X. J. Pu, "Queuing theory-based optimization research on the multi-objective transportation problem of fourth party logistics," *Proc. Inst. Mech. Eng. B*, vol. 235, no. 8, pp. 1327–1337, 2021.
- [14] S. H. Yang, J. Zhu, Q. Wang, M. Huan, "Routing Problem with Stochastic Delay Time Under Forth Party Logistics," *Proc. 32nd Chinese Control Decis. Conf.*, vol. 5, no. 6, 2020.
- [15] W. Y. Cai, X. F. Wang, X. H. Qian, M. Q. Yin and Y. X. Li, "A route problem with customers' preferences for a fourth party logistics provider," *CCDC. Kunming. China*, 2021, pp. 4185–4189, 2021.
- [16] A. Tatarczak, "A Framework to Support Coalition Formation in the Fourth Party Logistics Supply Chain Coalition," *Acta Univ. Lodz. Folia Oecon.*, vol. 5, no. 338, pp. 192–212, 2018.
- [17] S. H. Zhou, Q. Wang, Y. B. Sun, M. T. Yang, D. X. Li, M. Huang, "A Combinatorial Optimization Model for Customer Routing Problem Considering Cost Discount," *CCDC. Hefei. China*, pp. 2210–2214, 2020.
- [18] M. Huang, L. Ren, L. H. Lee, X. W. Wang, "4PL routing optimization under emergency conditions," *Knowl. - Based Syst.*, vol. 89, pp. 126–133, 2015.
- [19] M. Huang, L. Ren, L. H. Lee, X. W. Wang, "Model and algorithm for 4PLRP with uncertain delivery time," *Inf. Sci.* vol. 330, pp. 211–225, 2016.
- [20] F. Q. Lu, H. L. Bi, L. Huang, W. Bo, "Improved genetic algorithm based delivery time control for Fourth Party Logistics," *2017 13th IEEE CASE. Xi'an. China*, pp. 390–393, 2017.
- [21] F. Q. Lu, W. J. Feng, M. Y. Gao, H. L. Bi, S. X. Wang, "The Fourth-Party Logistics Routing Problem Using Ant Colony System-Improved Grey Wolf Optimization," *J. Adv. Transp.*, vol. 1, October 2020.
- [22] F. Q. Lu, W. D. Chen, W. J. Feng, H. I. Bi, "4PL routing problem using hybrid beetle swarm optimization," *SOFT COMPUT.*, vol. 27, pp. 17011–17024, 2023.
- [23] R. Liang, F. Qing, S. Yuan, L. Ao, "Fourth Party Logistics Routing Problem with Time Window in Uncertain Environments," *Inf. Control*, vol. 47, no. 5, pp. 583–588, 2018.
- [24] X. Y. Gao, X. Gao, Y. Liu, "Fourth Party Logistics Routing Problem Under Uncertain Time and Random Demand[J]. *Journal of Uncertain Systems*," 2024.
- [25] L. Ren, Z. R. Zhou, Y. P. Fu, A. Liu, Y. F. Ma, "Integrated optimization of logistics routing problem considering chance preference," *Mod. SCRC Appl.*, 2024.
- [26] L. S. Deng, M. Huang, D. Z. Wang, M. Q. Yin, "Multi-point to multi-point multi-task fourth party logistics routing problem considering tardiness risk," *IEEE CASE. C*, pp. 1350–1355, 2017.
- [27] G. H. Bo, M. Huang, "Model and Solution of Routing Optimization Problem in the Fourth Party Logistics with Tardiness Risk," *ComplexSyst. Complex. Sci.*, vol. 15, no. 03, pp. 66–74, 2018.
- [28] W. Wang, M. Huang, X. W. Wang, "The Optimization Model of 4PL Routing Problem for Risk-averse Customer," *CCDC. Hefei. China*, pp. 2215–2219, 2020.
- [29] X. Liu, G. H. Bo, "Q-Learning Algorithm for Fourth Party Logistics Route Optimization Considering Tardiness Risk," *Proceedings of the 2022 International Conference on Cyber-Physical Social Intelligence*, 2022.
- [30] J. Tu, L. D. Wan, Z. J. Sun, "Safety Improvement of Sustainable Coal Transportation in Mines: A Contract Design Perspective," *Sustainability*, vol. 15, no. 3, pp. 2085–2095, 2023.

Convolutional Neural Network and Bidirectional Long Short-Term Memory for Personalized Treatment Analysis Using Electronic Health Records

Prasanthi Yavanamandha^{1,3}, D. S. Rao²

Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,
Hyderabad-500075, Telangana, India^{1,2}

Department of CSE-AIML & IoT-Faculty, VNR Vignana Jyothi Institute of Engineering & Technology,
Hyderabad, India³

Abstract—Correct precision techniques have far not been introduced for modeling the modality risk in Intensive Care Unit (ICU) patients. Traditional mortality risk prediction techniques effectively extract the data in longitudinal Electronic Health Records (EHRs), that ignore the difficult relationship and interactions among variables and time dependency in longitudinal records. The proposed work, developed the Convolutional Neural Network – Bidirectional-Long Short-Term Memory (CNN-Bi-LSTM) method for personalized treatment analysis using EHR data. The CNN extracts the significant features from relevant features, focused on spatial-based relationships. Then, the Bi-LSTM layer captured the sequential dependencies and temporal relationships in patient histories that are essential to understand the treatment results. The Circle Levy flight – Ladybug Beetle Optimization (CL-LBO) integrates the circle chaotic map and Levy flight process in traditional LBO to select relevant features for classification. The proposed method reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% Area Under Curve (AUC) when compared to LSTM.

Keywords—*Bidirectional-long short-term memory; circle chaotic map; convolutional neural network; electronic medical records; Intensive Care Unit (ICU); ladybug beetle optimization*

I. INTRODUCTION

The healthcare centers are responsible for maintaining patient documents tracked to analyze the disease [1]. Manually handling the document is a difficult task, as the documents get lost, damaged, or destroyed [2]. Collecting confidential information during the readmission of patients became difficult and delayed the patient's treatment [3]. Securing the information of the medical data helps for the analysis of existing disease types, and maintains the confidential patient data [4-6]. Electronic Health Records (EHR) is a digital platform for the data storage and maintenance of medical patient history for future analysis [7]. The EHR allows the providers to access the details of the patients easily and quickly which provides more information of the patients to get better treatment [8-10]. The EHR platform also improves communication among healthcare that enhances coordination to reduce medical errors [11]. In the unavoidable and increasing digital transformation process of national healthcare system management, a significant size of structural EHR data is available [12].

The advancement in recent technologies has introduced many applications for the automatic recording and maintenance of medical patient history [13, 14]. Deep Learning (DL) and Machine Learning (ML) methods are introduced for automatic analysis of medical data. These methods undergo training to predict the result of data [15]. The Intensive Care Unit (ICU) technologies and automatic system development have effectively improved the efficiency of healthcare professionals including caregivers, nurses, and doctors [16]. The EHR is used for model training to learn details present in data [17]. The Medical Information Mart for Intensive Care III (MIMIC-III) dataset is considered as input for DL and ML methods. The study focused on solving the feature selection and dimensionality problems to maximize the performance of the model. The oversampling method is used in the pre-processing step to balance the data, fitness function-based optimization technique is used in the feature selection process to improve classification accuracy and prediction.

The essential contributions of the proposed framework are, the Synthetic Minority Over-Sampling Technique (SMOTE) is introduced in the pre-processing phase to balance the classes in data, which enhances the performance and generalization capability of the method, the Circle Levy flight – Ladybug Beetle Optimization (CL-LBO) based feature selection algorithm is developed to choose relevant and appropriate features whole feature subset, which helps to enhance the classification performance, the CL-LBO integrated the circle chaotic map and Levy flight in traditional LBO, which enhances the searchability and convergence rate of LBO to select relevant features, the Convolutional Neural Network – Bidirectional-Long Short-Term Memory (CNN-Bi-LSTM) is developed for classification to focus much on spatial-based relationships and to capture the sequential dependencies and temporal relationships in patient histories.

The previous research is analyzed in a literature survey is given in Section II. The proposed method explains the process of personalized treatment analysis using the proposed methodology presented in Section III. The results of the proposed model have been examined and compared with existing machine learning models in Section IV. Finally, the paper is concluded in Section V.

II. LITERATURE REVIEW

Fang Yang et al. [18] presented a deep morality risk prediction based on longitudinal EHR data. The LSTM model was used as a classifier to predict mortality rate. Visit-level and variable-level attention mechanisms were used to solve the gradient vanishing issue that occurred during the model training to predict appropriate outcomes that enhanced the efficiency of the model. However, the missing values presented in the data caused an overfitting problem during the process of training that affected the accuracy of prediction and reduced the performance of the model.

Shuai Niu et al. [19] presented a model to predict congestive heart failure mortality based on the feature selection method. Various ML techniques like Decision Tree (DT), Logistic Regression (LR), Random Forest (RF), and Gradient Boosting (GB) were used as a classifier to predict the mortality rate. The Partial Swarm optimization (PSO)-based feature selection algorithms were utilized to select significant details of features that enhanced the accuracy of prediction and classification. However, the model struggled to interpret the pattern of features due to huge variables present in the data affected the training of the model to predict accurate outcomes and reduced the accuracy of classification. Maria Bampa et al. [20] presented a multimodal clustering technique for understanding the patient phenotype. A Multimodal Autoencoder (MMAE) model was used as a classifier to group the disease based on similar features. The model strength was enhanced by combining multiple modalities to improve the clustering of data and enhance the accuracy of classification. However, the model suffered from getting the approximate hyperplane required for the classification due to a huge number of variables that reduced the accuracy of classification.

Belel Alsinglawi et al. [21] presented a ML framework for the prediction of lung disease in hospitals. The RF algorithm was used as a classifier to predict the type of lung disease. The SMOTE was utilized to solve imbalance issues of the data that enhanced the training of the model to predict accurate outcomes. However, the model failed to identify the significant details of the data due to a dimensionality issue that reduced the accuracy of classification. Hua Shen [22] presented the Recurrent Neural Network (RNN) model to improve the prediction of diseases in healthcare. An attention mechanism was used to reduce the gradient vanishing problem that occurred during the training process where the maximum gradients were retained which increased the accuracy of prediction to get the approximate outcome. However, the model suffered from gradient vanishing issues where the gradients vanished during the training process which affected the accuracy of classification.

Sapiah Sakri et al. [23] presented a hybrid model to predict sepsis disease based on extracted features. Convolutional Neural Network (CNN) and Bi-directional LSTM models were combined to classify sepsis disease. The spatial and temporal features of the data were used to select the significant details of the data where the model identified the pattern of selected features that enhanced the performance of the models. However,

the extra layers present in the neural network affected the interpretability of the model that reduced the accuracy of classification. Sarika R. Khope et al. [24] presented a featured engineering-based disease prediction Artificial Neural Network (ANN). The encoding method was utilized to eliminate the outer lines of the features and improve the accuracy of prediction and classification. The feature engineering technique was used to get significant details of the data that enhanced the accuracy of prediction. However, the overfitting issue occurred during the process because the irrelevant features present in the data fit the model and learned the noises that affected the accuracy of classification.

Chang Liu et al. [25] presented a different ML models like K Nearest Neighbor (KNN), DT, NB, LR, and RF for the early prediction of Multiple Organ Deficiency Syndrome (MODS). Kernel Shapley Additive exPlanations (SHAP) was utilized to enhance the quality of data. The ML methods used here enhanced the potential of the prediction by clustering the data which enhanced the accuracy of classification. However, the model struggled to identify the appropriate hyperplane required to classify the data approximately which reduced the accuracy of prediction and classification. Vinod Kumar Chauhan et al. [26] presented a DL-based attention model to enhance the classification of HER data. The LSTM model was used as a classifier along with a cross-attention-based transformer model for the prediction of the accurate disease. The LSTM model focused on identifying the significant features of data that enhanced classification accuracy. However, the model performance was reduced due to poor hyperparameter tuning due to a lack of efficient optimal value that reduced the accuracy of classification.

Ho-Joon Lee et al. [27] presented an ensemble consensus model based on HER for the classification of strokes. A stroke classifier technique was used to predict the outcome of the model. The Principal Component Analysis (PCA) technique was employed to minimize the dimensionalities of data and enhance the accuracy of classification. However, the model interpretability was challenging due to missing values presented in data that reduced the accuracy of classification.

From the overall analysis, the existing researches have limitations such as overfitting issues occurring during the process because the irrelevant features present in the data fit, dimensionality issues, struggle to interpret the pattern of features, and the missing values presented in the data caused an overfitting problem.

To mitigate these limitations, the CNN-Bi-LSTM method for personalized treatment analysis using EHR data is developed. The CNN extracts the significant features from relevant features, focused on spatial-based relationships. Then, the Bi-LSTM layer captured the sequential dependencies and temporal relationships in patient histories that are essential to understand the treatment results. The SMOTE technique is employed in the pre-processing phase to balance the classes in data.

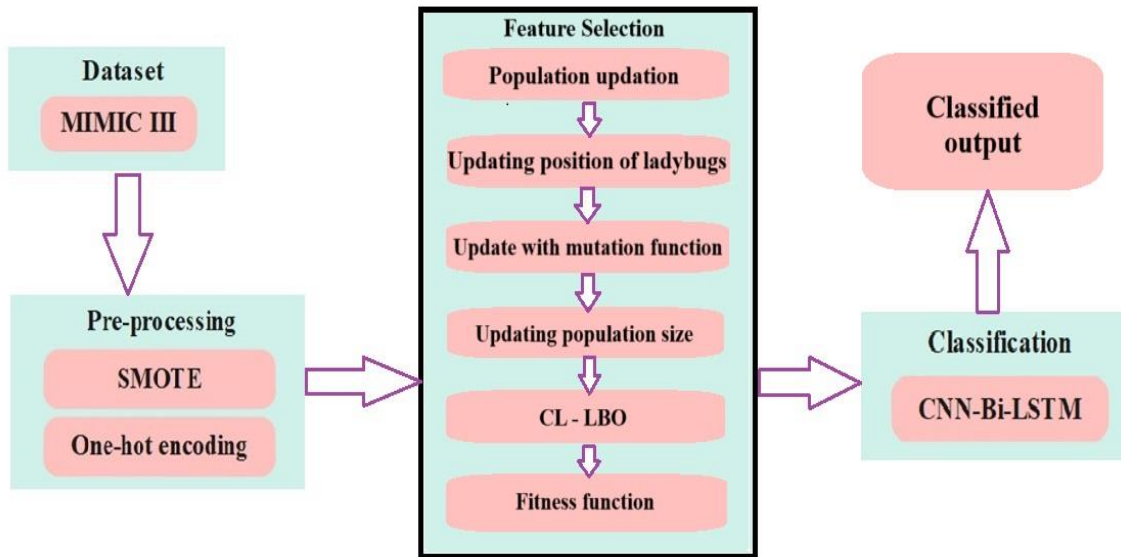


Fig. 1. Proposed architecture of personalized treatment analysis using EHR data.

The CL-LBO integrates the circle chaotic map and Levy flight process in traditional LBO to select relevant features for classification. This process improves the process of personalized treatment with high classification accuracy.

III. PROPOSED METHOD

The DL-based technique is developed in this work for personalized treatment analysis using EHR data. The MIMIC III dataset is used for personalized treatment and the data is pre-processed by using one-hot encoding and SMOTE techniques. Then, the relevant features are chosen by using CL-LBO that integrated the circle chaotic map and levy flight process. At last, the CNN-Bi-LSTM method is developed to classify features in data with classification accuracy. Fig. 1 describes the process of personalized treatment analysis using EHR data.

A. Dataset

The dataset used for this research is MIMIC III dataset [28] which includes 1177 cases for in-hospital mortality prediction and has a total of 51 features. The value of 0 represents life and 1 represents death. The BMI, gender, and Age are demographic factors studied. The essential signs are blood pressure, heart rate, respiratory rate, blood temperature, urine result, and saturation pulse oxygen. The characteristics of comorbidity include atrial fibrillation, depression, hyperlipidemia, hypertension, and diabetes. The white blood cells, basophils, lymphocytes, red blood cells, neutrophils, potassium, anion gap, sodium, bicarbonate, lactate, chloride, creatinine, calcium, magnesium, prothrombin time and creatine kinase are the laboratory variables.

B. Pre-processing

The pre-processing of data is employed to issue of imbalance samples is dissimilar from copy sample mechanism in random oversampling. The SMOTE technique synthesized the new samples among two minority samples by linear interpolation,

which efficiently mitigates overfitting issue caused through random oversampling that makes balanced class distribution and enhances generalization capability of classifier [29].

The basic principle of SMOTE technique initially, choose every sample x_i from minority samples as the origin sample of a new synthetic sample, next, in accordance with up-sampling magnification n , randomly choose any neighboring samples k of similar classes in the sample x_i is an auxiliary sample in producing new samples, repeated n times, next linear interpolation is done among each sample and every auxiliary sample by using Eq. (1), and at last n synthesized samples are produced.

$$x_{new,attr} = x_{i,attr} + rand(0,1) \times (x_{j,attr} - x_{i,attr}) \quad (1)$$

In Eq. (1), the $x_{new,attr}$ for $attr = 1, 2, \dots, d$ represents $attr - th$ attribute value in i th sample of the majority sample, the $rand(0,1)$ represents a random number between 0 and 1, the x_{ij} for $j = 1, 2, \dots, k$ represents j th nearest neighbor sample of x_i , the x_{new} is the new sample synthesized among j and x_{ij} .

1) *One-hot encoding*: One-hot encoding is called as the one-bit effective encoding. This technique utilizes N-bit status registers for encoding N states. Every state has an independent register bit and one bit is valid. This technique is a representation of categorical variables in binary vectors with the benefit which transforming sample dataset to develop which is easy for utilizing the ML, particularly in classification algorithm. This process effectively enhances the measuring speed and method's performance.

C. Feature Selection

The general procedure of numerous metaheuristic algorithms is same to each other. In the proposed procedure, initial population of algorithms is sorted and evaluated depended on its evaluation. Next, population is updated and reevaluated.

After repeating the necessary updating and evaluation process for population, the optimal solution is described.

The LBO is motivated by coordination movements of ladybugs to identify the position with much heat. For this process, initial population that has $N(0)$ ladybugs are taken and the last population contains $N(k_{max})$ ladybugs and optimum fitness function is determined. Then, the LBO modeling is processed in three phases.

1) *Population updation*: The initial population includes $N(0)$ ladybugs that are positioned randomly in search area depended in uniform distribution. The ladybugs population is estimated with determined fitness function and sorted. Next, population moved to position with much heat in accordance with coordination movement. Because of the ladybug's nature which always moved in coordination with a swarm of ladybugs when searching for an appropriate position, the ladybugs follow each other by signals emitted through group members. Hence, much inclined to move towards it from ladybugs. In the proposed model, front ladybugs who have cable identify the position with more heat than others. For exploration and exploitation of the algorithm, the mutation phase is taken for certain individuals of the population that are randomly assigned to certain individuals in every iteration. However, at every phase of position update in population, its location in the search area is updated in accordance with other positions or mutation phases that are represented below.

2) *Updating the position of other ladybugs*: In every phase, entire positions of ladybugs are evaluated and updated. The old and new position of ladybugs are combined and optimal groups are selected in accordance with its fitness function values. The new population is assigned for updating and evaluating in following iteration. For updating every member of population in every iteration, another population member is chosen by utilizing the technique which is described. For instance, consider that goal is to update i th population of ladybug and j th population of ladybugs have selected for updating the member. Taken the location of i th ladybug in k th iteration is represented as $x(k)$. This ladybug moves in the outcome of three vectors for updating in $(k + 1)$ iteration moves a little towards j th ladybug and move in the direction of j th ladybug towards $(j - 1)$ th ladybug. To avoid local optima for movements of i th ladybugs and to maintain the balance between exploration and exploitation, the third direction is also taken, that is coefficient of their present position. All the determined movements are required to be multiplied through random values to sustain the random search. Moreover, the third movement that is assigned for avoiding the local optima is multiplied through the ratio of individual heat value to all heat values of the population. As an outcome, population members that are trapped in local optima have a chance to outflow local optima due to its high coefficient of third movement. The mathematical formula for the new position of i th ladybug is given as Eq. (2),

$$x_i(k + 1) = x_i(k) + rand \times (x_j(k) - x_i(k)) + rand \times (x_j(k) - x_{j-1}(k)) + rand \times |C_i|^{\frac{k}{N(k)}} \times x_i(k) \quad (2)$$

In equation (2), the C_i represents the same proportion of i th ladybird cost to total cost of entire ladybirds in k th iteration of developed algorithms. The mathematical formula for parameter value is measured by using Eq. (3),

$$C_i = \frac{f(x_i(k))}{\sum_{t=1}^{N(k)} f(x_t(k))} \quad (3)$$

The Roulette-wheel selection is assigned for selecting the j th ladybug is utilized for updating the i th ladybug location using Eq. (3). In the general equation, for choosing j th ladybug from $N(k)$ ladybugs, the distance between 0 and 1 is separated to $N(k)$ unequal phases. Every phase corresponds to one of ladybugs and length of every phase is inversely relevance to fitness function of respective ladybug. For ladybug with much optimum fitness function, length of respective phase is higher. Next, the random number among 0 and 1 is selected. In present phase, random number determined in that phase of separation is positioned. The respective ladybug of the chosen phase is selected as j th ladybug. This is clear that ladybugs with warmer positions have a high chance of being selected. The P vector respective to the population with $N(k)$ ladybugs are determined. The Roulette-wheel selection expected a represents input vectors like P and its mathematical formula is given as Eq. (4),

$$P = [P_1, P_2, \dots, P_{N(k)}], P_i = e^{-\beta \frac{f(x_i(k))}{f_{worst}}} \quad (4)$$

In Eq. (4), the β represents pressure coefficient in Roulette-wheel selection technique and f_{worst} represents worst value of fitness function for present iteration in the process of algorithm. The higher β , the optimal ladybugs of population have a good chance of being chosen in Roulette-wheel selection. This is clear that new location of i th ladybug is defined through outcome of three vectors r_1, r_2 and r_3 .

3) *Updation in accordance with mutation process*: Considering the mutation is updation process of the population is critical for exploration of unidentified phases of search area and escapes from local optima. The mutation phase in search area causes the maximum speed of algorithm. Therefore, updation technique of every position of the ladybug includes in accordance with other ladybugs and mutation is randomly defined. Considers the i th ladybug is mutated. The amount of decision variables of i th ladybug is mutated and its mathematical formula is given as Eq. (5),

$$n_m = round(n \times \mu_m) \quad (5)$$

In Eq. (5), the μ_m represents mutation rate and the n represents length of decision variable. Hence, n_m represents variables presented in n variables of i th ladybug is chosen randomly. Next, random variables in the feasible phase are replaced for choosing location of i th ladybug.

4) *Updating the number of population size*: In searching for warm place that is common for ladybugs to disappear and lost.

The ladybug moves away from annihilate and others because of cold. The mathematical methods for the annihilation of ladybugs in search are taken in LBO. The mathematical formula for the number of ladybugs in different phases is measured by using Eq. (6),

$$N(k + 1) = \text{round} \left(N(k) - \text{rand} \times N(k) \left(\frac{NFE}{NFE_{max}} \right) \right) \quad (6)$$

In Eq. (6), the NFE represents the number of function estimations and the NFE_{max} represents a maximum of NFE. Whether the amount of function evaluation is stopping criteria of LBO. Whether several iterations are a condition to terminate the algorithm. The mathematical formula for the ladybug in every iteration is given by using Eq. (7),

$$N(k + 1) = \text{round} \left(N(k) - \text{rand} \times N(k) \left(\frac{k}{k_{max}} \right) \right) \quad (7)$$

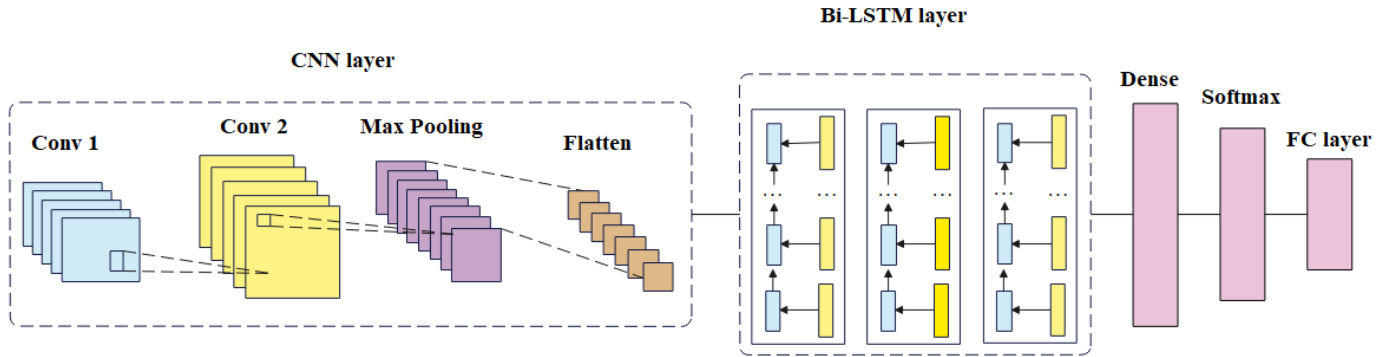


Fig. 2. Architecture of CNN-Bi-LSTM network.

In Eq. (7), the k represents iteration and the k_{max} represents maximum iteration.

5) *CL-LBO*: Enhanced non-linear LBO algorithm depended on Circle chaotic map and Levy flight process is introduced in the proposed work. Firstly, initialize the population by utilizing a circle chaotic map to maximize bee diversity. The integration of LBO and levy flight provides an algorithm with high capability in global exploration. Additionally, the non-linear adaptive weight operator is implemented for modifying the weight coefficient of bee following behavior in CL-LBO. The relationship between global exploration and local exploitation in the iteration process is effectively balanced. The chaotic values of actual circle operation are grouped in the range of [0.2, 0.5]. For making uniform chaotic value distribution, the mathematical method of the Circle chaotic mapping strategy is enhanced. The numerical expression for the circle chaotic map is given as Eq. (8),

$$x_{i+1} = \text{mod}(x_i + 0.2 - (0.5/2\pi) \sin(2\pi x_i), 1) \quad (8)$$

In Eq. (8), the x_i represents i th chaotic particle and the x_{i+1} represents $(i + 1)$ th chaotic particle. The frequency histogram and plot of the initial candidate solution of the circle chaotic mapping process.

a) *Levy flight*: The trajectory and movement of several little insects and animals in life have Levy flight characteristics. The insects and animals include flies and ants. Numerous animals in natural usage of Levy flight strategy are the essential path of foraging. The Levy flight is the process compatible with Levy distribution. The step size of the Levy flight is mixed and random with short and long distances that make it easy to search the huge scale and with unknown scope compared with Brownian motion. In the searching procedure, the Levy process utilized little steps for walking and long steps for jumping,

which allowed it for effective of local attraction points. Hence, in the random searching issue, numerous heuristic algorithms adopted this strategy for changing the iteration process that effectively supports the algorithm to get the influence of local attraction points. The mathematical formula for the strategy is given as Eq. (9),

$$L(s) \sim |s|^{-1-\beta} \quad (9)$$

In Eq. (9), the β in is range [0, 2], the s represents step size and the $L(s)$ represents the probability density of step size in accordance with Levy modeling.

6) *Fitness function*: For calculating the fitness value of generated CL-LBO agents, the Mean Square Error (MSE) fitness function is dependent on measuring the difference between original and predicted values through produced agents for training samples. The mathematical formula for MSE is in Eq. (10),

$$MSE = \frac{1}{n} \sum_{i=1}^n (y - \hat{y})^2 \quad (10)$$

In Eq. (10), the y describes the actual value, the \hat{y} describes the predicted value and the n describes a number of instances in the training set. From the CL-LBO algorithm, the 21 selected relevant features are given to the classification phase for further process.

D. Classification

The classification is performed by using the CNN-BiLSTM network. Initially, the architecture is developed for leveraging the advantages of CNN to filter and remove the noisy data and obtain substantial data from time series. The noisy data is eliminated by using dimensionality reduction. Hence, inappropriate data (noise) is not involved in minimized matrix. The significant data in the hidden interval is attained through employing the process of convolutional such as kernel matrix or filter passed by input matrix, various characteristics based on

kernel and that obtained the hidden data. Next, the design is developed to pull the ability of the Bi-LSTM setup for modeling and forecasting both short and long-term dependencies in temporal data. In this design, the input data is processed twice, concurrently from left to right and from right to left. Both context readings are combined into results and provide much more comprehensive data of information context than unidirectional LSTM. Here, The CNN functions as the encoder that identifies and extracts features from the input data, whereas the BiLSTM assists as the decoder, analyzing historical dependencies in the data stream. The CNN block contains the CNN layer, pooling layer, and flattening process. The output from the CNN block is conceded as input to the BiLSTM block. This block contains a BiLSTM layer, a dropout layer, and a dense layer. The convolutional and pooling layers are included to mine features from the data, operating it in a matrix format. Fig. 2. demonstrates the design of CNN-BiLSTM network architecture.

1) *CNN block*: The convolution and pooling layer are used to filter the incoming information for extracting significant data from the matrix. The convolution process is done through a convolution layer among input and small matrices known as filters and kernels. Generally, numerous filters are included as sliding windows with a certain height and width. These filters slide across the input matrix with a specified stride, assigning a convolution operation to each overlapping sub-region of the matrix. This procedure produces a convoluted matrix that bags a specific feature of the original input matrix. By using multiple convoluted features, the technique delivers a more specific representation of the input matrix. Consider H as an input matrix, the I represents kernel matrix and the indices m and n signifies the rows and columns of the subsequent matrix, respectively and its mathematical formula is given as Eq. (11) and Eq. (12),

$$R[m, n] = (H \cdot I)[m, n] \quad (11)$$

$$= \sum_j \sum_k I[j, k] \cdot H[m - j, n - k] \quad (12)$$

The convolution layer utilizes the ReLU activation function, which is a majorly used activation function. Its primary benefit is that it does not activate all neurons at the same time and transforms entire negative values to 0. Because of this, ReLU has high computation efficiency. The ReLU is six times quicker than other activation functions like tanh. The mathematical formula for the ReLU activation function is given as Eq. (13),

$$f(X) = x^+ = \max(0, x) \quad (13)$$

The max pooling layer will follow the convolution layer and turns as a sub-sampling technique to decrease the dimensions of the convolution matrix. It attains this by removing definite values while retaining key features recognized by each filter. For every patch of the matrix, it chooses the heights value, creating new matrices that helps as condensed notations of the convolution matrices. This pooling procedure improves the robustness of the technique. Finally, the pooling layer is tracked by a flattening procedure, which opens the values into a one-dimensional format to make the input for next layers.

2) *Bi-LSTM block*: The next segment of the method contains the BiLSTM module, which encompasses the BiLSTM network, a dropout layer, and a dense layer. To know the model and the performance of the BiLSTM network initially describes the one-directional LSTM network. The LSTM network is one of its kind of RNN, At this juncture they utilizes cyclic links in its inner layers which give short-term memory for the method and the capability to process thesequential data. Therefore, classical RNN have a problem known as the long-term dependencies issue, the poor memory of previous data as the neurons number increases. The LSTM network resolves this issue by storing relevant data by entire LSTM units in a type of conveyor belt called memory cell. Every LSTM unit includes a memory cell and 3 gates that regulate data flow by determining which data is forgotten and which one remains in the method. This is primary for learning the long-term dependencies through LSTM and resolving issues. In LSTM, three gates are there such as forget, input, and output gates. The parameters of the network used are Adam optimizer, softmax activation function, 50 epochs, and 64 batch sizes.

In forget gate, the sigmoid function is employed for data included in the present input X_t and past hidden state $(h - 1)$. This process is represented as f_t returned the value among 0 and 1 which represents the percentage of data. The input gate considers data from present data and past hidden states and is passed by the second sigmoid function, transforming the data to values between 0 and 1. The similar data passed by the tanh function that supports to regulation of the network returns the value among -1 and 1. Next, the sigmoid result i_t is multiplied through \tanh result for determining which data is significant to keep. Now, there is enough data for executing the cell state. Initially, past cell state is multiplied through forget result. Next, result of input gate is included, updates cell state with new values considered relevance through network. The outcome of these two processes is provided as new cell state. At last, the output gate is provided, which determines value of following hidden state. Initially, data from present input and from past hidden state passed by third sigmoid function. Next, new cell state passes by tanh function. The mathematical formula for LSTM cell is given from Eq. (14) to Eq. (19),

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (14)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (15)$$

$$\hat{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (16)$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \hat{C}_t \quad (17)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (18)$$

$$h_t = o_t \cdot \tanh(C_t) \quad (19)$$

In Eq. (14) to Eq. (19), the σ describes the sigmoid function, the \tanh describes the hyperbolic tangent function, x_t is input data, the h_t is a hidden state in time, the W_x and b_x are the weight matrix and bias vector. The Bi-LSTM network includes forward and backward LSTM networks. The forward LSTM utilizes the input sequence of values ranging from $t - k$ to t

when backward LSTM utilizes input sequence range from t to $t - k$. The result of the BiLSTM layer is attained by using Eq. (20),

$$Y_t = \sigma(\vec{h}_t, \vec{h}_t) \tag{20}$$

Here, the sigmoid function unified the results of unidirectional LSTM networks. The BiLSTM utilizes the dropout method for regulating the over-fitting.

IV. EXPERIMENTAL RESULTS

The performance of developed technique is simulated by using python environment and used system configurations are i5 processor, 8 GB RAM and windows 10 (64 bit) The evaluation metrics utilized to assess performance are accuracy, recall, f1-score, precision and AUC/ROC. The True Negative (TN) and True Positive (TP) values represents the ability of classifier method for predicting the presence or absence of sepsis in patient. The False Negative (FN) and False Positive (FP) values represents incorrect predictions identified by methods. The accuracy represents ratio of actual positive observations to total number of positive instances. The recall executes whole fraction of positive instances. The f1-score calculates mean of precision and recall. The mathematical formula for evaluation metrics is given from Eq. (21) to (24),

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \tag{21}$$

$$Precision = \frac{TP}{TP+FP} \times 100 \tag{22}$$

$$Recall = \frac{TP}{TP+FN} \times 100 \tag{23}$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100 \tag{24}$$

In Table I, the feature selection algorithm CL-LBO is evaluated with different metrics on the MIMIC III dataset. The Honey Badger Optimization (HBO), Ant Colony Optimization (ACO), and Squirrel Search Algorithm (SSA) are the other feature selection algorithms considered to evaluate performance of CL-LBO algorithm. The developed feature selection algorithm reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% AUC when compared with other algorithms.

TABLE I. PERFORMANCE OF FEATURE SELECTION ALGORITHM

Feature selection algorithms	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
HBO	94.75	94.60	94.50	94.55	94.85
ACO	95.65	95.50	95.40	95.45	95.70
DSSA	96.85	96.70	96.60	96.65	96.90
CL-LBO	99.85	99.60	99.50	99.55	99.95

In Table II, the performance of classifier is evaluated using whole feature set with different metrics on the MIMIC III dataset. The RNN, Gated Recurrent Unit (GRU), and LSTM are the other classifiers considered to evaluate the performance of the CNN-LSTM network. The classifier with whole feature set reached 96.45% accuracy, 96.58% precision, 96.98% recall, 96.51% f1-score, and 96.45% AUC.

TABLE II. PERFORMANCE OF CLASSIFIER USING WHOLE FEATURE SET

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
RNN	93.50	93.60	93.70	93.65	93.40
GRU	94.80	94.85	94.90	94.87	94.75
LSTM	95.60	95.65	95.75	95.70	95.50
Proposed CNN-LSTM	96.45	96.58	96.98	96.51	96.45

In Table III, the performance of the classifier is evaluated using selected relevant features with different metrics on the MIMIC III dataset. The RNN, GRU, and LSTM are the other classifiers considered to evaluate the performance of the CNN-LSTM network. The classifier with selected relevant features reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% AUC when compared to other classifiers.

TABLE III. PERFORMANCE OF CLASSIFIER USING SELECTED RELEVANT FEATURES

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
RNN	95.45	95.20	95.10	95.15	96.10
GRU	96.85	96.50	96.40	96.45	97.20
LSTM	97.75	97.60	97.50	97.55	98.10
Proposed CNN-LSTM	99.85	99.60	99.50	99.55	99.95

In Table IV, the performance of the activation function is evaluated with various metrics on the MIMIC III dataset. The Tanh, Rectified Linear Unit (ReLU) and sigmoid are other activation functions considered to evaluate the performance of the softmax function. The classifier with the softmax function reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% AUC when compared to other activation functions.

TABLE IV. PERFORMANCE OF ACTIVATION FUNCTION

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
Tanh	94.58	94.40	94.30	94.35	94.68
ReLU	95.75	95.60	95.50	95.55	95.89
Sigmoid	96.85	96.70	96.60	96.65	96.99
Softmax	99.85	99.60	99.50	99.55	99.95

In Table V, the performance of the optimizer is evaluated with various metrics on the MIMIC III dataset. The AdaGrad, Adamax, and Stochastic Gradient Descent (SGD) are other optimizers considered to evaluate the performance of the Adam optimizer. The classifier with Adam optimizer reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% AUC when compared to other optimizers.

TABLE V. PERFORMANCE OF OPTIMIZER

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
AdaGrad	94.53	94.47	94.32	94.39	94.61
Adamax	95.76	95.63	95.54	95.58	95.82
SGD	96.87	96.73	96.65	96.69	96.91
Adam	99.85	99.60	99.50	99.55	99.95

In Fig. 3, the accuracy vs. epoch graph is represented for the developed classifier. The training accuracy, after some epochs, is close to 100% represents that the classifier has learned to classify the data accurately. The validation accuracy enhances the training process, represents that it stabilizes the model during training. There is only a small gap between training and validation accuracy representing that the model has better generalization capability and there is no overfitting. The validation accuracy stabilizes the model without any drop. In Fig. 4, the loss vs. epochs is represented for the developed classifier. The training loss continues to decline plateaus at less value, representing that the model has minimized errors in training data. The validation loss minimized in the initial phase represents that the method maximizes its performance on unseen data. After certain epochs, validation loss stabilizes the model. There is less gap between training and validation loss, which represents that the model is not overfitting.

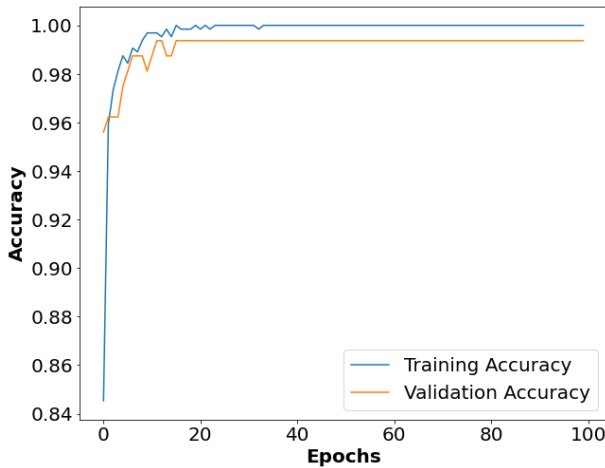


Fig. 3. Accuracy vs. Epochs.

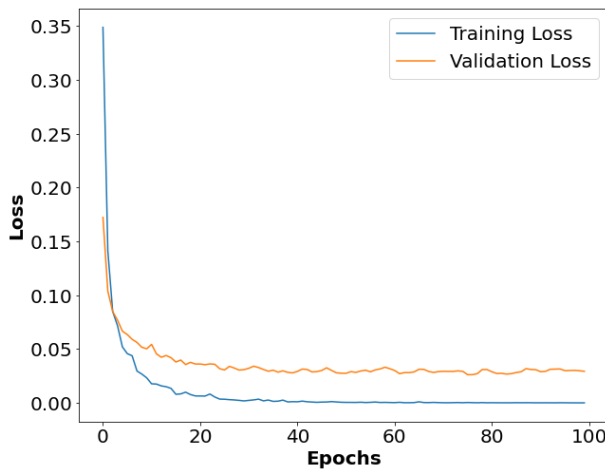


Fig. 4. Loss vs. Epochs.

A. AUC/ROC Curve

By using the AUC/ROC curve, Fig. 5 represents the connection between True Positive Rate (TPR) and False Positive Rate (FPR) by utilizing the AUC/ROC function. The ROC determines the ability of classifier for differentiating their

classes. While AUC is substantial, the prediction of the method is correct.

B. Comparative Analysis

In Table VI, performance of the implemented technique is compared to existing techniques like LSTM [18], Ensemble ML [19], and CNN-Bi-LSTM [23] with different metrics on MIMIC III dataset. The CNN extracts the significant features from relevant features, focused on spatial-based relationships. Then, the Bi-LSTM layer captured the sequential dependencies and temporal relationships in patient histories that are essential to understand the treatment results. CL-LBO integrates the circle chaotic map and Levy flight process in traditional LBO to select relevant features for classification. The proposed technique reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% AUC when compared to existing techniques.

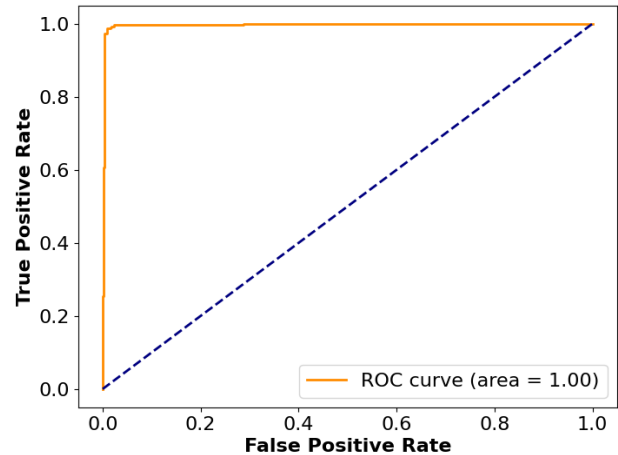


Fig. 5. ROC curve.

TABLE VI. COMPARATIVE ANALYSIS

Methods	Dataset	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	AUC (%)
LSTM [18]	MIMIC III	-	77.00	79.87	78.24	85.01
Ensemble ML [19]		92.80	-	-	-	83.00
CNN-Bi-LSTM [23]		99.15	99.16	99.15	98.85	-
Proposed technique		99.85	99.60	99.50	99.55	99.95

C. Discussion

The outcomes of the proposed technique are evaluated with the MIMIC III dataset using various evaluation metrics. The developed CL-LBO feature selection algorithm is evaluated with different optimization algorithms of HBO, ACO, and DSSA. The developed classifier is evaluated with default features and with selected relevant features. Additionally, the performance of the classifier is evaluated with an activation function and different optimizers. Moreover, the performance of the developed technique is compared with LSTM [18],

Ensemble ML [19], and CNN-Bi-LSTM [23] with different metrics on the MIMIC III dataset. These existing algorithms have drawbacks such as overfitting issues occurring during the process because of the irrelevant features present in the data fit, dimensionality issues, struggled to interpret the pattern of features and the missing values presented in the data caused an overfitting problem. To mitigate these drawbacks, the CNN-Bi-LSTM method for personalized treatment analysis uses EHR data. The CNN extracts the significant features from relevant features, focused on spatial-based relationships. Then, the Bi-LSTM layer captured the sequential dependencies and temporal relationships in patient histories that are essential to understand the treatment results. The SMOTE technique is employed in the pre-processing phase to balance the classes in data. The CL-LBO integrates the circle chaotic map and Levy flight process in traditional LBO to select relevant features for classification. This process improves the process of personalized treatment with high classification accuracy. By using this technique, in this article, the developed technique reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% AUC when compared to existing techniques.

V. CONCLUSION

The traditional mortality risk prediction techniques effectively extract the data in longitudinal EHRs that ignore the difficult relationship and interactions among variables and time dependency in longitudinal records. In the proposed work, the CNN-Bi-LSTM method is developed for personalized treatment analysis using EHR data. The MIMIC III dataset is used in this article and the data is balanced by using SMOTE and the balanced data is encoded by using the one-hot encoding technique. Then, the relevant features from pre-processed data are selected by using the developed CL-LBO algorithm. Here, the circle chaotic map and levy flight process are integrated with the traditional LBO algorithm to enhance the search capability and convergence rate of the algorithm. Then, the CNN extracts significant features from relevant features, focused on spatial-based relationships. Then, the Bi-LSTM layer captured the sequential dependencies and temporal relationships in patient histories that are essential to understand the treatment results. The proposed method reached 99.85% accuracy, 99.60% precision, 99.50% recall, 99.55% f1-score, and 99.95% AUC when compared to LSTM. In the future, different DL-based will be used to further enhance the process of personalized treatment analysis.

REFERENCES

- [1] W. Wang, P. Mohseni, K. L. Kilgore, and L. Najafizadeh, "PulseDB: A large, cleaned dataset based on MIMIC-III and VitalDB for benchmarking cuff-less blood pressure estimation methods," *Front. Digit. Heal.*, vol. 4, p.1090854, February 2023, doi: 10.3389/fgth.2022.1090854.
- [2] R. Zhang et al., "Independent effects of the triglyceride-glucose index on all-cause mortality in critically ill patients with coronary heart disease: analysis of the MIMIC-III database," *Cardiovasc. Diabetol.*, vol. 22, no. 1, p.10, January 2023, doi: 10.1186/s12933-023-01737-3.
- [3] J. Xu, H. Cai, and X. Zheng, "Timing of vasopressin initiation and mortality in patients with septic shock: analysis of the MIMIC-III and MIMIC-IV databases," *BMC Infect. Dis.*, vol. 23, p.199, April 2023, doi: 10.1186/s12879-023-08147-6.
- [4] S. Peng, J. Peng, L. Yang, and W. Ke, "Relationship between serum sodium levels and all-cause mortality in congestive heart failure patients: A retrospective cohort study based on the MIMIC-III database," *Front. Cardiovasc. Med.*, vol. 9, p.1082845, January 2023, doi: 10.3389/fcvm.2022.1082845.
- [5] W. Liao and J. Voldman, "A Multidatabase ExTRaction PipELine (METRE) for facile cross validation in critical care research," *J. Biomed. Inform.*, vol. 141, p. 104356, May 2023, doi: 10.1016/j.jbi.2023.104356.
- [6] Y. Hakverdi et al., "Enhancing ICU Management and Addressing Challenges in Türkiye Through AI-Powered Patient Classification and Increased Usability With ICU Placement Software," *IEEE Access*, vol. 12, pp. 146121–146136, 2024, doi: 10.1109/access.2024.3426919.
- [7] J. Chen, T. Di Qi, J. Vu, and Y. Wen, "A deep learning approach for inpatient length of stay and mortality prediction," *J. Biomed. Inform.*, vol. 147, p. 104526, November 2023, doi: 10.1016/j.jbi.2023.104526.
- [8] S. Wei et al., "Machine learning-based prediction model of acute kidney injury in patients with acute respiratory distress syndrome," *BMC Pulm. Med.*, vol. 23, no. 1, p.370, October 2023, doi: 10.1186/s12890-023-02663-6.
- [9] N. Ashrafi, Y. Liu, X. Xu, Y. Wang, Z. Zhao, and M. Pishgar, "Deep learning model utilization for mortality prediction in mechanically ventilated ICU patients," *Informatics Med. Unlocked*, vol. 49, p. 101562, 2024, doi: 10.1016/j.imu.2024.101562.
- [10] L. Liu, O. Perez-Concha, A. Nguyen, V. Bennett, and L. Jorm, "Automated ICD coding using extreme multi-label long text transformer-based models," *Artif. Intell. Med.*, vol. 144, p. 102662, October 2023, doi: 10.1016/j.artmed.2023.102662.
- [11] M. Bernardini, A. Doynychko, L. Romeo, E. Frontoni, and M.-R. Amini, "A novel missing data imputation approach based on clinical conditional Generative Adversarial Networks applied to EHR datasets," *Comput. Biol. Med.*, vol. 163, p. 107188, September 2023, doi: 10.1016/j.compbiomed.2023.107188.
- [12] S. S. Samy, S. Karthick, M. Ghosal, S. Singh, J. S. Sudarsan, and S. Nithiyanantham, "Adoption of machine learning algorithm for predicting the length of stay of patients (construction workers) during COVID pandemic," *Int. J. Inf. Technol.*, vol. 15, no. 5, pp. 2613–2621, June 2023, doi: 10.1007/s41870-023-01296-6.
- [13] M. Fathima Begum and S. Narayan, "A pattern mixture model with long short-term memory network for acute kidney injury prediction," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 4, pp. 172–182, April 2023, doi: 10.1016/j.jksuci.2023.03.007.
- [14] C. Yu and Q. Huang, "Towards more efficient and robust evaluation of sepsis treatment with deep reinforcement learning," *BMC Med. Inform. Decis. Mak.*, vol. 23, no. 1, p.43, March 2023, doi: 10.1186/s12911-023-02126-2.
- [15] A. Ahmed, X. Zeng, R. Xi, M. Hou, and S. A. Shah, "MED-Prompt: A novel prompt engineering framework for medicine prediction on free-text clinical notes," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 36, no. 2, p. 101933, February 2024, doi: 10.1016/j.jksuci.2024.101933.
- [16] X. Li, Y. Zhang, X. Li, H. Wei, and M. Lu, "DGCL: Distance-wise and Graph Contrastive Learning for medication recommendation," *J. Biomed. Inform.*, vol. 139, p. 104301, March 2023, doi: 10.1016/j.jbi.2023.104301.
- [17] H. Dong et al., "Ontology-driven and weakly supervised rare disease identification from clinical notes," *BMC Med. Inform. Decis. Mak.*, vol. 23, no. 1, p.86, May 2023, doi: 10.1186/s12911-023-02181-9.
- [18] F. Yang, J. Zhang, W. Chen, Y. Lai, Y. Wang, and Q. Zou, "DeepMPM: a mortality risk prediction model using longitudinal EHR data," *BMC Bioinformatics*, vol. 23, no. 1, p.423, October 2022, doi: 10.1186/s12859-022-04975-6.
- [19] N. Tasnim, S. Al Al Mamun, M. S. Shahidul Islam, M. S. Kaiser, and M. Mahmud, "Explainable Mortality Prediction Model for Congestive Heart Failure with Nature-Based Feature Selection Method," *Appl. Sci.*, vol. 13, no. 10, p. 6138, May 2023, doi: 10.3390/app13106138.
- [20] M. Bampa, I. Miliou, B. Jovanovic, and P. Papapetrou, "M-ClustEHR: A multimodal clustering approach for electronic health records," *Artif. Intell. Med.*, vol. 154, p. 102905, August 2024, doi: 10.1016/j.artmed.2024.102905.
- [21] B. Alsinglawi et al., "An explainable machine learning framework for lung cancer hospital length of stay prediction," *Sci. Rep.*, vol. 12, no. 1, p.607, January 2022, doi: 10.1038/s41598-021-04608-7.

- [22] S. Niu, J. Ma, L. Bai, Z. Wang, L. Guo, and X. Yang, "EHR-KnowGen: Knowledge-enhanced multimodal learning for disease diagnosis generation," *Inf. Fusion*, vol. 102, p. 102069, February 2024, doi: 10.1016/j.inffus.2023.102069.
- [23] S. Sakri et al., "Sepsis Prediction Using CNNBDLSTM and Temporal Derivatives Feature Extraction in the IoT Medical Environment," *Comput. Mater. & Contin.*, vol. 79, no. 1, pp. 1157–1185, 2024, doi: 10.32604/cmc.2024.048051.
- [24] S. R. Khope and S. Elias, "Simplified & Novel Predictive Model using Feature Engineering over MIMIC-III Dataset," *Procedia Comput. Sci.*, vol. 218, pp. 1968–1976, 2023, doi: 10.1016/j.procs.2023.01.173.
- [25] C. Liu et al., "Early prediction of MODS interventions in the intensive care unit using machine learning," *J. Big Data*, vol. 10, no. 1, p.55, May 2023, doi: 10.1186/s40537-023-00719-2.
- [26] V. K. Chauhan, A. Thakur, O. O'Donoghue, O. Rohanian, S. Molaei, and D. A. Clifton, "Continuous patient state attention model for addressing irregularity in electronic health records," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, p.117, May 2024, doi: 10.1186/s12911-024-02514-2.
- [27] H.-J. Lee et al., "StrokeClassifier: ischemic stroke etiology classification by ensemble consensus modeling using electronic health records," *npj Digit. Med.*, vol. 7, no. 1, p.130, May 2024, doi: 10.1038/s41746-024-01120-w.
- [28] A. Johnson, T. Pollard, and R. Mark, "MIMIC-III Clinical Database." *PhysioNet*, 2023, doi: 10.13026/C2XW26.
- [29] J. He, Y. Hao, and X. Wang, "An Interpretable Aid Decision-Making Model for Flag State Control Ship Detention Based on SMOTE and XGBoost," *J. Mar. Sci. Eng.*, vol. 9, no. 2, p. 156, February 2021, doi: 10.3390/jmse9020156.

Spam Detection Using Dense-Layers Deep Learning Model and Latent Semantic Indexing

Yasser D. Al-Otaibi¹, Shakeel Ahmad², Sheikh Muhammad Saqib³

Department of Information Systems-Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Jeddah 21589, Saudi Arabia¹

Department of Computer Science-Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University, Jeddah 21589, Saudi Arabia²

Department of Computing and Information Technology, Gomal University, D.I.KHAN, Pakistan³

Abstract— In the digital age, online shoppers heavily depend on product feedback and reviews available on the corresponding product pages to guide their purchasing decisions. Feedback is used in sentiment analysis, which is helpful for both customers and company management. Spam feedback can have a negative impact on high-quality products or a positive impact on low-quality products. In both cases, the matter is bothersome. Spam detection can be done with supervised or unsupervised learning methods. We suggested two direct methods to detect feedback orientation as 'spam' or "not spam", also called "ham," using the deep learning model and the LSI (Latent Semantic Indexing) technique. The first proposed model uses only dense layers to detect the orientation of the text. The second proposed model uses the concept of LSI, an effective information retrieval algorithm that finds the closest text to a provided query, i.e., a list containing spam words. Experimental results of both models using publicly available datasets show the best results (89% accuracy and 89% precision) when compared to their corresponding benchmarks.

Keywords—Spam; supervised learning methods; unsupervised learning methods; LSI; dense; deep learning

I. INTRODUCTION

The company is shocked when there is a complaint about their high-quality product. How did they find the complaint? Obviously, from the webpage concerned with the feedback of the customer, if a complaint is factual, it can be found in a product; if not, these types of comments degrade the product's star rating or sentiment score. Before sentiment analysis, these comments or reviews should be filtered. These reviews can be considered spam. Researchers have done much work to separate spam from the given reviews [1]. Most spam is delivered as emails, so there should be a strong filter to detect the spam.

Classifiers trained using a combination of features are more effective than those learned using only one type of feature [2]. A machine learning algorithm has also been investigated for filtering spam emails [3]. Most supervised machine learning algorithms are not suitable for spam detection due to the lack of features or words that indicate the hint that the review is actual or not [4]. Although Support Vector Machines (SVM) is an important and powerful technique for detecting reviews as spam. However, for big data, the efficiency of SVM is reduced because of the many data processing complexities [5].

Unsupervised learning algorithms has also been investigated for spam detection such as clustering algorithm [6] has proved that these algorithms are well suited for spam clustering. A novel unsupervised text mining model was developed and integrated into a semantic language model for detecting untruthful reviews [7]. Unsupervised methods are currently unable to match the performance of supervised learning methods, research is limited, and results are inconclusive, warranting further investigation [8].

Deep learning is a new trend, through which classification can be done in a very descent way. This learning can be learned automatically, without predefined knowledge explicitly coded by the programmers. Although previous work which has been done on spam detection is based on bidirectional LSTM (Long Short Term Memory) [9] a resource hunger technique. The proposed work develops a simple sequential dense layer's model using scaling of data to detect spam text and found best performance, which require less computational processing.

The supervised learning approach employs training data based on labels ("ham" and "spam") sent to a classifier, which detects "spam" using this learned corpus. Unsupervised learning, on the other hand, necessitates the discovery of rules and patterns from supplied data. Both techniques need a significant amount of work, but, in this case, the suggested methodology does not necessitate the use of training data or rules. Latent Semantic Indexing was used to filter the "spam" and "ham" (not-spam), reviews. LSI is simple to comprehend, execute, and employ. When compared to other approaches, the results of LSI are far more precise and speedier. It seeks the most representational, rather than the most discriminative, qualities for document representation [10] The manually compiled Spam Words (SW) list includes 956 entries, which might be used in spam reviews [11]. This study makes the following key contributions:

- A method has been suggested to make a sequential deep learning model with scaling that can tell when text is spam.
- A method is proposed to detect reviews as either "spam" or "ham" using Latent Semantic Indexing (LSI) with an Automatic Generated Query (AGQ) that serves as a major input to LSI. Hence, there is no need to provide a separate query.

A. Significance of the Study

This study emphasizes the critical importance of maintaining the integrity of customer feedback systems by filtering spam reviews. Genuine feedback is essential for ensuring accurate product sentiment scores and star ratings. By addressing the issue of spam detection, the research highlights how filtering out non-factual reviews can prevent negative impacts on product perception. Additionally, the study advances the field by exploring modern machine learning and deep learning techniques, demonstrating their potential to improve spam detection accuracy while overcoming computational challenges in handling large datasets. These innovations contribute to more reliable systems for analyzing customer reviews.

B. Key Contributions

The study makes several notable contributions to the field of spam detection. First, it proposes a sequential deep learning model with data scaling, which not only achieves high performance but also requires significantly lower computational resources compared to traditional methods. Second, it introduces the integration of Latent Semantic Indexing (LSI) with an Automatic Generated Query (AGQ), enabling the filtering of "spam" and "ham" reviews without the need for a predefined query. This innovation simplifies the detection process and improves efficiency. Lastly, the research demonstrates that LSI outperforms existing approaches by providing more precise and faster results, focusing on representational features for document representation rather than solely discriminative ones.

C. Research Gap

Despite substantial progress in spam detection, significant gaps remain in existing methodologies. Supervised machine learning methods often fail to achieve optimal results due to insufficient features or words that indicate the authenticity of reviews. Additionally, Support Vector Machines (SVM), though recognized as a powerful technique, face reduced efficiency when dealing with large datasets because of data processing complexities. While unsupervised learning methods, such as clustering algorithms, have shown potential, their performance is still inferior to supervised techniques, and research in this area is limited and inconclusive. Furthermore, current deep learning approaches, such as bidirectional LSTMs, are resource-intensive, highlighting the need for simpler yet effective models that can be practically implemented for large-scale spam detection tasks.

II. RELATED STUDY

Deep learning has gained significant prominence across various research domains, including applications in natural image processing [12], electricity theft detection [13], diagnosis of human and animal diseases [14][15], and sentiment analysis [16].

For Sentiment analysis, filtration of objective reviews not necessary but also filtration of spam reviews will also increase the accuracy of sentiment analysis. With respect to sentence polarity, there is lot of studies which is about determining the sentiment orientation of a review or comment [17]. Sentiment orientation means that a positive opinion will be an exact

positive, and a negative opinion will be an exact negative [18]. The view, assessment or feeling of a person towards a product, aspect [19], or service is known as a sentiment. Most of the work on reviews or blogs is based on sentiment analysis based on binary classification i.e. positive or negative classes [20]. As text classification is done using machine learning based [21], deep learning based [22] and score based approaches [23]. Training data is used in machine learning and deep learning approaches while different rules based on attributes and entities are used in other methods. To find polarity of opinion based on aspects, lot of researches has been done to extract aspect and aspect based sentiment analysis [24]. Besides machine learning, lot of sentiment analysis work has also been done by deep learning from different dimensions [25]. Work of [26] used word2vec to reduce number of parameters by considering of bag of words in deep learning. Authors [27] investigated the impact on performance over multiple runs by changing hyper parameters for convolutional neural network. OpCNN model based on k-max pooling was presented in [28] by considering word order problem of Chinese. Sentiment classification on tweets to detect tweet as either positive or negative was implemented by LSTM neural network [29].

Different research techniques are available for spam filtration such as filtering Technique based on content [30], spam filtering technique based on heuristic rules [31], spam filtering technique based on previous likeness [32], adaptive spam detection [33] etc. There are many proposed email classification techniques, which detect the spam emails such as case-based technique [9], ANN (Artificial Neural Networks) [34] and SVM (Support-Vector-Machine) [35]. For this purpose LSI (Latent Semantic Indexing) is better [36]. For clustering purpose LSI has also been considered to filter unwanted emails in Chinese and English [37][38].

III. PROPOSED METHODOLOGY

Generally, tasks of proposed work consist of different steps shown in Fig. 1. Manual Feature extraction work excluded in deep learning because it is responsibility of deep learning model-training to handle it automatically.

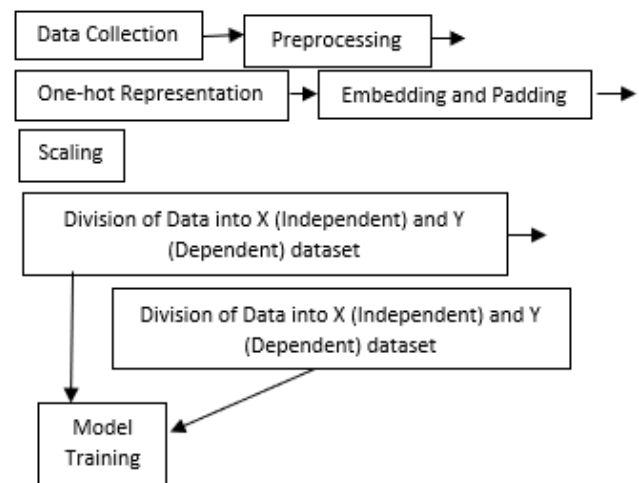


Fig. 1. Steps of proposed work.

The used dataset consists of two columns; one contains text, and the other has a decision as spam or ham [2]. The data set contains 20718 texts, of which 10369 are spam and 10349 are ham. The preprocessing process removes irrelevant opinions, duplicate words, extra spaces, and stop words. It also involves tokenization, converting all words into lower cases, contractions, stemming, and lemmatization. The one-hot encoding process is used to convert categorical variables into a form that the machine can easily read. The one-hot encoding performs better in prediction. The proposed work package from TensorFlow. The next step is embedding and padding, where a large sparse vector represents each word with a score (representing an entire vocabulary), and padding adds zeros at the end or start of the sequence to make the sample the same size as the sequence. The proposed work uses embedding and pad sequences packages from TensorFlow.

To train proposed model, we require training data which is a complete set of dependent (Y) and independent (X) variables, across a model can learn. Proposed model has used train_test_split package for this purpose.

A. Model Training on Dataset

This model consists of only dense layers. The name indicates that layers are fully connected through the neurons in a network layer. Each neuron in a layer collects input from all the neurons that appeared in the previous layer, thus, they are densely attached. Fitting of model is shown in Fig. 2.

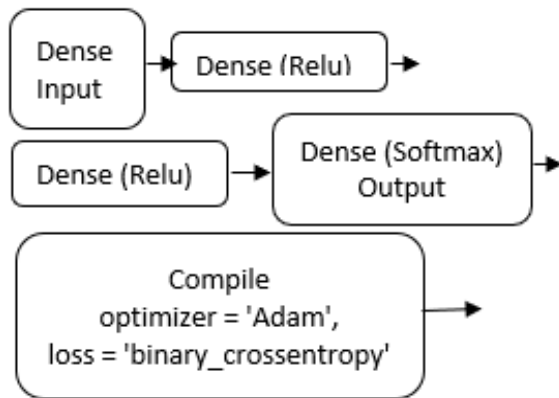


Fig. 2. Model structure of dense layers.

B. Training of Model without Scaling

A simple sequential model has been created with four dense layers. First and last layers are input and output layers and remaining two are hidden layers. Description of model is given below in Table I.

TABLE I. SUMMARY OF MODEL

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 400)	160400
dense_1 (Dense)	(None, 20)	8020
dense_2 (Dense)	(None, 15)	315
dense_3 (Dense)	(None, 1)	16
Total params: 168751		
Trainable params: 168751		
Non-trainable params: 0		

First three layers are using activation function 'relu' and output layer is using activation function 'softmax'. A 'binary_crossentropy' loss function is used, because 'spam' or 'ham' is a binary problem. Proposed model has been compiled using 'adam' optimizer, because we are using batch option and there also there is neither 'vanishing gradient problem' will occur nor 'dead neurons' will occur. Different attempts have been made to achieve high accuracy based on epoch and batch size. But achieved 69% at 100 epochs with 40 batch-size shown in Table II.

TABLE II. ACCURACIES OF MODEL (WITHOUT SCALING) WITH DIFFERENT ATTEMPTS

Epochs	Batch Size	Accuracy
10	40	0.63
50	40	0.68
100	40	0.87
100	100	0.86

C. Training of Model with Scaling

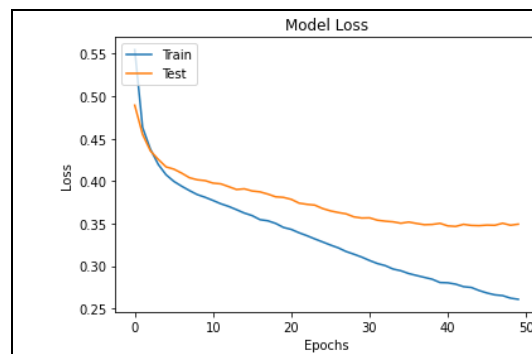
Here scaling concept is used to enhance the accuracy.

To normalize the range of independent variables, scaling feature is used. Its basic purpose is to convert the whole independent variables into range 0 to 1, because this range is very suitable for deep learning models [39].

TABLE III. ACCURACIES OF MODEL (WITH SCALING) WITH DIFFERENT ATTEMPTS

Epochs	Batch Size	Accuracy
10	40	0.80
50	40	0.88
50	100	0.84
Previous Work[28]	0.82	0.82

Different attempts have been made to achieve high accuracy based on epoch and batch size, shown in Table III. And achieved 88% at 100 epochs with 40 batch-size. Remaining measures of confusion matrix are given below in Table IV. Performance of model with respect to loss and with respect to accuracy is shown in Fig. 3.



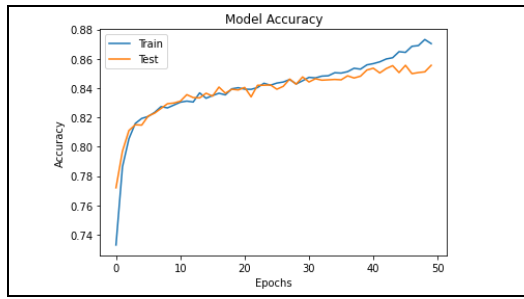


Fig. 3. Performance of model with respect to Loss and Accuracy.

TABLE IV. DIFFERENT MEASURES OF CONFUSION MATRIX

Proposed Work	Precision	Recall	F1-score
0	0.93	0.83	0.87
1	0.80	0.92	0.86
weighted avg	0.87	0.88	0.87
Previous Work[28]	0.82	0.78	0.80

opCNN [28] achieved 84% accuracy while proposed deep learning model achieved 88% accuracy. Furthermore, since deep learning models require a lot of resources such as Keras, TensorFlow, Activation Functions, etc., further supervised and unsupervised machine learning models also earn unsatisfactory accuracy for spam detection. In contrast, the proposed LSI technique is comparatively effective.

IV. SPAM DETECTION USING LSI

Major inputs to the proposed model are "reviews" and "automatically generated queries" (AGQ). After processing through LSI, the output is measured in terms of scores. A decision based on these scores will be made, i.e., whether the review is spam or not. Here, the classification category is "spam" and "ham" (not spam). The decision depends upon the pivot value; in the result section, we made different attempts to find the pivot value. If the score of each review is greater than the pivot value, it will be considered "spam," otherwise "ham." The whole process is depicted in Fig. 4.

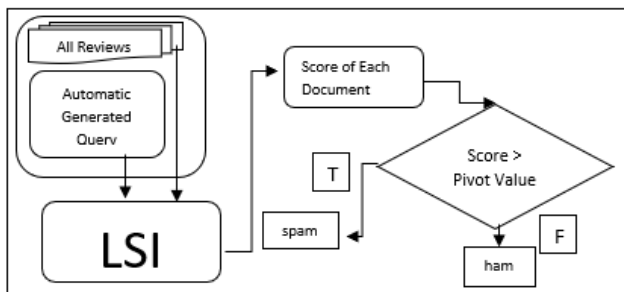


Fig. 4. Proposed framework.

A. Latent Semantic Indexing

The LSI proposed by [40] is an efficient information retrieval algorithm. In LSI, there is a cosine similarity measurement between the coordinates of a document vector and the coordinates of a query vector. The result of cosine similarity measurement "1" means the document is 100% and the result of cosine similarity measurement "0" means the

document is very far from a query. A feature matrix from the frequencies of all words in the documents and query will be formed, and singular value decomposition (SVD) will be calculated from this matrix. Singular value decomposition (SVD) can be used to determine the coordinates of the documents and the query. Three matrices, S, V, and U can be easily extracted from SVD. The document coordinates will be determined from S and V as depicted in the algorithm shown in Fig. 5. Finally, a cosine similarity function is applied to these coordinates to find the texts that best match the spam query [41]. Based on LSI techniques algorithm for proposed work is shown in Table V.

TABLE V. ALGORITHM FOR SPAM DETECTION USING LSI

Function LSI (AllReviews, AGQ)
1. Matrixf: Frequency Matrix from AllReviews
2. Matrixq: Query Matrix from AGQ from List of Spam words and Reviews
3. V, S, U = numpy.linalg.svd(Matrixf)
4. UK = Rank 2 Approximation of U
5. VK = Rank 2 Approximation of V
6. SK = Rank 2 Approximation by taking two columns and two rows of S
7. CoorR: Each row of V relates to Coordinates of a Review
8. Query Coordinates: Coorq = (Matrixq)TUKSk-1
9. Find dot product of Coorq with each document coordinates CoorR
10. $U_{x=1}^m(CoorR, Coorq) = \frac{\sum_{i=1}^n CoorR(i) * Coorq(i)}{\sqrt{\sum_{i=1}^n (CoorR(i))^2} \sqrt{\sum_{i=1}^n (Coorq(i))^2}}$
11. Return (Score of all Documents)
End Function

B. Preprocessing

First of all, it is very necessary to remove noise from the reviews, Eq. (1), Eq. (2), Eq. (3) and Eq. (4) are used to filter the reviews based on stop words and negations.

$$R = U_{x=1}^n R_x \quad (1)$$

$$C(x) = U_{i=1}^n R_i \quad (2)$$

$$FC(x) = U_{i=1}^n \{Antonyme(C_i), \text{ if } C_{i-1} \notin Negations\} \quad (3)$$

$$FC(x) = U_{i=1}^n \{T_i, \text{ if } T_i \notin StopW\} \quad (4)$$

where $x = 1, 2, 3...n$, StopW means stop words, R represents the total number of reviews, C(x) represents the chunks of the xth reviews, and FC(x) represents the filtered chunks of the xth reviews.

C. AGQ (Automatic Generated Query)

All reviews I and a list of spam words (SW already identified in the introduction) are major inputs for AGQ. The intersection of chunks of each review and spam words (SW) will be determined. Then this list will be updated with AGQ as a union. If a chunk does not belong to SW, then it will be checked in the dictionary (WordNet). If this chunk is not present in the dictionary, then it will also be added in AGQ as a union. Because sometimes spam reviews also contain meaningless words. Since the motive of the proposed model is to find those reviews close to spam words, i.e., AGQ, the whole process is shown in Fig. 6, and the creation of AGQ has been portrayed in Eq. (5) and Eq. (6).

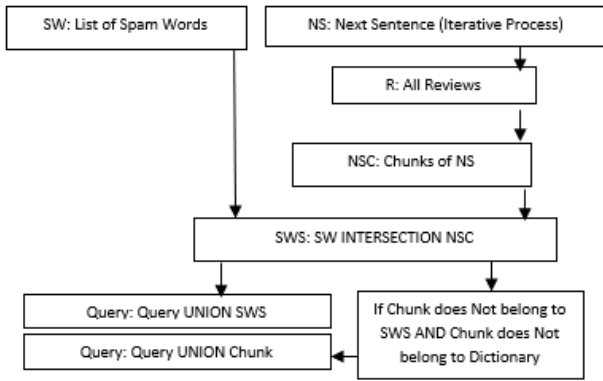


Fig. 5. Process for generating automatic query.

$$AGQ = \bigcup_{x=1}^n \{FC(x)_i, \bigcup_{i=1}^n. \text{if } FC(x)_i \in SW \quad (5)$$

$$AGQ = AGQ \text{ UNION } \bigcup_{x=1}^n \{FC(x)_i, \bigcup_{i=1}^n. \text{if } FC(x)_i \neq SW \text{ AND } FC(x)_i \neq \text{WordNet} \quad (6)$$

where $x = 1, 2, 3 \dots n$ and $FC(x)_i$ means i th words of x th review. AGQ contains those words from all reviews which belongs to SW (spam words) and those which does not belong to wordnet dictionary.

D. Scoring

Already we have determined $FC(x)$ and AGQ. Now Eq. (7) will find the score of each review $FC(x)$ with spam-query AGQ.

$$LSI(\text{Score})_x = \bigcup_{x=1}^n (LSI_x(FC(x), AGQ)) \quad (7)$$

Decision

If LSI score is greater than pivot value, it means review is ‘spam’ because it is closest with spam query otherwise considered as ‘ham’. Following equation Eq. (8) is used for filtering purpose.

$$R_{\text{decision}}(x) = \bigcup_{x=1}^n \begin{cases} \text{spam}_x, & \text{if } (LSI(\text{Score})_x) > \text{Pivot Value} \\ \text{ham}_x, & \text{else} \end{cases} \quad (8)$$

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

The SMS Spam Collection is a set of SMSs tagged messages that have been collected for SMS Spam research [42]. It consists of column v1 and v2. Column v1 contains 5,574 English messages with label ‘ham’ and ‘spam’ and column v2 contains the text of message. The sample listing of the said datasets is presented in Table VI.

TABLE VI. SAMPLE SMS FROM DATASET

v1	v2
ham	Absolutely wonderful – silky and sexy and comfortable
ham	This dress is perfection! So pretty and flattering.
Ham	Super cute and comfy pull over. Sizing is accurate. Material has a little bit of stretch.
Ham	Loved this top and was really happy to find it on sale!
Spam	100 dating service call 09064012103 box334sk38ch
spam	FREE entry into our 250 weekly competition just text the word WIN to 80086 NOW 18 T&C www.txttowincouk
spam	XXXXMobileMovieClub To use your credit click the WAP link in the next txt message or click here httpwap xxxmobilemovieclubcom?n=QJKGIGHJJGCBL
spam	500 New Mobiles from 2004 MUST GO Txt NOKIA to No 89545 & collect yours today From ONLY 1 www.4tbiz 2optout 08718726270150gbpmtmsg18

A confusion matrix [43] is formed from the four outcomes produced as a result of binary classification. A binary classifier predicts all data instances of a test dataset as either ‘spam’ or ‘ham’. This classification (or prediction) produces four outcomes -true spam (TS), -false spam (FS), -true ham (TH) and -false ham (FH).

Here, in start 0.7 was considered as pivot value, then achieved accuracy was 84%, at pivot values 0.8 & 0.9 accuracy was 88% while at 0.99 accuracy was 54%. So, 0.8 or 0.9 can be considered as pivot value as shown in Fig. 7. Graphs of Confusion matrices at different scores in Fig. 6, also predict that values greater or equal to 0.8 and less or equal to 0.9 can be considered as pivot value.

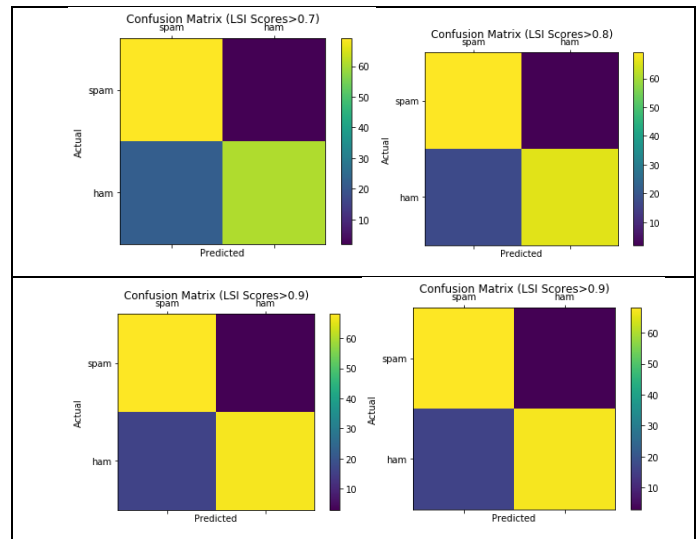


Fig. 6. Confusion matrices at all selected scores.

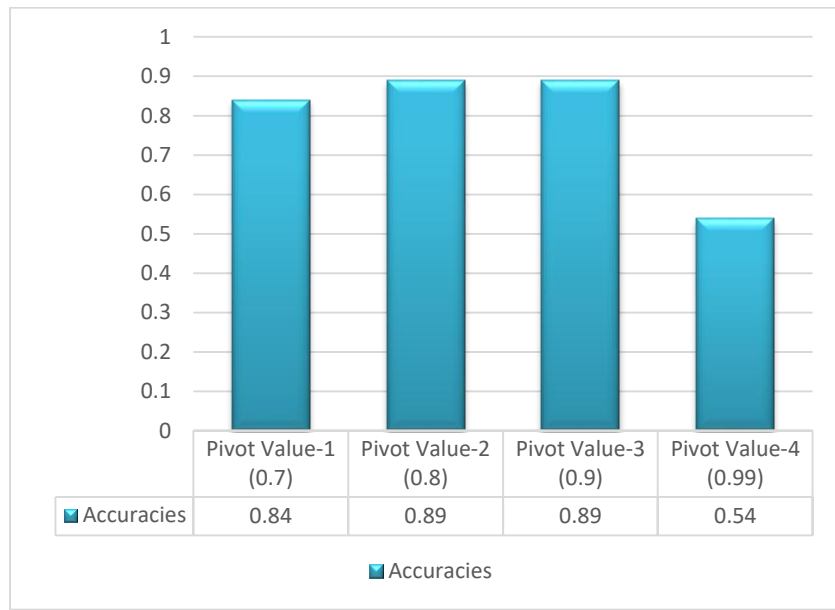


Fig. 7. Accuracies at different scores.

Table VII shows some of sampled documents based on proposed method to detect text as ‘ham’ or ‘spam’ using LSI-score greater than pivot values 0.8 and 0.9.

TABLE VII. SAMPLE OF DOCUMENTS WITH LSI SCORES GREATER THAN 0.8 AND 0.9

S. No	Reviews	LSI Score	Actual	Detected Based on Score > 0.8	Detected Based on Score > 0.9
1	d[0]	0.99982	Ham	Spam	spam
2	d[1]	0.337082	Ham	Ham	ham
3	d[2]	0.492025	Ham	Ham	ham
4	d[3]	0.491148	Ham	Ham	ham
5	d[4]	0.989738	Ham	Spam	spam
6	d[5]	0.551706	Ham	Ham	ham
7	d[6]	-0.18123	Ham	Ham	ham
8	d[7]	-0.21559	Ham	Ham	ham
9	d[8]	0.271523	Ham	Ham	ham
10	d[9]	0.431423	Ham	Ham	ham
20	d[82]	0.975451	Spam	Spam	spam
21	d[83]	0.976627	Spam	Spam	spam
22	d[84]	0.976538	Spam	Spam	spam

Now at detected pivot values, proposed model achieved 0.89 precision and 0.88 recall as shown in Table VIII.

TABLE VIII. STATISTICAL RESULTS AT DIFFERENT PIVOT VALUES

Pivot Values	Class	Precision		Recall		F1-Score
0.8	Ham	0.97	0.80	0.80	0.80	0.87
	Spam	0.80	0.80	0.97	0.80	0.88
	Avg	0.89	0.80	0.88		0.88
0.9	Ham	0.96		0.81		0.88
	spam	0.81		0.96		0.88
	avg	0.89		0.88		0.88

Recently accuracies of supervised learning approaches have been increased, while unsupervised approaches are still working on increasing the efficiency. Because major used source are spam words, which are not only present in spam-text while also in ham-text. Table IX is showing that proposed model gained high performance with respect Supervised, Unsupervised, Combined and Active Learning.

TABLE IX. COMPARISON WITH DIFFERENT APPROACHES

Methods	Precision	Accuracy
Supervised Learning Methods	49%	78%
Unsupervised Learning Methods	42%	80%
Combined Approach	64%	83%
Adaptive Resonance Theory (ART)	75%	89%
Active Learning	87%	88%
Proposed Work	89%	89%

VI. CONCLUSION

Spam is a serious issue that is not just annoying to the end-user but also financially damaging and security risks. In this paper, state-of-the-art models and LSI model were experimented against the task of detecting spam emails. To validate the generalization capabilities of the proposed method, its experimental results have been compared with CNN and OPCNN models. CNN achieved 82% and opCNN achieved 84% accuracy while proposed deep learning model achieved 88% accuracy. Although accuracy of proposed model is less than the accuracy of another base line whose accuracy is 96% (Spam with Deep Model). But this work uses bidirectional LSTM, which is computationally expensive and also uses reviews with sequence length less than 300 with dataset length 5000. Proposed work has been implemented on

20718 lengths of dataset also containing reviews with length more than 300.

Another less computationally expensive proposed model has been implemented using LSI concept to detect ‘spams’ from given data. The major theme of this work is to avoid laborious work for detecting patterns and making rules and implementation from machine learning methods. Based on the experimental results through confusion matrix, it found that results generated from the proposed method show a significant improvement from existing techniques related not only to precision and accuracy, but also to recall and f1-score which are 88% shown in Table IX. Fig. 8 shows that the proposed work provides better results than previous approaches.

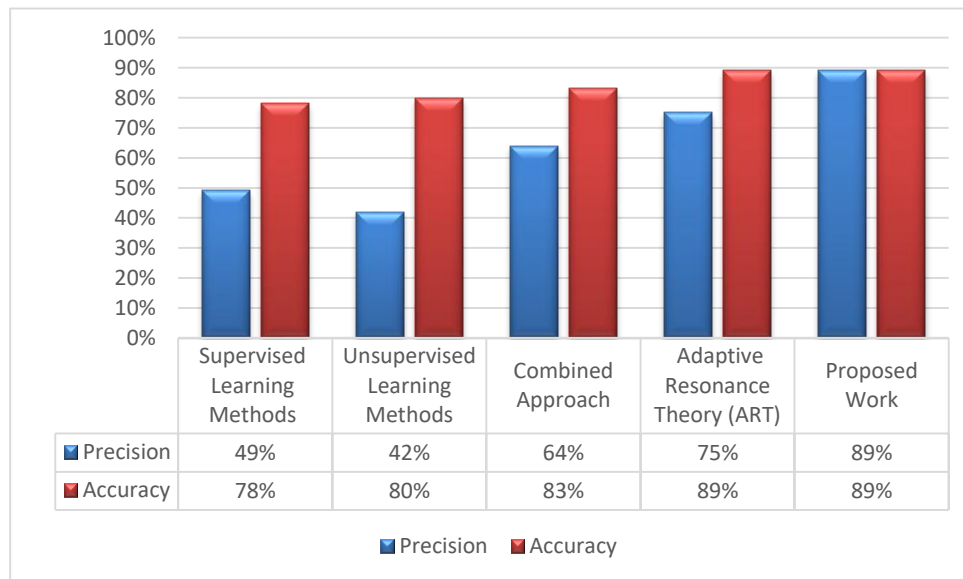


Fig. 8. Comparison of proposed LSI work with alternative approaches.

VII. LIMITATIONS AND FUTURE WORK

This study demonstrates significant progress in spam detection; however, certain limitations remain that open avenues for future research. The dataset used in this study comprises 5,574 English-language messages, which, while effective for the current analysis, limits the generalizability of the findings. Future work could focus on increasing the dataset size and incorporating data from other languages to improve model robustness and applicability across diverse linguistic contexts. Additionally, the automatic query process in this work relies on the WordNet dictionary. While effective, the exploration of other dictionaries or lexical resources could enhance the flexibility and accuracy of the query generation process. The pivot value of 0.7, achieved with the current dataset size, may vary with larger or smaller datasets, suggesting the need for further investigation into optimal pivot values for datasets of different sizes.

ACKNOWLEDGMENT

This project was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University, Jeddah, under

Grant No. 94-830-1442. The authors, therefore, acknowledge with thanks DSR for technical and financial support.

REFERENCES

- [1] A. Qazi, N. Hasan, R. Mao, M. E. M. Abo, S. K. Dey, and G. Hardaker, "Machine Learning-Based Opinion Spam Detection: A Systematic Literature Review," IEEE Access, 2024, doi: 10.1109/ACCESS.2024.3399264.
- [2] H. Khan, M. U. Asghar, M. Z. Asghar, G. Srivastava, P. K. R. Maddikunta, and T. R. Gadekallu, "Fake Review Classification Using Supervised Machine Learning," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 12664 LNCS, pp. 269–288, 2021, doi: 10.1007/978-3-030-68799-1_19.
- [3] S. Si, Y. Wu, L. Tang, Y. Zhang, and J. Wosik, "Evaluating the Performance of ChatGPT for Spam Email Detection," Comput. Lang., 2024, [Online]. Available: <http://arxiv.org/abs/2402.15537>.
- [4] M. Gaur, R. Aggrawal, A. Garg, and A. Singh, "Fake Profile Detection on Social sites," Proc. - 2024 6th Int. Conf. Comput. Intell. Commun. Technol. CCICT 2024, pp. 530–537, 2024, doi: 10.1109/CCICT62777.2024.00089.
- [5] Z. S. Torabi, M. H. Nadimi-Shahraki, and A. Nabiollahi, "Efficient Support Vector Machines for Spam Detection: A Survey," IJCSIS Int. J. Comput. Sci. Inf. Secur., vol. 13, no. January, pp. 10–28, 2015, [Online]. Available: <http://sites.google.com/site/ijcsis/>.
- [6] J. S. Whissell and C. L. A. Clarke, "Clustering for semi-supervised spam

- filtering,” *ACM Int. Conf. Proceeding Ser.*, pp. 125–134, 2011, doi: 10.1145/2030376.2030391.
- [7] R. Y. K. Lau, S. Y. Liao, R. Chi-Wai Kwok, K. Xu, Y. Xia, and Y. Li, “Text mining and probabilistic language modeling for online review spam detection,” *ACM Trans. Manag. Inf. Syst.*, vol. 2, no. 4, 2011, doi: 10.1145/2070710.2070716.
- [8] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, “Survey of review spam detection using machine learning techniques,” *J. Big Data*, vol. 2, no. 1, 2015, doi: 10.1186/s40537-015-0029-9.
- [9] I. AbdulNabi and Q. Yaseen, “Spam email detection using deep learning techniques,” *Procedia Comput. Sci.*, vol. 184, no. 2019, pp. 853–858, 2021, doi: 10.1016/j.procs.2021.03.107.
- [10] A. Sharma and S. Kumar, “Machine learning and ontology-based novel semantic document indexing for information retrieval,” *Comput. Ind. Eng.*, vol. 176, 2023, doi: 10.1016/j.cie.2022.108940.
- [11] “<https://blog.prospect.io/455-email-spam-trigger-words-avoid-2018>.”
- [12] S. M. Saqib, M. Z. Asghar, A. Al-rasheed, M. A. Khan, and Y. Ghadi, “DenseHillNet : a lightweight CNN for accurate classification of natural images,” *PeerJ Comput. Sci.*, pp. 1–21, 2024, doi: 10.7717/peerj-cs.1995.
- [13] S. M. Saqib et al., “Deep learning-based electricity theft prediction in non-smart grid environments,” *Heliyon*, vol. 10, no. 15, 2024, doi: 10.1016/j.heliyon.2024.e35167.
- [14] S. M. Saqib et al., “Cataract and glaucoma detection based on Transfer Learning using MobileNet,” *Heliyon*, vol. 10, no. 17, 2024, doi: 10.1016/j.heliyon.2024.e36759.
- [15] S. M. Saqib et al., “Lumpy skin disease diagnosis in cattle: A deep learning approach optimized with RMSProp and MobileNetV2,” *PLoS One*, vol. 19, no. 8 August, 2024, doi: 10.1371/journal.pone.0302862.
- [16] S. Ahmad, S. Muhammad, and A. Hassan, “CNN and LSTM based hybrid deep learning model for sentiment analysis on Arabic text reviews,” *Mehran Univ. Res. J. Eng. Technol.*, vol. 43, no. 2, pp. 183–194, 2024, doi: <https://doi.org/10.22581/muet1982.3130>.
- [17] F. Wu, Y. Huang, and Z. Yuan, “Domain-specific sentiment classification via fusing sentiment knowledge from multiple sources,” *Inf. Fusion*, vol. 35, pp. 26–37, 2017, doi: 10.1016/j.inffus.2016.09.001.
- [18] B. Liu, *Sentiment Analysis and Opinion Mining*, vol. 5, no. 1. 2012.
- [19] L. Shu, H. Xu, and B. Liu, “Lifelong Learning CRF for Supervised Aspect Extraction,” in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 2017, pp. 148–154, doi: 10.18653/v1/P17-2023.
- [20] A. A. A. Esmín, R. L. De Oliveira, and S. Matwin, “Hierarchical classification approach to emotion recognition in twitter,” in *Proceedings - 2012 11th International Conference on Machine Learning and Applications, ICMLA 2012*, 2012, vol. 2, no. March, pp. 381–385, doi: 10.1109/ICMLA.2012.195.
- [21] N. Sureja, N. Chaudhari, P. Patel, J. Bhatt, T. Desai, and V. Parikh, “Hyper-tuned Swarm Intelligence Machine Learning-based Sentiment Analysis of Social Media,” *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 4, pp. 15415–15421, 2024, doi: 10.48084/etasr.7818.
- [22] D. Elangovan and V. Subedha, “Adaptive Particle Grey Wolf Optimizer with Deep Learning-based Sentiment Analysis on Online Product Reviews,” *Eng. Technol. Appl. Sci. Res.*, vol. 13, no. 3, pp. 10989–10993, 2023, doi: 10.48084/etasr.5787.
- [23] E. Aljohani, “Enhancing Arabic Fake News Detection: Evaluating Data Balancing Techniques Across Multiple Machine Learning Models,” *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 4, pp. 15947–15956, 2024, doi: 10.48084/etasr.8019.
- [24] S. Gojali and M. L. Khodra, “Aspect based sentiment analysis for review rating prediction,” 2016, doi: 10.1109/ICAICTA.2016.7803110.
- [25] K. Dashtipour, M. Gogate, A. Adeel, H. Larijani, and A. Hussain, “Sentiment analysis of persian movie reviews using deep learning,” *Entropy*, vol. 23, no. 5, pp. 1–16, 2021, doi: 10.3390/e23050596.
- [26] R. Johnson and T. Zhang, “Effective use of word order for text categorization with convolutional neural networks,” in *NAACL HLT 2015 - 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, 2015, no. 2011, pp. 103–112, doi: 10.3115/v1/n15-1011.
- [27] Y. Zhang and B. Wallace, “A Sensitivity Analysis of (and Practitioners’ Guide to) Convolutional Neural Networks for Sentence Classification,” in *Proceedings of the 8th International Joint Conference on Natural Language Processing*, 2015, pp. 253–263, [Online]. Available: <http://arxiv.org/abs/1510.03820>.
- [28] S. Zhao, Z. Xu, L. Liu, M. Guo, and J. Yun, “Towards Accurate Deceptive Opinions Detection Based on Word Order-Preserving CNN,” *Math. Probl. Eng.*, vol. 2018, pp. 1–8, 2018, doi: 10.1155/2018/2410206.
- [29] D. Tang, B. Qin, and T. Liu, “Document modeling with gated recurrent neural network for sentiment classification,” in *Conference Proceedings - EMNLP 2015: Conference on Empirical Methods in Natural Language Processing*, 2015, no. September, pp. 1422–1432, doi: 10.18653/v1/d15-1167.
- [30] V. Christina, “Email Spam Filtering using Supervised Machine Learning Techniques,” *Int. J. ...*, vol. 02, no. 09, pp. 3126–3129, 2010, [Online]. Available: <http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jml=09753397&AN=58495579&h=iA3%2FX2tuV1JYXccJpWbuYj6fk0pDvAcACLAXisHLHSX%2FL4Hz9xiQMueUTWkrzsKKireW27Sl2NterjVC9NCQ%3D%3D&cr=c>.
- [31] J. R. Méndez, F. Fdez-Riverola, F. Díaz, E. L. Iglesias, and J. M. Corchado, “A comparative performance study of feature selection methods for the anti-spam filtering domain,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4065 LNAI, pp. 106–120, 2006, doi: 10.1007/11790853_9.
- [32] G. Sakkis, I. Androustopoulos, G. Paliouras, V. Karkaletsis, C. D. Spyropoulos, and P. Stamatopoulos, “Stacking classifiers for anti-spam filtering of e-mail,” 2001, [Online]. Available: <http://arxiv.org/abs/cs/0106040>.
- [33] L. Pelletier, J. Almhana, and V. Choulakian, “Adaptive filtering of SPAM,” *Proc. - Second Annu. Conf. Commun. Networks Serv. Res.*, pp. 218–224, 2004, doi: 10.1109/DNSR.2004.1344731.
- [34] G. M. Shahariar, S. Biswas, F. Omar, F. M. Shah, and S. Binte Hassan, “Spam Review Detection Using Deep Learning,” *2019 IEEE 10th Annu. Inf. Technol. Electron. Mob. Commun. Conf. IEMCON 2019*, no. October, pp. 27–33, 2019, doi: 10.1109/IEMCON.2019.8936148.
- [35] N. Bouguila and O. Amayri, “A discrete mixture-based kernel for SVMs: Application to spam and image categorization,” *Inf. Process. Manag.*, vol. 45, no. 6, pp. 631–642, 2009, doi: 10.1016/j.ipm.2009.05.005.
- [36] S. M. Saqib, K. Mahmood, and T. Naeem, “Comparison of LSI algorithms without and with pre-processing : using text document based search,” *Accent. Trans. Inf. Secur.*, vol. 1, no. 4, pp. 44–51, 2016.
- [37] A. Huang, D. Milne, E. Frank, and I. H. Witten, “Clustering Documents using a Wikipedia-based Concept Representation,” in *PAKDD 2009: Advances in Knowledge Discovery and Data Mining*, 2009, pp. 628–636.
- [38] Q. YANG, “SUPPORT VECTOR MACHINE FOR CUSTOMIZED EMAIL FILTERING BASED ON IMPROVING LATENT SEMANTIC INDEXING,” in *Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 2005*, pp. 18–21.
- [39] “<https://www.atoti.io/when-to-perform-a-feature-scaling/> (Visited On 3 July, 2021).”
- [40] F. Horasan, “Latent Semantic Indexing-Based Hybrid Collaborative Filtering for Recommender Systems,” *Arab. J. Sci. Eng.*, vol. 47, no. 8, pp. 10639–10653, 2022, doi: 10.1007/s13369-022-06704-w.
- [41] G. J. Phadnis N, “Framework for document retrieval using latent semantic indexing,” *Int. J. Comput. Appl.*, vol. 94, no. 14, 2014.
- [42] “<https://www.kaggle.com/ishansoni/sms-spam-collection-dataset>.”
- [43] S. Muthulakshmi, “Comparative Study on Classification Meta Algorithms,” *Ijircce.Com*, 2013, <https://www.rroij.com/open-access/comparative-study-on-classification-metaalgorithms.php?aid=43760>.

A Deep Learning for Arabic SMS Phishing Based on URLs Detection

Sadeem Alsufyani, Samah Alajmani

Department of Information Technology, College of Computers and Information Technology, Taif University, Taif, Saudi Arabia.

Abstract—The increasing use of SMS phishing messages in Arab communities has created a major security threat, as attackers exploit these SMS services to steal users' sensitive and financial data. This threat highlights the necessity of designing models to detect SMS messages and distinguish between phishing and non-phishing messages. Given the lack of sufficient previous studies addressing Arabic SMS phishing detection, this paper proposes a model that leverages deep learning models to detect Arabic SMS messages based on the URLs they contain. The focus is on the URL aspect because it is one of the common indicators in phishing attempts. The proposed model was applied to two datasets that were in English, and one dataset was in Arabic. Two datasets were translated from English to Arabic. Three datasets included a number of Arabic SMS messages, mostly containing URLs. Three deep learning models—CNN, BiGRU, and GRU—were implemented and compared. Each model was evaluated using metrics such as precision, recall, accuracy, and F1 score. The results showed that the GRU model achieved the highest accuracy of 95.3% compared to other models, indicating its ability to capture sequential patterns in URLs extracted from Arabic SMS messages effectively. This paper contributes to designing a phishing detection model designed for Arab communities to enhance information security within Arab communities.

Keywords—Phishing; URL phishing; SMS phishing; GRU; BiGRU; CNN

I. INTRODUCTION

Cybersecurity refers to one or more of the following three things as a set of security measures and other activities aimed at first protecting computer hardware and networks, related hardware and software, as well as the information they contain and transmit, including software and data. This protection includes protection against attacks, disruptions, or threats. Second, the status or quality of protection from threats. Third, expand the scope for public discussions aimed at the process of implementing and improving those activities and quality [1]. Therefore, a cyberattacks is an attempt by attackers to infiltrate information systems at the level of an individual or organization in a deliberate way. A cyberattack aims to disrupt the resources of the target victim's system by stealing his confidential information and disrupting the main functions of his system. After that, network assaults come in several types. The attackers search for the type of ransom after carrying out network attacks on organizations. This threat is not limited only to large companies but also includes medium and small organizations. The reason lies in the fact that medium and small organizations do not have high-level security measures, and this makes attackers also focus on medium and small companies and find out their vulnerabilities [2]. Cyberattacks have become widespread in our daily lives, affecting government institutions,

from the economic side to trade, as well as banks and hospitals. Malware, phishing, social engineering attacks, botnets, password attacks, man-in-the-middle attacks, and other types of cyberattacks are among the most prevalent types [3]. Therefore, preserving private data against social engineering threats such as phishing attempts is the primary objective of information security. To protect private data from these types of social engineering assaults, consumers, website developers, and experts have been particularly concerned about security vulnerabilities in every company [4]. Social engineering refers to the tactic of manipulating individuals to obtain unauthorized access to data. This method falls under the broader category of information security. People are the weakest point, the focal point of most organizations because they pose a threat to their organizations. If sensitive information about the organization is compromised, it falls into the wrong hands. Organizations usually use advanced security measures to minimize the chances of unauthorized individuals accessing that information. An organization needs to prevent its employees from succumbing to social engineering attacks. Most humans react emotionally, so they are more vulnerable than machines most often. The greatest threat that an organization poses to having sensitive information is human, not technical protection because they constitute the essence of an organization. As a result, attackers have concluded that using a human to get unauthorized access to an organization's data and communication technology infrastructure is more straightforward than attempting to breach security mechanisms [5]. Phishing attacks are prevalent in social engineering attacks. The attacker entices users to send fake messages such as winning a prize, sending a message from a fake social media account, or hacking passwords. These messages seem to be an order from a trusted entity, such as a bank disclosing information to achieve financial gain. Social engineering techniques with some fraudulent tactics are ingeniously used to entice users to acquire information. Fraudulent methods can connect to a message, phone, or fake email. Scammers send fake messages to many internet users. These attacks target people who lack sufficient knowledge about cyberattacks and their security. They are led to assume that the messages are from a legitimate organization. The core goal of phishing attacks is to search for the vulnerabilities of the intended user. The attacker always finds ways to make the targeted victims visit a phishing site. By designing fake messages in a way that makes them appear reliable, including a link that transports them to this fraudulent site, it is easy to target and deceive victims [4]. Phishing includes several types of attacks targeting users, such as voice phishing, email phishing, SMS phishing, website phishing, and social media phishing [6]. SMS phishing, also referred to as smishing, is a kind of phishing

that utilizes short message service (SMS) technology. This type of phishing exploits SMS messages on smartphones. The smishing happens in two major methods. The first is when an SMS message is sent from a reliable source, such as a bank or system administrator. The second way occurs when the victim receives an SMS message containing private content, such as an account block or stolen identity. Subsequently, the victim will be sent to a deceptive website or contacted via phone number to confirm their data [7].

Therefore, with the increasing prevalence of phishing attacks received through SMS, research targeting Arabic-speaking communities to detect Arabic SMS messages containing URLs remains insufficient. This gap poses a significant security threat to Arabic-speaking users and can lead to the loss of sensitive personal and financial data. While existing phishing detection solutions have targeted SMS messages in English, Arabic SMS messages based on URLs are relatively unexplored. The significance of this paper lies in its contribution to addressing the gap in phishing detection models specifically designed for Arabic SMS messages containing URLs. By leveraging deep learning models, this paper aims to mitigate the risk of phishing threats faced by Arabic-speaking users. Thus, the key challenges addressed in this work are: (1) how can URLs extracted from Arabic SMS messages be analyzed to determine whether the message is phishing or non-phishing? (2) what is the appropriate deep learning model for effective classification and detection of Arabic SMS phishing based on URLs? Accordingly, the paper proposes a proposed model to bridge this gap through several contributions. First, provide a model for detecting Arabic SMS messages based on URLs and determine the type of message, whether phishing or non-phishing, depending on the analysis of the URL contained in the Arabic SMS message. Second, we created a dataset of 16,521 Arabic SMS messages, which helps provide a dataset for future research in the field of Arabic SMS detection, then extracted and analyzed URLs from Arabic SMS using deep learning models: CNN, BiGRU, and GRU. Finally, evaluate and compare the performance of models for deep learning in the classification of Arabic SMS messages based on URLs, whether phishing or non-phishing. The results of this paper can benefit financial institutions and telecommunications service providers by providing an effective tool to protect sensitive data and reduce financial losses caused by SMS-based phishing threats.

The rest of this paper is organized as follows: Section II provides a review of related works. Section III presents the methodology. Section IV presents the results and discussions. Finally, Section V presents the conclusion and future work of the paper.

II. RELATED WORKS

In this section, we review several related works on SMS phishing detection with deep learning and machine learning methods.

Mishra & Soni, 2023 [8] presented a two-stage SMS phishing detection model. The first stage was URL validation domain checking. The second stage was SMS classification. The URL domain was checked, and SMS classification categorized the messages' text content and extracted some useful features. Finally, the system used a backpropagation algorithm to classify

the messages and was evaluated using the SMS dataset. The results showed an accuracy rate of 97.93%.

Mishra & Soni, 2020 [9] suggested the Smishing Detector model, which detected SMS phishing messages with minimal false positives. The model analyzed content through a Naïve Bayes classification. Four modules provided this model: APK Download Detector, SMS Content Analyzer, URL Filter, and Source Code Analyzer. The results demonstrated an overall accuracy rate of 96.29%.

Agrawal et al., 2023 [10] developed a model to detect fraudulent SMS. The model's design involved two phases: the first phase used a hybrid model for SMS message classification, whereas the examination of URLs was the second phase. Random Forest, Naïve Bayes, and Extra Tree classifiers were used in their hybrid model. The results showed that the Random Forest, Multinomial Bayes classifier, and Extra Tree classifier achieved 96.25% accuracy and 99.38% precision.

Prasanna Bharathi et al., 2021 [11] applied two well-known algorithms to categorize spam SMS: a Support Vector Machine and Naïve Bayes. 96.19% accuracy percent was achieved by the Naïve Bayes algorithm. 98.77% accuracy was achieved with the Support Vector Machine algorithm approach.

Wu et al., 2018 [12] introduced a novel approach to detecting SMS phishing utilizing oversampling technology to enhance feature selection and improve accuracy. They utilized three types of features, namely symbol features, subject features, language query features, and word calculation (LIWC). They applied one of the oversampling methods called the Adasyn adaptive synthetic sampling approach. The BPSO binary particle swarm was used to analyze the three feature types and then select the optimal combination of all the features. The experiment was performed on the Almeida et al. dataset, which contained 5574 messages in English. They used the Random Forest classification algorithm to obtain detection findings. The findings showed that the two methods offered by ADASYN and BPSO achieved the highest accuracy rate of 99.01%.

Oswald et al., 2022 [13] proposed an intent-based approach that efficiently handled the filtering of SMS spam, textual and semantic features of SMS messages were created using 13 pre-defined intent labels. Multiple pre-trained NLP models were applied to generate textual contextual embeddings. For the pre-defined labels, intent scores were computed. Several supervised learning classifiers were used to filter spam or ham. The results showed that the DistilBERT+SVM (Poly) model performed well with an accuracy (98.07%), precision, and recall (~0.97).

Tuan et al., 2023 [14] evaluated five algorithms on three various Vietnamese datasets: Support Vector Machine, Random Forests, Naïve Bayes, Convolutional Neural, and Long Short-Term Memory to evaluate the efficiency of spam detection in Vietnamese SMS. The results showed that the CNN and LSTM, supported by the transformer PhoBERT model, were more effective than the conventional models for machine learning. The LSTM model obtained the greatest accuracy of 97.77%, on the Vietnamese full-dialect dataset, while the CNN and PhoBERT models showed a high accuracy of 95.56% on the non-diacritic Vietnamese dataset.

The University of Baghdad et al., 2021 [15] suggested a new approach for detecting SMS spam that focused on improving the binary particle swarm based on fuzzy rule selection. Initially, the significant features of the SMS spam dataset were extracted. Then, a fuzzy collection of rules was produced using the features that were extracted. The most reliable fuzzy rules, which lowered complexity and enhanced model performance, were finally chosen using a binary particle swarm. The findings demonstrated that the suggested model achieved an F-measure of 94.6%, recall of 98.8%, accuracy of 98.5%, and precision of 90.8%.

Amir Sjarif et al., 2020 [16] proposed a method for classifying spam SMS messages through a variety of techniques for data mining. Algorithms such as Multinomial Naïve Bayes, Support Vector Machine, Naïve Bayes, and K Nearest Neighbor with different values of K = 1, 3, and 5 were trained and assessed using the dataset from the UCI machine learning repository. Each algorithm's performance was compared to determine which best-fitting classifier performed better in terms of accuracy, error, processing time, kappa statistics, and the lowest number of false positives. The SVM algorithm outperformed the other classifiers in terms of accuracy, with an average accuracy of 98.9% for detecting and labeling spam text messages. In terms of the error coefficient, the KNN algorithm had the highest error with K = 3 and K = 5, while SVM had the lowest error, followed by the Multinomial Naïve Bayes algorithm.

Uddin et al., 2024 [17] addressed spam detection using a transformer-based Large Language Models (LLMs) approach that was refined and optimized. The benchmark SMS spam dataset was used to detect spam messages. The imbalance problem in the data was mitigated by implementing methods for data augmentation, such as back translation. In addition, calculated the scores of positive and negative coefficients that detected and explained the transparency of the fine-tuned model in detecting spam messages using explainable artificial intelligence (XAI) techniques. Traditional models for machine learning and transformer-based models' performances were compared. The experiments showed that the refined and optimized BERT model with the variant model RoBERTa obtained the highest accuracy of 99.84%.

Ali et al., 2023 [18] proposed a new model for detecting SMS spam using (MLP) Multiple Linear Regression to extract seven features from each message. The message detection process was entrusted to the feature weight and Extreme Learning Machine (ELM). MLR was used to weigh the seven extracted features. The SMS was classified as spam or ham by ELM. The suggested model was evaluated for recall, F-measure, precision, and accuracy, and showed scores of 98.7%, 95.9%, 93.3%, and 98.2%, respectively.

Sonowal, 2020 [19] identified the greatest collection of features for the detection of SMS phishing by employing four ranking algorithms: Spearman's rank correlation, Pearson rank correlation, Kendall rank correlation, and Point biserial rank correlation, along with machine learning algorithms. According to the findings, the AdaBoost classifier provided the highest accuracy. When compared to other correlation algorithms, the Kendall rank correlation algorithm provided the best accuracy. Therefore, this finding proved that the ranking algorithm could

provide 98.40% accuracy and 61.53% reduction in feature dimensions.

Giri et al., 2023 [20] suggested various deep neural networks for spam SMS classification. The Tiago dataset was used, and some steps were taken to start with preprocessing and then feeding these preprocessed messages into two different models of deep learning (Long Short-Term Memory Network with Convolution Neural Network) with simple architectures. Word embedding techniques (BUNOW and GloVe) were combined to enhance the two basic architectures' accuracy with the deep learning models. The results after using the two-word embedding techniques in text categorization, demonstrated an accuracy of 98.44% with the CNN LSTM BUNOW model.

Table I summarizes the previous studies that contributed to SMS phishing message detection solutions.

TABLE I. COMPARISON OF PREVIOUS STUDIES ON SMS PHISHING USING MACHINE LEARNING AND DEEP LEARNING DETECTION

Ref.	Model architecture used	Dataset language	Result %
[8]	Backpropagation algorithm	English	97.93%
[9]	Naive Bayes	English	96.29%.
[10]	Random Forest, Naive Bayes, and Extra tree classifiers	English	96.25%
[11]	Support vector machine, and Naive Bayes.	English	98.77%
[12]	ADASYN, BPSO	English	99.01%
[13]	DistilBERT+SVM (Poly)	English	98.07%
[14]	Support Vector Machine, Random Forests, Naive Bayes, LSTM, and CNN	Vietnamese	97.77%, on the Vietnamese full-dialect and 95.56% on the non-diacritic Vietnamese
[15]	Binary particle swarm based on fuzzy rule selection.	English	98.5%
[16]	Support Vector Machine, Multinomial Naïve Bayes, Naïve Bayes, and K Nearest Neighbor with different values	English	98.9%
[17]	explainable artificial intelligence (XAI), transformer-based Large Language Models, refined and optimized BERT model with the variant model RoBERTa	English	99.84%
[18]	Multiple linear regression, extreme learning machine ELM.	English	98.2%
[19]	Spearman's rank correlation, Pearson rank correlation, Kendall rank correlation, and Point biserial rank correlation	English	98.40%
[20]	CNN LSTM BUNOW	English	98.44%

III. METHODOLOGY

The proposed methodology is based on the process of detecting Arabic SMS phishing messages based on URLs, which uses models for deep learning such as GRU, CNN, and BiGRU to examine and categorize these Arabic SMS messages. The process begins with the step of identifying SMS messages that contain URLs, which are often indicative of phishing tries.

Once identified, the Arabic SMS messages are classified based on the presence of URLs for further analysis. This analysis step is fundamental for understanding the type of content of Arabic SMS messages, with a focus on the URLs embedded within the Arabic SMS messages. The methodology evaluates patterns and characteristics that are usually related to phishing, such as suspicious domains or malicious URLs. The cleaned dataset experiences an inspecting process, where each Arabic SMS message is evaluated for the presence of a URL. After that, the URLs are extracted from the Arabic SMS messages and passed to the URL-based classification. The classification step utilizes models for deep learning, such as GRU, CNN, and BiGRU, to process and analyze the extracted URL dataset. These models were selected for their capability to capture sequential patterns, spatial features, and contextual dependencies, which are important in the process of detecting phishing tries.

Fig. 1 illustrates the overall methodology proposed in this paper, illustrating the steps in classifying Arabic SMS messages based on URLs based on the proposed models for deep learning.

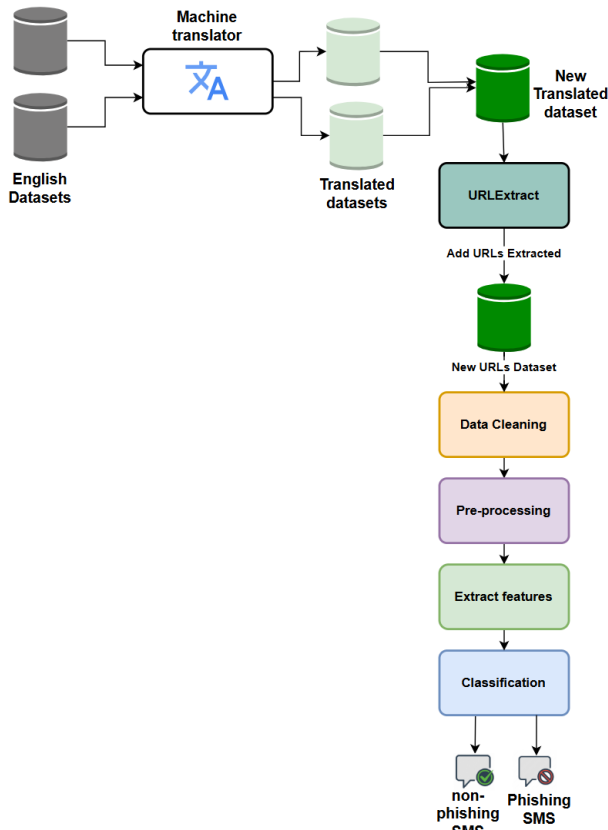


Fig. 1. The process of proposed Arabic SMS messages based on URLs detection.

The steps are explained as follows:

A. Step 1: SMS Messages Dataset and Translation

The datasets used in this process were obtained from three different sources: the first dataset from [21], the second dataset from the UCI Repository [22] and the third dataset from Kaggle [23]. The two sources contain [22] and [23] English datasets, which necessitate the use of machine translation, such as Google Translate, to convert SMS messages from English to Arabic.

This translation is necessary for the process of checking whether the SMS messages contain URLs or not, to classify Arabic SMS messages based on the type of the URL, whether it is phishing or non-phishing. SMS messages can be automatically translated using machine translation, a technology built on artificial intelligence systems.

Google Translate was utilized in our proposed model to translate two English datasets into Arabic. The file upload feature in Google Translate allowed us to translate the content of both datasets completely and comprehensively and convert them into Arabic to support the proposed model. After the translation process was successful, we downloaded the translated dataset file. This step is essential to support our proposed model to ensure the availability of an Arabic SMS dataset. We translated the two datasets [22] and [23] that were originally in English and then translated into Arabic using Google Translate.

B. Step 2: Merge the Dataset

This step creates a dataset containing a large number of SMS messages translated into Arabic, some of which include URLs in their content. The merging process was according to some steps:

1) *First step:* Column data type is identical. This rule indicates that the first column represents the label of each Arabic SMS message, whether it is non-phishing or phishing. The second column contains the text of the Arabic SMS message, which is of type text.

2) *Second step:* The number of columns is identical. This rule indicates that all the datasets have the same number of columns. Our dataset contains only two columns: The first column represents the label of the Arabic SMS message, categorizing it as either non-phishing or phishing. The second column represents the text of the Arabic SMS message.

After the merging process, we reached three datasets comprising a total of 16,521 Arabic SMS messages. Most of these messages include URLs, which will achieve our goal of detecting Arabic SMS phishing messages based on the URL they contain.

C. Step 3: URLs Dataset

We collected a dataset of URLs to support expanding the URL dataset to train and accurately classify deep learning models. The auxiliary dataset was collected from [24], which includes 20,000 URLs, categorized as either non-phishing or phishing. The dataset consists of two columns: the first column represents lists of the URLs, and the second column represents the type of URLs, whether non-phishing or phishing.

D. Step 4: URL Extraction and Merging Dataset

The dataset is processed by the URL classification component, starting with the extraction of URLs from the SMS messages using the URLExtract library, a Python library that extracts URLs from Arabic SMS messages. All URLs extracted from Arabic SMS messages are saved and merged with a new dataset [24] containing a large number of URLs. Merging these URLs with the new URL dataset enhances the expansion of the

URL dataset, which improves the accuracy of the training and classification using deep learning models.

E. Step 5: Data Cleaning

After completing the previous step of preparing the URL dataset, the next step is cleaning the dataset. Any unnecessary elements are removed. The cleaning process includes the following:

1) *Removing duplicate URLs*: This refers to eliminating duplicate entries where the same URL or the same rows are repeated.

2) *Removing null values*: This refers to removing cells that contain null or missing values, i.e., the URL is not provided or the label is missing.

F. Step 6: Pre-Processing Process

Preprocessing is an important step in improving data quality. This directly contributes to enhancing the accuracy and capability of the model. By thoroughly preparing the dataset, we ensure that it enhances feature extraction, helping the model provide more reliable and precise classification results.

One of the fundamental preprocessing tasks includes dealing with the URLs within the dataset. This includes various tasks that contribute to normalizing and standardizing URLs to remove inconsistencies and duplication. Key steps include:

1) *Converting characters to lowercase*: All characters in URLs are changed to lowercase. This helps to avoid handling the same URL with different letter cases as individual entities.

For example: ForExample.com

After the conversion process, it forexample.com

2) *Removal of numbers*: Any numeric values within URLs that are not appropriate to the classification task are removed to streamline the dataset and reduce noise.

3) *Removal of extra spaces*: Unnecessary spaces within URLs are removed to ensure consistency in data formatting and prevent errors during processing and classification.

4) *Removal of symbols*: Many of the elements that do not significantly contribute to the classification process are removed to avoid unnecessary complexity.

By performing these preprocessing steps, the dataset becomes more accurate, allowing the model to focus on the important aspects of the data. This approach lays the foundation for a more effective feature extraction process, leading to enhanced performance in data classification.

G. Step 7: Extraction Features

Features are extracted using lexical features. Lexical features are derived from the textual and structural components of URLs. The motivation for using lexical features is to rely on the appearance of a URL to determine the type of phishing or non-phishing. These features are commonly used in phishing detection systems and machine models to classify URLs as phishing or non-phishing [25].

1) *Features based on length*: These features depend on the length of many URL components:

a) *URL length*: Refers to the overall number of characters in the URL, including the protocols, hostname, path, queries, and any additional parameters.

b) *Path length*: Refers to the length of the URL path, which indicates a specific page or resource on the site.

c) *Hostname length*: Refers to the length of the part of the URL that identifies the server or site.

d) *Top-level domain length*: Refers to the top-level domain's length, indicating the type or geographic region of the site.

e) *First directory length*: Refers to the length of the first directory, which is the first part after '/' in the path.

2) *Features based on count*: Refers to the dependence of features on the number of times certain patterns appear within URLs. They are useful for analyzing URLs and discovering patterns that indicate phishing or non-phishing.

a) *Number of dashes*: Refers to the number of '-' symbols repeated within the URL. Its significance lies in identifying suspicious URLs that use many dashes in the process of dividing long parts of the URL.

b) *Number of @ in the URLs*: Refers to the total number of '@' symbols that appear. Its importance lies in the fact that some URLs contain the '@' symbol, which indicates attempts to redirect users within the URL.

c) *Number of question marks*: Refers to the count of '?' symbols in the URL. These are often used for creating queries, which attackers may use to collect user data.

d) *Number of percentage signs*: Refers to counting the number of '%' symbols in the URL, often used in encoding. Attackers may exploit this to hide parts of the URL or include special characters.

e) *Number of HTTP instances*: Count how many times HTTP appears in the URL. Some phishing URLs misuse HTTP to redirect the user.

f) *Number of HTTPS instances*: Count how many times HTTPS appears in the URL.

g) *Number of WWW instances*: Count the repetitions of WWW in the URL.

h) *Number of dots in the URLs*: Refers to the count of '.' dots. Phishing URLs may use excessive dots in domain names to appear similar to legitimate sites.

i) *Number of equal signs in the URLs*: Refers to the number of '=' symbols repeated in the URL, often used in query transactions. A high frequency may indicate data-collection attempts.

3) *Features based on binary*: Malicious URLs often use techniques to obscure their true identity, complicating detection by users and security systems. One popular tactic is replacing domain names with IP addresses (IPv4 or IPv6).

For example, instead of using a domain like: <http://forexample.com/phishing>

An attacker may use an IPv4 address and convert it to:
<http://196.168.1.1/phishing>

This is for the case where the domain name changed from the identifiable to IPv4.

As for if the attacker uses an IPv6 instead of using the domain name.

For example: <http://forexample.com/phishing>

and converted it to: <http://2001:db8:ff00:42:8329/phishing>

Attackers exploit users' limited familiarity with IP addresses compared to domain names, making them more likely to click on these URLs without suspicion. However, using IP addresses instead of domain names can bypass detection systems that rely on domain-based pattern analysis, enhancing phishing or malware distribution capabilities.

H. Step 8: Classification Process

After completing the previous steps, which include dataset splitting, model building, and classification, the process is as follows:

1) *Data splitting*: The dataset is split into two groups:

a) *Training Group*: Refers to the data used to train the models.

b) *Testing Group*: Refers to the data used to evaluate the performance of the model.

c) The dataset is split with 70% for training and 30% for testing.

2) *Model building*: In our proposed model, the datasets are passed to different deep learning models, namely GRU, CNN, and BiGRU.

a) *CNN model*: This model is known as Convolutional Neural Network. In terms of data, it was developed as a method to handle it in various types. The structure and operation of the brain's visual cortex served as the model's inspiration [26].

b) *BiGRU model*: This model stands for Bidirectional GRU and contains a two-layer reinforcement neural network. This design allows the two layers of the output layer to fully integrate the contextual data of the input data sequence at every moment. The basic concept behind this model is that the input sequence is processed by both the forward and backward neural networks [27].

c) *GRU model*: It is a type of RNN, GRU short for Gated Recurrent Units. It contains GRU units, which are used for deep learning, particularly effective in processing sequential data for applications [28].

3) *Classification*: URLs are classified using deep learning models such as GRU, CNN, and BiGRU.

The results then indicate that this URL-based SMS is either phishing or a non-phishing message.

I. Step 9: Model Evaluation

The model evaluation process plays a crucial role in the evaluation performance of three models for deep learning in our proposed model. The evaluation process is based on four major criteria: precision, accuracy, recall, and F1 score. These criteria

are essential to providing a comprehensive understanding of the model's ability to classify Arabic SMS messages containing URLs as non-phishing or phishing. By analyzing these criteria, we can determine the model that performs best in the particular task of detecting phishing in Arabic SMS. Following is an explanation of each of the four evaluation criteria:

First, the parameters used to evaluate the performance of models for deep learning are explained:

- **True Positive (TP)**: Represents the number of URLs that were correctly classified as positive, indicating that phishing URLs were correctly classified as phishing.
- **True Negative (TN)**: Represents the number of URLs that were correctly classified as negative, indicating that non-phishing URLs were correctly classified as non-phishing.
- **False Positive (FP)**: Represents the number of URLs that were incorrectly classified as positive category, i.e., non-phishing URLs that were incorrectly classified as phishing.
- **False Negative (FN)**: Represents the number of URLs incorrectly classified as belonging to the negative category, i.e., phishing URLs that were incorrectly classified as non-phishing.

Next, we will explain the four evaluation criteria:

- **Accuracy**: This metric is used to evaluate the quality of classification. It considers the rate of correct classification across all categories, rather than the distribution of the dataset. It reflects the number of correct predictions made by the model, whether the classifications are positive, i.e., identifying URLs as phishing, or negative, i.e., identifying URLs as non-phishing. A higher accuracy value indicates that the model is effectively classifying Arabic SMS messages based on URLs. It is represented by the following Eq. (1):

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{FP} + \text{FN} + \text{TP}} \quad (1)$$

- **Recall**: It represents the percentage of actual phishing URLs correctly identified by the model. It indicates the model's ability to detect all phishing instances. A higher recall rate means that the model is less likely to ignore phishing URLs. It is represented by the following Eq. (2):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2)$$

- **F1 Score**: It refers to the average between precision and recall, providing an integrated view of model performance, commonly used to assess the performance of the model in unbalanced classification problems. It is represented by the following Eq. (3):

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

- **Precision**: It refers to correct positive predictions, i.e., URLs that were correctly identified as phishing. An

increase in precision indicates the model is less likely to mistakenly classify non-phishing URLs as phishing. It is represented by the following Eq. (4):

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4)$$

IV. RESULTS AND DISCUSSIONS

A. Results

In this section, we present the performance of the proposed models for detecting Arabic SMS messages based on URLs. Three deep learning models were utilized: CNN, BiGRU, and GRU. Fig. 2 illustrates a comparison of the deep learning models' accuracy.

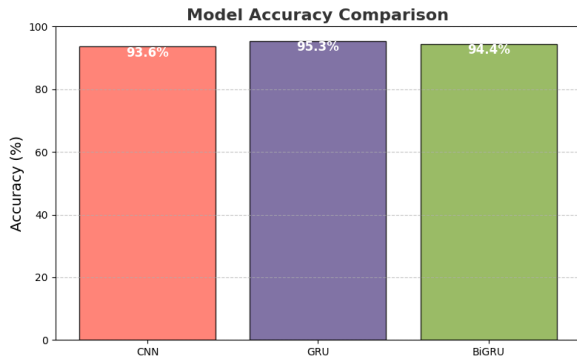


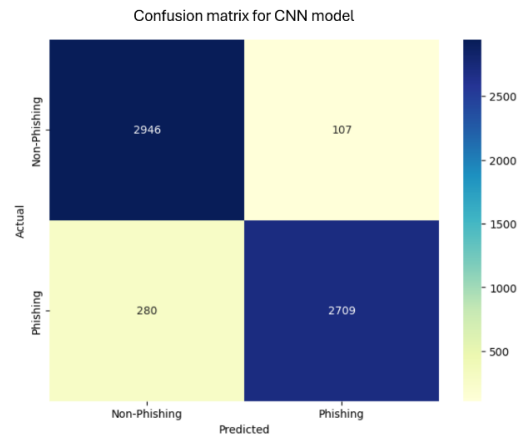
Fig. 2. Accuracy performance of three models for deep learning: CNN, GRU, and BiGRU.

Fig. 2 provides a detailed comparison of the performance of the three models proposed for deep learning in our proposed model for detecting Arabic SMS messages based on URLs. The classification focuses on determining whether an Arabic SMS message is phishing or non-phishing based on the type of URL extracted. The comparison was based on the accuracy achieved by each model. The GRU model demonstrates the best performance among the three models, achieving a superior accuracy rate of 95.33%. This high accuracy focuses on the GRU model's ability to effectively learn temporal dependencies in sequential data, making it particularly suitable for analyzing datasets. While the BiGRU model ranked second with an accuracy rate of 94.42%, slightly lower than the GRU model, the BiGRU's bidirectional architecture enables it to capture context in both the forward and backward directions. The CNN model achieved an accuracy rate of 93.59%, which, although lower than the GRU and BiGRU models.

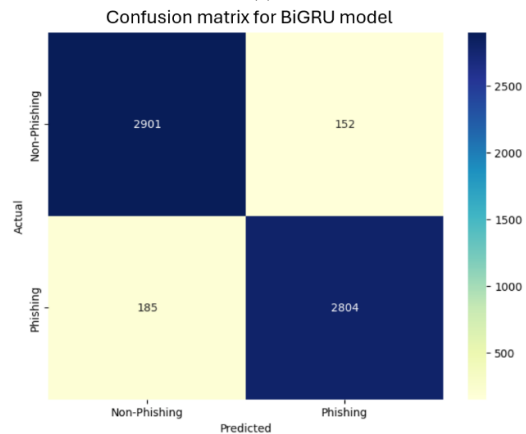
From the graph in Fig. 2, it is clear that the GRU model outperforms the others in terms of accuracy. This superior performance demonstrates that the GRU is the most effective for the task of classifying Arabic SMS messages based on URLs in this paper.

Fig. 3, presents a confusion matrix for the three deep learning models, CNN, BiGRU, and GRU, used to classify Arabic SMS messages based on URL type as phishing or non-phishing. Confusion matrix (a) shows the performance of the CNN model. It correctly classified 2709 phishing URLs as phishing and correctly classified 2946 non-phishing URLs as non-phishing. However, it incorrectly classified 107 non-phishing URLs as phishing and incorrectly classified 280 phishing URLs as non-phishing.

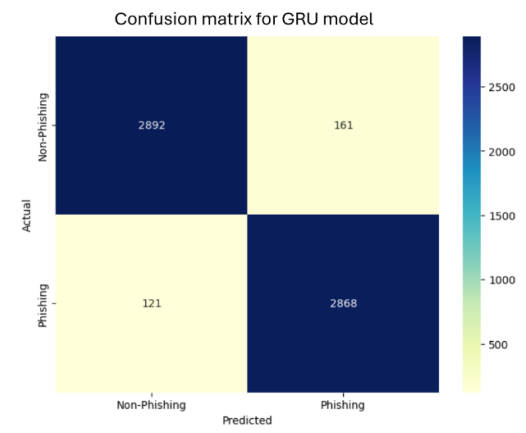
Confusion matrix (b) displays the performance of the BiGRU model, which correctly classified 2901 non-phishing URLs as non-phishing and correctly classified 2804 phishing URLs as phishing. Nevertheless, it incorrectly classified 152 non-phishing URLs as phishing and incorrectly classified 185 phishing URLs as non-phishing. Confusion matrix (c) illustrates the results of the GRU model, which correctly classified 2892 non-phishing URLs as non-phishing and correctly classified 2868 phishing URLs as phishing. However, incorrectly classifying 161 non-phishing URLs as phishing and incorrectly classifying 121 phishing URLs as non-phishing.



(a)



(b)



(c)

Fig. 3. Confusion matrix, (a) CNN, (b) BiGRU, (c) GRU.

TABLE II. A COMPREHENSIVE COMPARISON OF THE THREE PROPOSED MODELS, NAMELY CNN, BiGRU, AND GRU

Models	Evaluation Metrics %				Time (seconds)	
	Accuracy	Precision	Recall	F1 score	Train	Test
CNN	93.59	96.20	90.63	93.33	8.19	0.70
GRU	95.33	94.68	95.95	95.31	32.16	1.33
BiGRU	94.42	94.86	93.81	94.33	190.69	3.27

Based on Table II, a comprehensive comparison of the three proposed models, namely CNN, BiGRU, and GRU, is presented. This comparison is based on several performance metrics, namely precision, F1, recall, and accuracy, as well as two additional elements: the time required for training and testing. This analysis helps to understand the strengths and weaknesses of each model. The GRU model achieved the highest accuracy rate compared to the other two models at 95.33%, which indicates its strength in classification. While the precision was 94.68%, slightly lower than that of the CNN model, the recall was 95.95%, the highest among the models, indicating the model's ability to detect Arabic SMS phishing messages more effectively based on the URLs. The F1 score also had the highest result at 95.31%, reflecting a strong and balanced performance. As for the training and testing time performance, the training time was 32.16 seconds, and the testing time was 1.33 seconds, which is slower than CNN but faster than BiGRU. The BiGRU model followed, achieving a lower accuracy than GRU, at 94.42%, which is the average between the two models, CNN and GRU. It achieved a recall percentage that was lower than GRU but higher than CNN, which was 93.81%. As for the precision, it was also average among the other models, at 94.86%, slightly lower than GRU. The F1 score was 94.33%. In terms of training and testing time, it achieved a training time of 190.69 seconds and a testing time of 3.27 seconds, making it the slowest model in both training and testing, due to the processing of texts in both directions, i.e., reading from beginning to end and from end to beginning. The last model was the CNN model, which achieved an accuracy rate of 93.59%, which is a reasonable performance, but it is lower compared to the other models. In terms of recall, it was the lowest, indicating that it may miss some phishing messages, which was 90.63%. However, in terms of precision, it was the best among the other models, indicating that the model avoids false positives to a large extent, with a precision of 96.20%. The F1 score was 93.33%, reflecting a balanced result between precision and recall. In terms of training and testing time, the training time was 8.19 seconds, and the testing time was 0.70 seconds. This indicates that it is the fastest model in both training and testing among all the models, making it an excellent choice for practical applications in time-critical situations.

B. Discussions

The results indicate that GRU is the most effective model for classifying Arabic SMS messages based on URLs, due to its high accuracy, recall, and balanced F1 score. This makes it able to learn temporal dependencies perfectly in analyzing sequential data, such as Arabic SMS messages containing URLs. Although

the BiGRU model was able to capture context from both directions, forward and backward, it took longer training and testing time, which can limit its practical application in real-time scenarios. While the CNN model was the fastest, it performed poorly in accuracy and recall, making it an excellent choice when speed is considered. Therefore, the results emphasized that the GRU model outperforms both BiGRU and CNN in terms of accuracy and recall, which indicates its strength in processing sequential data and provides the best balance between speed and classification, making it the superior choice for detecting Arabic SMS phishing messages based on URLs. However, the CNN model provided the fastest training and testing time, but it provided the lowest recall, indicating a higher probability of missing Arabic SMS phishing messages based on URLs, which reduces its reliability compared to GRU and BiGRU. The BiGRU model is an alternative solution when the demand for contextual understanding is high. Therefore, based on the paper's goal of choosing a deep learning model that provides better accuracy in detecting Arabic SMS phishing messages based on URLs, the GRU model is the most suitable choice that achieves this goal based on the previous results.

V. CONCLUSION AND FUTURE WORK

The rapid development and widespread use of smartphones have led to an increase in cyber-attacks targeting smartphones, including SMS phishing attacks. This paper proposed a model for detecting Arabic SMS phishing messages based on URLs using models for deep learning, namely GRU, CNN, and BiGRU. We assessed the performance of these deep learning models and compared their accuracy and effectiveness in the detection process. The GRU model illustrates superior performance with an accuracy of 95.3%, demonstrating its capability to effectively process data sequences and capture contextual relations within the dataset. This high level of accuracy makes the GRU model an excellent candidate for applications where accuracy is critical. Although the CNN model achieved a slightly lower accuracy of 93.6%, it was capable of better in faster training time compared to the GRU model. This makes CNN a strong option for real-time scenarios requiring faster processing as a priority. The BiGRU model achieved an accuracy of 94.4%, which is lower than GRU but higher than CNN, although it did not outperform GRU in terms of performance. Its bidirectional structure allowed it to capture contextual data in both forward and backward directions, making it the suitable option in certain applications. These results emphasized the significance of selecting the appropriate model based on certain requirements, such as accuracy or speed.

This paper has achieved valuable objectives, but it has some limitations. First, the dataset used was relatively small and translated from English to Arabic due to the lack of a supporting Arabic dataset in this field, which may affect the results. Second, the models were assessed based on URLs as a phishing indicator, excluding other indicators that may be used as phishing processes, such as email and phone numbers. In future work, we aim to expand the Arabic dataset, compare the proposed models with other deep learning techniques to mitigate phishing detection in Arabic SMS messages and extend the proposed model to include other indicators such as email and phone numbers. These proposals aim to create a more

comprehensive solution to mitigate SMS phishing messages in Arabic-speaking communities.

REFERENCES

- [1] E. A. Fischer, "Cybersecurity Issues and Challenges: In Brief," 2014.
- [2] K. M. Sudar, P. Deepalakshmi, P. Nagaraj, and V. Muneeswaran, "Analysis of Cyberattacks and its Detection Mechanisms," in 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Bangalore, India: IEEE, Nov. 2020, pp. 12–16. doi: 10.1109/ICRCICN50933.2020.9296178.
- [3] Ö. Aslan, S. S. Aktuğ, M. Ozkan-Okay, A. A. Yilmaz, and E. Akin, "A Comprehensive Review of Cyber Security Vulnerabilities, Threats, Attacks, and Solutions," *Electronics*, vol. 12, no. 6, p. 1333, Mar. 2023, doi: 10.3390/electronics12061333.
- [4] S. Gupta, A. Singhal, and A. Kapoor, "A literature survey on social engineering attacks: Phishing attack," in 2016 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India: IEEE, Apr. 2016, pp. 537–540. doi: 10.1109/CCAA.2016.7813778.
- [5] F. Mouton, L. Leenen, M. M. Malan, and H. S. Venter, "Towards an Ontological Model Defining the Social Engineering Domain," in ICT and Society, K. Kimppa, D. Whitehouse, T. Kuusela, and J. Phahlamohlaka, Eds., Berlin, Heidelberg: Springer, 2014, pp. 266–279. doi: 10.1007/978-3-662-44208-1_22.
- [6] R. Alabdan, "Phishing Attacks Survey: Types, Vectors, and Technical Approaches," *Future Internet*, vol. 12, no. 10, Art. no. 10, Oct. 2020, doi: 10.3390/fi12100168.
- [7] College of Computers and Information Technology, Taif University, Saudi Arabia et al., "Four Most Famous Cyber Attacks for Financial Gains," *IJEAT*, vol. 9, no. 2, pp. 2131–2139, Dec. 2019, doi: 10.35940/ijeat.B3601.129219.
- [8] S. Mishra and D. Soni, "DSmishSMS-A System to Detect Smishing SMS," *Neural Comput & Applic*, vol. 35, no. 7, pp. 4975–4992, Mar. 2023, doi: 10.1007/s00521-021-06305-y.
- [9] S. Mishra and D. Soni, "Smishing Detector: A security model to detect smishing through SMS content analysis and URL behavior analysis," *Future Generation Computer Systems*, vol. 108, pp. 803–815, Jul. 2020, doi: 10.1016/j.future.2020.03.021.
- [10] N. Agrawal, A. Bajpai, K. Dubey, and B. D. Patro, "An Effective Approach to Classify Fraud SMS Using Hybrid Machine Learning Models," 2023, p. 6. doi: 10.1109/I2CT57861.2023.10126300.
- [11] P. Prasanna Bharathi, G. Pavani, K. Krishna Varshitha, and V. Radhesyam, "Spam SMS Filtering Using Support Vector Machines," in *Intelligent Data Communication Technologies and Internet of Things*, vol. 57, J. Hemanth, R. Bestak, and J. I.-Z. Chen, Eds., in *Lecture Notes on Data Engineering and Communications Technologies*, vol. 57, Singapore: Springer Singapore, 2021, pp. 653–661. doi: 10.1007/978-981-15-9509-7_53.
- [12] T. Wu, K. Zheng, C. Wu, and X. Wang, "SMS Phishing Detection Using Oversampling and Feature Optimization Method," *dtcse*, no. iece, Dec. 2018, doi: 10.12783/dtcse/iece2018/26634.
- [13] C. Oswald, S. E. Simon, and A. Bhattacharya, "SpotSpam: Intention Analysis-driven SMS Spam Detection Using BERT Embeddings," *ACM Trans. Web*, vol. 16, no. 3, pp. 1–27, Aug. 2022, doi: 10.1145/3538491.
- [14] V. M. Tuan, N. X. Thang, and T. Q. Anh, "Evaluating the Efficiency of Vietnamese SMS Spam Detection Techniques," *ISJ*, vol. 1, no. 18, Jun. 2023, doi: 10.54654/isj.v1i18.932.
- [15] University of Baghdad, S. Hameed, Z. Ali, and Mustansiriyah University, "SMS Spam Detection Based on Fuzzy Rules and Binary Particle Swarm Optimization," *IJIES*, vol. 14, no. 2, pp. 314–322, Apr. 2021, doi: 10.22266/ijies2021.0430.28.
- [16] N. N. Amir Sjarif, Y. Yahya, S. Chuprat, and N. H. F. Mohd Azmi, "Support Vector Machine Algorithm for SMS Spam Classification in The Telecommunication Industry," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 10, no. 2, p. 635, Apr. 2020, doi: 10.18517/ijaseit.10.2.10175.
- [17] M. A. Uddin, M. N. Islam, L. Maglaras, H. Janicke, and I. H. Sarker, "ExplainableDetector: Exploring Transformer-based Language Modeling Approach for SMS Spam Detection with Explainability Analysis," May 12, 2024, arXiv:2405.08026. Accessed: Oct. 04, 2024. [Online]. Available: <http://arxiv.org/abs/2405.08026>
- [18] Z. H. Ali, H. M. Salman, and A. H. Harif, "SMS Spam Detection Using Multiple Linear Regression and Extreme Learning Machines," *Iraqi Journal of Science*, pp. 6342–6351, Oct. 2023, doi: 10.24996/ij.s.2023.64.10.45.
- [19] G. Sonowal, "Detecting Phishing SMS Based on Multiple Correlation Algorithms," *SN COMPUT. SCI*, vol. 1, no. 6, p. 361, Nov. 2020, doi: 10.1007/s42979-020-00377-8.
- [20] S. Giri, S. Das, S. B. Das, and S. Banerjee, "SMS Spam Classification–Simple Deep Learning Models with Higher Accuracy using BUNOW and GloVe Word Embedding", doi: [http://dx.doi.org/10.6180/jase.202310_26\(10\).0015](http://dx.doi.org/10.6180/jase.202310_26(10).0015).
- [21] A. Ibrahim, S. Alyousef, H. Alajmi, R. Aldossari, and F. Masmoudi, "Phishing Detection in Arabic SMS Messages using Natural Language Processing," in 2024 Seventh International Women in Data Science Conference at Prince Sultan University (WiDS PSU), Riyadh, Saudi Arabia: IEEE, Mar. 2024, pp. 141–146. doi: 10.1109/WiDS-PSU61003.2024.00040.
- [22] J. H. Tiago Almeida, "SMS Spam Collection." UCI Machine Learning Repository, 2011. doi: 10.24432/C5CC84.
- [23] "Spam / Ham SMS DataSet." Accessed: Oct. 04, 2024. [Online]. Available: <https://www.kaggle.com/datasets/vivekchutke/spam-ham-sms-dataset>
- [24] E. S. Aung and H. Yamana, "Segmentation-based Phishing URL Detection," in *IEEE/WIC/ACM International Conference on Web Intelligence, ESSENDON VIC Australia: ACM*, Dec. 2021, pp. 550–556. doi: 10.1145/3486622.3493983.
- [25] D. Sahoo, C. Liu, and S. C. H. Hoi, "Malicious URL Detection using Machine Learning: A Survey," Aug. 21, 2019, arXiv: arXiv:1701.07179. doi: 10.48550/arXiv.1701.07179.
- [26] S. Min, B. Lee, and S. Yoon, "Deep Learning in Bioinformatics," vol. 47(6), pp. 366–382, Dec. 2023, doi: <https://doi.org/10.55730/1300-0152.2671>.
- [27] P. Li et al., "Bidirectional Gated Recurrent Unit Neural Network for Chinese Address Element Segmentation," *IJGI*, vol. 9, no. 11, p. 635, Oct. 2020, doi: 10.3390/ijgi9110635.
- [28] "Gated Recurrent Unit Definition | DeepAI." Accessed: Oct. 03, 2024. [Online]. Available: <https://deepai.org/machine-learning-glossary-and-terms/gated-recurrent-unit>.

Jordanian Currency Recognition Using Deep Learning

Salah Alghyaline

Department of Computer Science, the World Islamic Sciences and Education University, Amman, Jordan

Abstract—Automatic Currency Recognition (ACR) has a significant role in various domains, such as assessment of visually impaired people, banking transactions, counterfeit detection, digital transformation, currency exchange, vendor machines, etc. Therefore, developing an accurate ACR system enhances efficiency across several domains. The contribution of this paper is three-fold; it proposed a large dataset of 2799 images and seven denominations for Jordanian currency recognition. The second contribution proposed an efficient multiscale VGG net to recognize Jordanian currency. Third, popular CNN architectures on the proposed dataset will be evaluated, and the result will be compared with the proposed architectures. Four metrics were used in the evaluation. The experimental result showed the accuracy of the proposed Multiscale VGG outperformed VGG16, DenseNet121, ResNet50, and ResNet101 and achieved 99.88%, 99.88%, 99.89%, and 99.98% accuracy, precision, sensitivity, and specificity.

Keywords—Automatic currency recognition; deep learning; VGG

I. INTRODUCTION

Currency recognition uses image processing techniques to identify currency. People use currency in their daily lives, and it is important to develop an automatic way to recognize it [1]. Currency recognition systems have many applications, such as ATM machines, vendor machines, money exchange shops, bank systems, blind people's assistants, and the detection of fake currency [2]. According to a study in 2020, 43.3 million people were blind, and 258 million had low vision ability, most of them from developing and poor countries [3].

Deep learning approaches showed super results in the image processing field [4]. Many approaches in the literature were proposed to develop currency recognition systems for different currencies worldwide. There are more than 180 currencies worldwide, and each country has its specifications for currency in terms of size, paper, pattern, and color [5]. Here is a shortage of publicly available datasets for Jordanian currency¹.

This study reviews the literature on Jordanian currency recognition and proposes a new dataset that includes 2799 images representing seven denominations of Jordanian currency. The proposed dataset includes paper banknotes and metals. Moreover, the proposed dataset is evaluated on different Deep learning architectures, and some of these architectures were modified to improve recognition accuracy.

The paper is organized as follows: Section II reviews the related work in automatic currency recognition. Section III presents the materials and methodology. This includes the

proposed dataset, the CNN architectures used, and the proposed Multiscale VGG. Implementation Details are described in Section IV. It consists of an experimental environment, model parameter settings, performance metrics, and experimental results. Finally, Section V concludes the paper.

II. RELATED WORKS

A. Recognition of Currency

Due to rapid technological advancement, many applications require automatic currency recognition, such as detecting fake currency in vending machines and ATMs and assisting visually impaired persons [6].

Many approaches were proposed in the literature to address automatic currency recognition; some of them used traditional methods such as histogram analysis, edge detection, descriptors-based features like Histogram of Oriented Gradients (HOG) [7], Speeded-Up Robust Features (SURF) [8] and Scale-Invariant Feature Transform (SIFT) [9]. After extracting the image features, the simple way to predict the class label is to use Template matching between the extracted features and the predefined feature. Support Vector Machines (SVM), k-nearest Neighbors (k-NN), and Random Forests are also used by many approaches to classify the extracted features. These traditional approaches are usually computationally efficient but could struggle with challenged images with noise such as lighting, clutter, occlusion, orientation and background; moreover, they extracted a limited number of features [10] [11].

Deep learning has shown superior results in image recognition during the last few years. Many CNN architectures have been proposed and used broadly in image processing and computer vision applications, such as VGG [12], DenseNet121 [13], Resnet [14], and Inception V3 [15]. It is reported that deep learning achieved high accuracy in automatic currency recognition [16] [17] [18][19].

B. Literature Review

Automatic currency recognition is significant in people's daily lives, helping blind people or those with vision problems, automatic selling machines, detecting fake bank notes, and banking applications. Therefore, many approaches were proposed to address this problem.

The study in [2] proposed a system to recognize Indian real currency from fake; the system starts with noise removal by converting the image into a grayscale image and resizing it into a fixed dimension. Then, the image histogram is extracted, and template matching is used to determine whether the currency is real or not.

¹<https://drive.google.com/drive/folders/1faAhWB7B8CohvTBTAxhNp5C2m4HuPa8A>

The study in [20], developed a mobile application to recognize Yemeni paper currency. 1600 images and four currency denominations were used. The images trained on the MobileMe architecture, a group of images ranging from 18 to 29, were used to evaluate the model accuracy; the model achieved 100% accuracy.

The study in [21], proposed an approach to recognize three Nigerian paper currencies. The approach is based on a color histogram for feature extraction and a rule-based technique for image classification. The dataset includes 300 images, and the approach achieved 98.66% accuracy.

The study in [16] is employed to recognize Indian banknotes. Four CNN architectures were implemented using a basic sequential model, VGG 16, AlexNet, and MobileNet architectures were trained and tested on a dataset of 1270 images. The approach achieved 97.98%, 92.81%, 89%, and 71.66% on the four architectures, respectively.

The research [22] developed an approach to recognizing US banknotes. The approach is based on the Speeded-UP Robust Features (SURF) descriptor and uses template matching to recognize currency.

The research [23] developed a banknote recognition system for three countries: the US, Egypt, and Saudi Arabia. SURF was used for key-point detection, a histogram of oriented gradients (HOG), and the scale-invariant feature transform (SIFT) were used to describe the features. A support Vector Machine (SVM) was used to classify the features into 12 classes of currency denominations. The system achieved a 99.2% accuracy rate.

The study in [24] developed a system using Python and Raspberry Pi to distinguish original and counterfeit 100 NTD; Mean Gray Values (MGVs) were analyzed in specific regions; these regions represent note security regions.

The research in [25] proposed an open dataset for the Indian currency. The dataset includes 5125 images resized into 1280 × 768 pixels, captured by a Galaxy A33 5G and Apple iPhone 6. The dataset includes four denominations (10, 20, 50 and 100

Rupees), including the new and old shapes of these denominations.

Faster R-CNN and YOLOv5 [17] were trained to recognize rupiah banknotes. A dataset of 1120 images that represent eight classes is used to make the comparison. R-CNN and YOLOv5 achieved accuracies of 98.65% and 82.1%, respectively.

The study in [18] proposed CNN architecture to predict four currencies: the US dollar, Euro, Jordan dinner, and Korean won. The architecture is an improved version of YOLO-v3 architecture. It consists of 69 conventional layers. The model was evaluated on a dataset of 21,020 images and achieved an accuracy of 83.96%. The Jordanian currency includes nine denominations, and the images are 1024×1024 pixels.

III. MATERIALS AND METHODOLOGY

A. Proposed Dataset

The dataset was collected from college students whose ages ranged from 18-22. The students used their phones, which have various characteristics. Each person was asked to capture the currency on both sides, and each side was captured from two angles.

Fig. 1 shows sample pictures from each denomination. The captured images were taken in different conditions of lighting, orientation, background, sizes, and quality. The images include many challenges, such as background objects, some parts of the currency not being captured, being taken from different distances, some parts of the currency being damaged, being captured in different lighting conditions, and being captured from different phones. The captured images were resized into 448×448 pixels. The dataset includes 2799 images with JPEG formats represent seven denominations of the Jordanian currency as follows: 50JD: 409 images, 20JD: 397 images, 10JD: 419 images, 5JD:419 images, 1JD: 425 images, 50 piasters: 363 images, 25 piasters: 367, as shown in Fig 2. The Jordanian banknotes and coins have old and new shapes, and both are used now. There are differences in the color, security features, and regions, as shown in Table I which makes it difficult to recognize them.

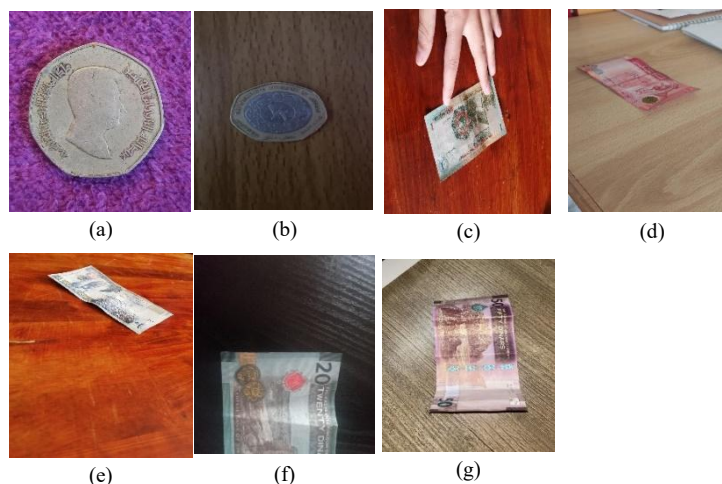


Fig. 1. Sample pictures from the proposed Jordanian currency dataset including images taken under different conditions: (a) 25 piasters (b) 50 piasters (c) 1JD (d) 5 JD (e) 10JD (f) 20 JD (g) 50 JD.

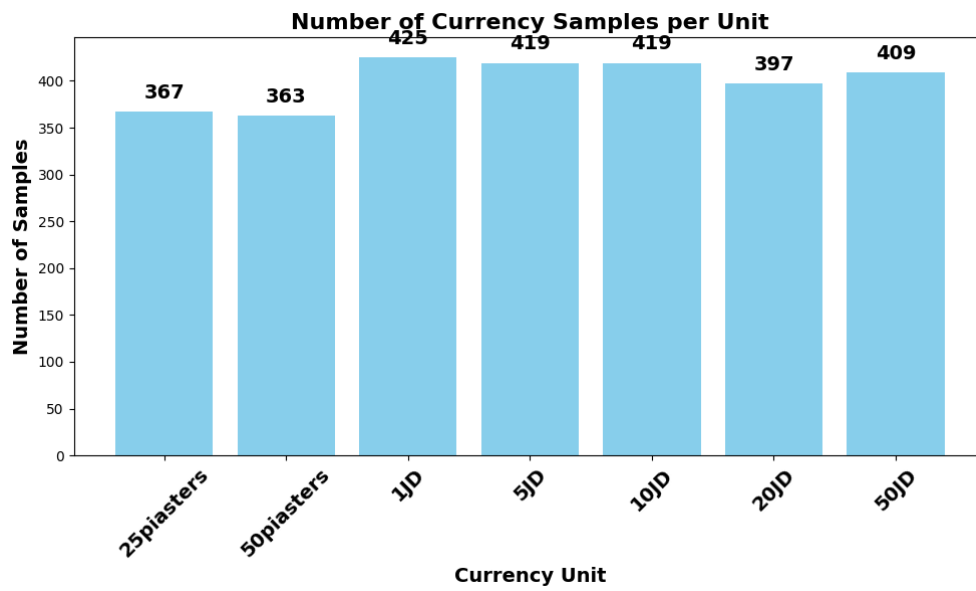


Fig. 2. Number of collected samples for each currency unit.

TABLE I. SAMPLE PICTURES FROM OLD AND NEW SHAPES FOR THE CURRENCIES IN JORDAN

Denomination	Old shape	New shape
25 piasters		
50 piasters		
1JD		
5JD		
10JD		
20JD		
50JD		

B. CNN Architectures

A set of popular CNN architectures was adopted in this paper. VGG has a simple structure and is considered a baseline for image classification tasks. ResNet shows a significant result when training deep networks with residual connections. DenseNet121 improved feature reuse by concatenating features from different layers.

The research in study [12] VGG developed at the University of Oxford by the Visual Geometry Group (VGG) in 2015. They studied the effect of increasing the number of convolutional layers from 16 to 19. The architecture achieved state-of-the-art on the ImageNet Challenge 2014. VGG16 has a simple and effective structure; it includes a sequence of five blocks. Each block consists of two to four convolutional layers followed by a max pooling. Finally, the features are flattened using four fully connected layers. The last layer represents the number of classes. The convolutional layers are 3x3 windows, and filter sizes range from 64 to 512. The max pooling is a window of 2x2 to reduce the feature size to half. Relu is used for activations, Table II explains the architecture of VGG16 model.

TABLE II. THE ARCHITECTURE OF VGG16

Layer Type	Filter Size	Number of Filters	Output Size	Details
Input	-	-	224x224x3	RGB image input
Convolution (x2)	3x3	64	224x224x64	Two 3x3 conv layers with ReLU
Max Pooling	2x2	-	112x112x64	Stride 2
Convolution (x2)	3x3	128	112x112x128	Two 3x3 conv layers with ReLU
Max Pooling	2x2	-	56x56x128	Stride 2
Convolution (x3)	3x3	256	56x56x256	Three 3x3 conv layers with ReLU
Max Pooling	2x2	-	28x28x256	Stride 2
Convolution (x3)	3x3	512	28x28x512	Three 3x3 conv layers with ReLU
Max Pooling	2x2	-	14x14x512	Stride 2
Convolution (x3)	3x3	512	14x14x512	Three 3x3 conv layers with ReLU
Max Pooling	2x2	-	7x7x512	Stride 2
Fully Connected (x2)	-	4096	1x1x4096	Two fully connected layers with ReLU
Output (SoftMax)	-	Number of classes	1x1xNumber of classes	Classification layer

DenseNet introduced [13] the dense connection between CNN layers. Each CNN layer is connected with other layers in a forward manner. The architectures achieved state-of-the-art results on ImageNet, SVHN, CIFAR-10, and CIFAR-100 datasets. The architecture includes four Dense Blocks and three Transition Layers. The Dense Block consists of a group of convolutional layers that are densely connected. Each dense

layer includes Batch Normalization, ReLU, 1x1 Convolution and 3x3 Convolution. At the same time, the Transition layers include 1x1 Convolution and 2x2 Average Pooling layers to reduce the size of the feature map, Table III presents the architecture of DenseNet121 model.

ResNet baseline architecture [14] was derived from the VGG architecture. It uses the same filler size of 3x3 and a pooling layer to down sample the feature map. However, it reduced the number of filters compared with VGG, which reduced the model size and introduced the residual networks instead of learning unreferenced data from the network layer. The network is 8x deeper than VGG; however, it is reported that the mAP of VGG16 and ResNet101 on PASCAL VOC 2007/2012, was 70.4% and 73.8%, with 3.2% improvement. In comparison, the VGG16 gained a 6% improvement (21.2% TO 27.2%) on the COCO dataset compared to VGG16. Stated the first place in the ILSVRC 2015 classification competition. The residual is calculated according to Eq. (1) if the input and output feature map have the same dimensions; if not, Eq. (2) is used.

$$y = \mathcal{F}(x, W_i) + x \quad (1)$$

TABLE III. ARCHITECTURE OF DENSENET121

Layer Type	Output Size	Filter Size / Details
Input	224x224x3	RGB image
Convolution	112x112x64	7x7 conv, stride 2
Max Pooling	56x56x64	3x3, stride 2
Dense Block 1	56x56x256	6 bottleneck layers (growth=32)
Transition Layer 1	28x28x128	1x1 conv, 2x2 avg pool
Dense Block 2	28x28x512	12 bottleneck layers
Transition Layer 2	14x14x256	1x1 conv, 2x2 avg pool
Dense Block 3	14x14x1024	24 bottleneck layers
Transition Layer 3	7x7x512	1x1 conv, 2x2 avg pool
Dense Block 4	7x7x1024	16 bottleneck layers
Global Avg Pooling	1x1x1024	
Fully Connected Layer	Number-of-classes	

where \mathcal{F} is the residual mapping, x and y are the input and the output. W_i is the weights of the convolutional layers in the residual block.

$$y = \mathcal{F}(x, W_i) + W_s x \quad (2)$$

W_s is the weight of the 1x1 convolution used for projection.

InceptionV3 [26] reported that it is computationally more efficient compared with VGG architecture. Therefore, it can be trained and tested using bigger data. The architecture includes a sequence of inception modules. Each inception module represents several convolutional and pooling layers that operate in a parallel fashion and are then concatenated together. As shown in Fig. 3 [27], there are 1x1 convolutional layers used to reduce the feature size and feature extraction: 3x3 and 5x5 convolutional layers to capture spatial features. Max pooling and average pooling are used for feature-down sampling.

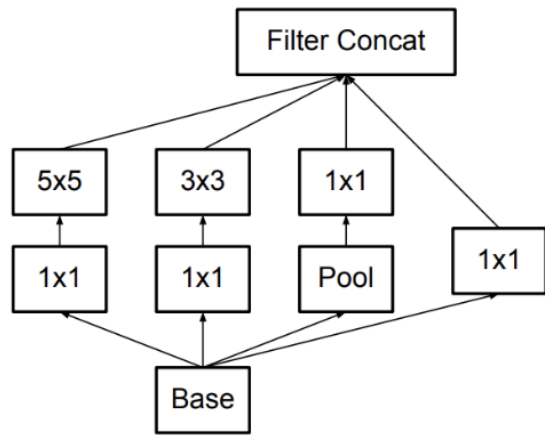


Fig. 3. The inception module.

C. Multiscale VGG

VGG net has a simple structure of a sequence of CNN blocks. Each block includes several convolution layers with different filters with a fixed size, followed by a max pooling layer to reduce the feature size by 50% compared with the previous block. Fig. 4 and Fig. 5 show the proposed multiscale VGG architecture to recognize currency. The proposed model improved the VGG model architecture by implementing a

multiscale VGG net to recognize Jordanian currency. The multiscale VGG net captures features with different respective fields compared with VGG architecture. Using multiple scales images, the last max-pooling layers of VGG extract four scales of feature map 7×7 , 5×5 , 3×3 , and 1×1 . The proposed multiscale VGG improves recognition rates by handling the variation of captured image sizes. The model can be applied to other currencies. The model learns text, color, security patterns and symbol features that exist in all other currencies. The trained model can also be fine-tuned and trained one other currencies, which can reduce the training time. The proposed dataset includes images with 448×448 size; the images are cropped from the center into 224×224 . The input image with 224×224 size is resized into four scales: scale 1: 100%, scale 2: 75%, scale 3: 50%, and scale 4: 25% and passed into parallel four VGG nets. The outputs of the four scales are flattened and concatenated, then passed into a fully connected layer of 256 size, followed by a 50% dropout layer, and finally, a fully connected layer with size 7 (number of classes). The four scales capture more fine-grained local features (e.g., edges or textures) compared with one scale feature. The final extracted feature before the first flattened layer encodes the spatial and semantic and includes the most informative features for classification. The sizes of extracted features from the last block of convolutions are 7×7 for scale 1, 5×5 for scale 2, 3×3 for scale three and 1×1 for scale 4.

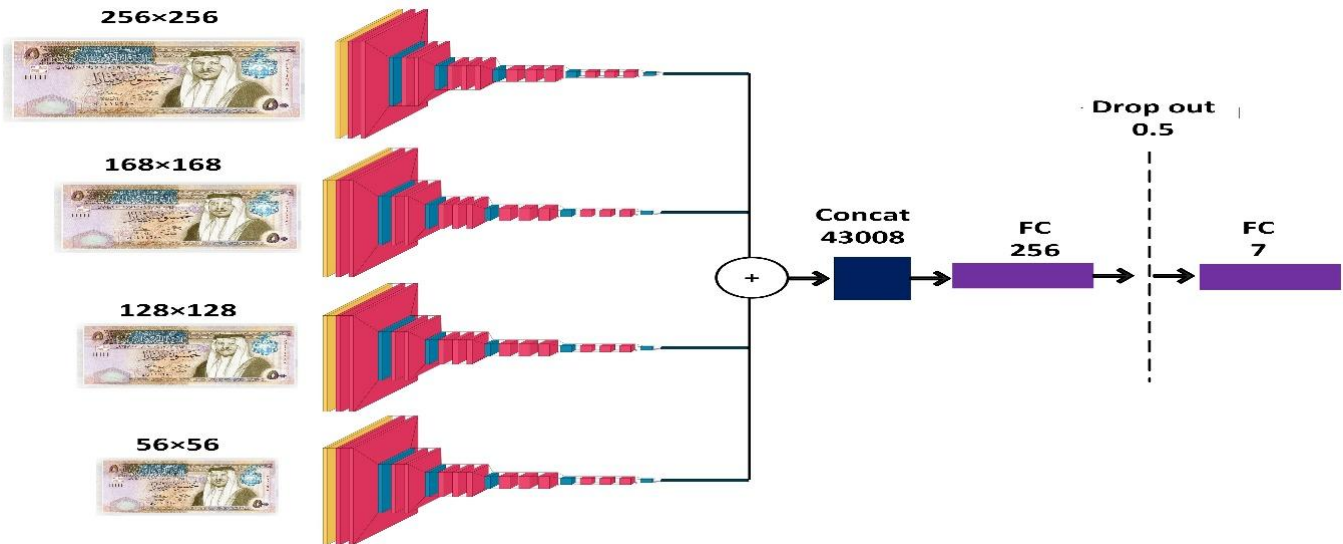
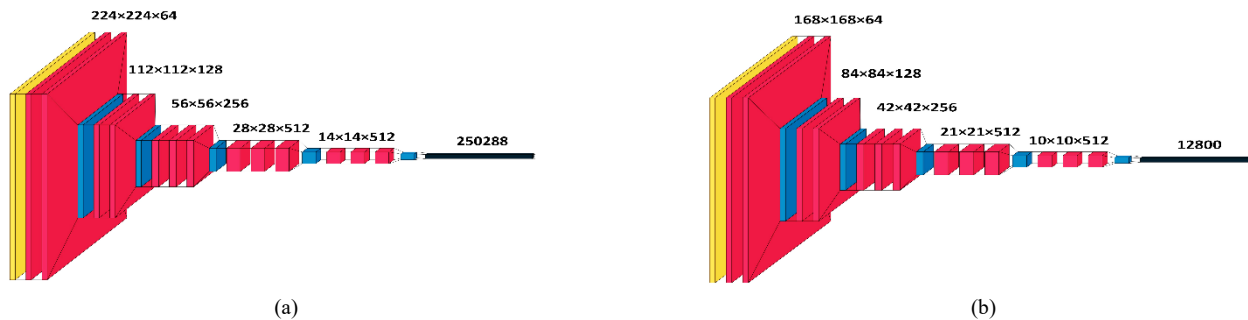


Fig. 4. The architecture of the proposed automatic multiscale currency recognition system.



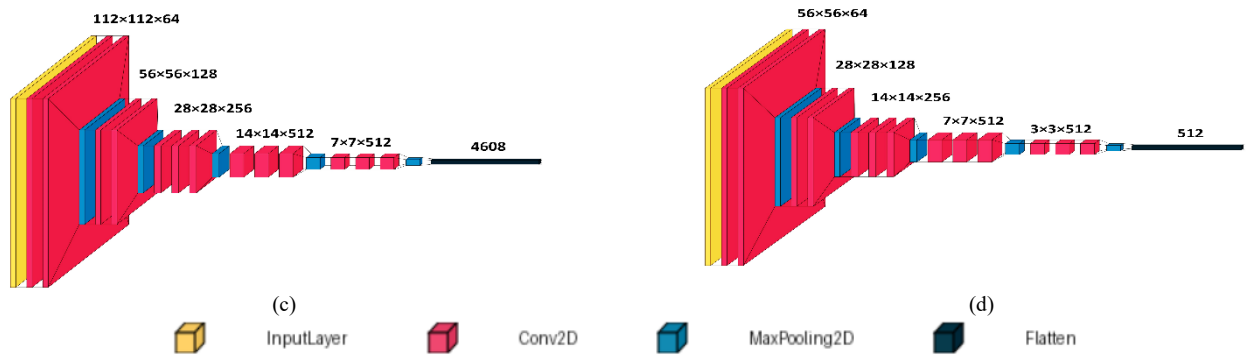


Fig. 5. Four scales of VGG16 architecture and the input images are (a) 224×224 (b) 168×168 (c) 112×112 (d) 56×56.

IV. IMPLEMENTATION DETAILS

A. Experimental Environment

The experiments were done in a Windows 10 and Jupyter Notebook environment. The processor is an Intel(R) Core (TM) i5-8600K CPU @ 3.60GHz. The installed RAM is 40.0 GB, and the GPU is a GTX1080 with 8 GB memory.

B. Model Parameter Settings

Table IV shows the parameters that were used to perform data augmentation while training the model.

TABLE IV. DATA AUGMENTATION USED DURING TRAINING

Augmentation Parameter	Value	Description
Rotation Range	20	up to 20 degrees
Width Shift Range	0.2	Horizontal shift by up to 20% of the image width
Height Shift Range	0.2	Vertical shift by up to 20% of the image height
Shear Range	0.2	up to 20%
Zoom Range	0.2	Zoom in or out by up to 20%
Horizontal Flip	TRUE	Randomly flips the image horizontally

C. Performance Metrics

Four metrics were used to evaluate the performance of the different CNN models in the dataset.

Accuracy: It measures the total of correctly predicted images compared to the total number of tested images.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (3)$$

Precision: The percentage of true positive compared with all positive prediction; high precision indicates fewer false positive

$$\text{Precision} = \frac{TP}{TP+FP} \times 100 \quad (4)$$

Sensitivity (Recall): It measures the model's capacity to detect every real positive case. Therefore, the ratio of the true positives to the total of the true positives and false negatives is used to compute it.

$$\text{Sensitivity} = \frac{TP}{TP+FN} \times 100 \quad (5)$$

Specificity: It calculates the proportion of accurately categorized negative events to all negative instances that really occurred:

$$\text{Specificity} = \frac{TN}{TN+FP} \times 100 \quad (6)$$

where *TP* denotes True Positive, *TN* is True negative, *FP* is False Positive and *FN* represent False negative.

D. Experimental Results

This section shows the performance of the proposed Multiscale VGG net and compares it with the VGG16, DenseNet12, ResNet50, ResNet101, and InceptionV3. Four metrics are used in the evaluation: accuracy, precision, sensitivity, and specificity. 70% of the dataset samples were used in training the model, whereas 30% were used to test the model. As shown in Table V, using multiple scales of the VGG16 improved the recognition rates compared with the base model VGG16. VGG16 achieved 98.92% accuracy, whereas adding features from multiple scales improved accuracy in all combinations. The highest accuracy occurred when combining features from scale 1 and scale 0.75, where the accuracy reached 99.88%, with a 0.96% improvement compared with VGG16. The precision, sensitivity, and specificity reached 99.89%, 99.89%, and 99.98%, respectively.

TABLE V. PERFORMANCE EVALUATION OF PROPOSED MULTISCALE VGG AT DIFFERENT SCALE VALUES 100%, 75%, 50% AND 25%

Models	Accuracy	Precision	Sensitivity	Specificity
VGG16	98.92%	98.96%	98.90%	99.82%
VGG16 SCALES 1,0.75,0.5,0.25	99.64%	99.66%	99.64%	99.94%
VGG16 SCALES 1,0.75	99.88%	99.89%	99.88%	99.98%
VGG16 SCALES 1,0.75,0.5	99.88%	99.88%	99.89%	99.98%
VGG16 SCALES 1,0.5,0.25	99.76%	99.78%	99.76%	99.96%
VGG16 SCALES 1,0.5	99.76%	99.78%	99.75%	99.96%

Table VI shows that the proposed method and InceptionV3 achieved 99.88% accuracy, after DenseNet121 with 99.52% accuracy, followed by ResNet101 with 99.28% accuracy. The precision, sensitivity, and specificity metrics showed that the

proposed multiscale and InceptionV3 achieved the best result, followed by DenseNet121 and ResNet101, respectively. Table VII shows the inference time and the model parameters. The VGG16 SCALES 1,0.75 runs in real-time and requires 200 ms to predict the class label for a patch of 32 images. At the same time, the model used 39,130,759 parameters.

TABLE VI. PERFORMANCE EVALUATION OF PROPOSED MULTISCALE VGG WITH OTHER CNN KNOWN CNN ARCHITECTURES

Models	Accuracy	Precision	Sensitivity	Specificity
VGG16	98.92%	98.96%	98.90%	99.82%
VGG16 SCALES 1,0.75	99.88%	99.88%	99.89%	99.98%
DenseNet121	99.52%	99.52%	99.52%	99.92%
ResNet50	99.16%	99.18%	99.17%	99.86%
ResNet101	99.28%	99.31%	99.30%	99.88%
InceptionV3	99.88%	99.89%	99.88%	99.98%

TABLE VII. INFERENCE TIME IN TERMS OF MS PER PATCH (PATCH=32 IMAGES) AND NUMBER OF PARAMETERS USED BY THE MODEL

Models	MS/Patch	#Parameters
VGG16	81	21,139,271
VGG16 SCALES 1,0.75	200	39,130,759
DenseNet121	78	19,884,615
ResNet101	146	68,350,343
InceptionV3	97	22,329,127

836 images were used to test the proposed architecture. The confusion matrix in Fig. 6 shows that the Multiscale VGG architecture accurately predicted 835 pictures, and only one image for 10 JD was predicted to be 50 JD.

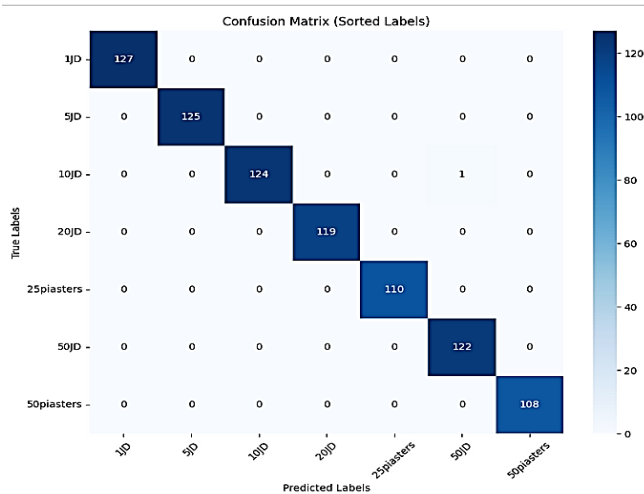


Fig. 6. Confusion matrix for the proposed multiscale VGG.

V. CONCLUSION

This paper proposed a dataset for Jordanian currencies and proposed Multiscale VGG architectures to recognize Jordanian currencies automatically. The dataset includes 2799 images for seven denominations (25piasters, 50piasters, 1JD, 5JD, 10JD, 20JD, 50JD). The images were captured in different conditions of lighting, backgrounds, clutter, occlusion, and orientation. Moreover, the experimental results showed that the proposed

Multiscale VGG architectures achieved the best accuracy, precision, sensitivity, and specificity compared with VGG16, DenseNet121, ResNet50, and ResNet101.

REFERENCES

- [1] D. S. Aljutaili, R. A. Almutlaq, S. A. Alharbi, and D. M. Ibrahim, "A Speeded up Robust Scale-Invariant Feature Transform Currency Recognition Algorithm," vol. 12, no. 6, pp. 346–351, 2018.
- [2] P. Garkoti, P. Mishra, N. Rakesh, Payal, M. Kaur, and P. Nand, "Indian Currency Recognition System Using Image Processing Techniques," in 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom), 2022, pp. 628–631.
- [3] R. Bourne et al., "Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study," *Lancet Glob. Heal.*, vol. 9, no. 2, pp. e130–e143, Feb. 2021.
- [4] K. Sharifani and M. Amini, "Machine Learning: A Review of Methods and Applications," *World Inf. Technol. Eng. J.*, vol. 10, no. 7, pp. 3897–3904, 2023.
- [5] S. Salih and T. Nasih, "Image-Based Processing of Paper Currency Recognition and Fake Identification: A Review," *Technium*, vol. 3, no. 7, pp. 46–63, 2021.
- [6] J. Lee, H. Hong, K. Kim, and K. Park, "A Survey on Banknote Recognition Methods by Various Sensors," *Sensors*, vol. 17, no. 2, p. 313, Feb. 2017.
- [7] N. Dalal, B. Triggs, N. Dalal, and B. Triggs, "Histograms of Oriented Gradients for Human Detection To cite this version: Histograms of Oriented Gradients for Human Detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005, pp. 886–893.
- [8] A. Xu and G. Namit, "SURF: Speeded-Up Robust Features COMP 558-Project Report," pp. 2–29, 2008.
- [9] D. G. Lowe, "Distinctive Image Features from Scale Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
- [10] D. Rika Widianita, "Combining Handcrafted and Deep Features For Scene Image Classification," *J. Data Acquis. Process.*, vol. 38, no. 3, p. 2158, 2023.
- [11] M. N. Abdi and M. Khemakhem, "Arabic writer identification and verification using template matching analysis of texture," *Proc. - 2012 IEEE 12th Int. Conf. Comput. Inf. Technol. CIT 2012*, pp. 592–597, 2012.
- [12] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–14, 2015.
- [13] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, vol. 39, no. 9, pp. 2261–2269.
- [14] K. He, X. Zhang, S. Ren, and S. Jian, "Deep Residual Learning for Image Recognition," in *In Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [15] C. Szegedy et al., "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.
- [16] K. Reddy, G. Ramesh, C. Raghavendra, C. Sravani, M. Kaur, and R. Soujanya, "An Automated System for Indian Currency Classification and Detection using CNN," *E3S Web Conf.*, vol. 430, p. 01077, Oct. 2023.
- [17] M. Z. Hanif, W. A. Saputra, Y. H. Choo, and A. P. Yunus, "Rupiah Banknotes Detection : Comparison of The Faster R-CNN Algorithm and YOLOv5," *J. INFOTEL*, vol. 16, no. 3, pp. 502–517, 2024.
- [18] C. Park and K. R. Park, "MBDM: Multinational Banknote Detecting Model for Assisting Visually Impaired People," *Mathematics*, vol. 11, no. 6, 2023.
- [19] S. Alghyaline, "Optimised CNN Architectures for Handwritten Arabic Character Recognition," *Comput. Mater. Contin.*, vol. 79, no. 3, pp. 4905–4924, 2024.
- [20] E. AL-Edreesi and G. Al-Gaphari, "Real-time Yemeni Currency Detection," *arXiv Prepr. arXiv:2406.13034*, 2024.

- [21] I. O. A. Omeiza, O. Ogunbiyi, O. Y. Ogundepo, A. O. Otuoze, D. O. Egbune, and K. Osunsanya, "A Method of Colour-Histogram Matching for Nigerian Paper Currency Notes Classification," *Jordan J. Electr. Eng.*, vol. 9, no. 1, pp. 42–59, 2023.
- [22] M. C. GENÇAL, "U.S. Banknotes Recognition By SURF Features," 2nd Int. Congr. Innov. Technol. Eng., pp. 68–72, 2022.
- [23] G. S. Hussein, S. Elseuofi, W. H. Dukhan, and A. H. Ali, "A Novel Method for Banknote Recognition Using a Combined Histogram of Oriented Gradients and Scale-Invariant Feature Transform," *Inf. Sci. Lett.*, vol. 12, no. 9, pp. 2121–2131, 2023.
- [24] A. Mukundan, Y. M. Tsao, W. M. Cheng, F. C. Lin, and H. C. Wang, "Automatic Counterfeit Currency Detection Using a Novel Snapshot Hyperspectral Imaging Algorithm," *Sensors*, vol. 23, no. 4, pp. 1–14, 2023.
- [25] V. Meshram, V. Meshram, K. Patil, Y. Suryawanshi, and P. Chumchu, "A comprehensive dataset of damaged banknotes in Indian currency (Rupees) for analysis and classification," *Data Br.*, vol. 51, p. 109699, 2023.
- [26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *arXiv Prepr. arXiv1512.00567*, 2015.
- [27] C. Szegedy, Y. J. Wei Liu, P. Sermanet, V. Scott Reed, Dragomir Anguelov, Dumitru Erhan Vincent, and A. Rabinovich, "Going Deeper with Convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

Foreground Feature-Guided Camouflage Image Generation

Yuelin Chen¹, Yuefan An², Yonsen Huang³, Xiaodong Cai⁴

School of Mechanical and Electrical Engineering, Guilin University of Electronic Technology, Guilin, China^{1,2}

School of Information and Communication, Guilin University of Electronic Technology, Guilin, China^{3,4}

Abstract—In the field of visual camouflage, generating a high-quality background image that seamlessly blends with complex foreground objects and diverse background environments is a critical task. When dealing with such complex scenes, the existing techniques have insufficient foreground feature extraction, resulting in insufficient fusion of the generated background image with the foreground objects, making it difficult to achieve the desired camouflage effect. In order to solve this problem and achieve the goal of higher quality visual camouflage effect, this paper proposes a new foreground feature-guided camouflage image generation method (Object Enhancement Module - Diffusion Refinement, OEM-DR), which generates camouflage images by enhancing the foreground features to guide the background. The method firstly designs a new object enhancement module to optimize the attention mechanism of the model, and eliminates the attention weights that have less influence on the output through pruning strategy, so that the model focuses more on the key features of the foreground objects, and thus guides the generation of the background more effectively. Second, a novel detail optimization framework based on diffusion strategy is constructed, which maintains the integrity of the global structure of the image while performing fine optimization processing on the local details of the image. In experiments on standard camouflaged image datasets, the proposed method in this study achieves significant improvement in both FID (Fréchet Inception Distance) and KID (Kullback-Leibler Divergence) evaluation metrics, which verifies the feasibility of the method. This suggests that by strengthening foreground features and detail optimization, the fusion between background images and foreground objects can be effectively improved to achieve higher quality visual camouflage effects.

Keywords—Camouflage image; foreground features; object enhancement; detail optimization

I. INTRODUCTION

In the field of visual perception, camouflage image generation is a challenging task that aims at generating background images that can skillfully mask foreground objects for visual concealment. This technique plays an important role in several practical application areas such as pest detection, healthcare [2], and autonomous driving [8]. With the advancement of computer vision techniques, especially in the fields of style migration [9], image editing [10] and image generation [11], new ideas have been provided to address the challenges of camouflage image generation. The Poisson image editing method proposed by Di Martino et al. [10] brought innovations in the field of image editing by allowing researchers to work with the image in a natural and intuitive manner content. The pioneering work of Chu et al. [12] on camouflage image

generation demonstrated how to generate hard-to-detect images by mimicking the camouflage mechanisms of natural organisms. The work of Huang and Belongie [9] further advanced the development of style-migration techniques, which allow us to change the style of an image to fit different contexts while keeping its content intact. Zhang et al. [13] proposed generating camouflaged images that can blend in with complex backgrounds by learning a large amount of natural image data. Li et al. [14] further proposed a camouflaged image generation network that does not require specific positional information. The work of Zheng et al. [15] provides a new solution for high fidelity image complementation by bridging global contextual interactions. Lugmayr et al. [16] provided a new idea for background complementation of camouflaged images by using denoising diffusion probabilistic model for image restoration.

However, despite the many advancements in existing technologies, several key issues remain. Firstly, most methods rely on manually selected backgrounds, which not only limits the diversity of generated samples but also significantly increases the cost of data collection. Secondly, these methods may perform poorly in complex and variable environments, as they often depend heavily on the precise extraction of background and foreground features. The LAKE-RED model [17], although innovative in generating camouflage images by fusing training backgrounds with extracted foreground features, may face challenges in complex or changing environments due to its reliance on precise feature extraction.

To address these issues, this study proposes a new object enhancement strategy. Inspired by the work of Dhariwal P et al. [19], this study utilizes weight sparsification and pruning to enhance the model's understanding and learning of deep features of target objects. This strategy aims to reduce dependence on precise feature extraction and improve the model's adaptability in complex environments.

Furthermore, camouflage images often lack detail optimization during generation, which can weaken their camouflage effect. Unnatural transitions in texture, color, or edge areas may reduce their concealment. To solve this, inspired by the work of Yang L et al. [18], this study designs a method to enhance the model's feature expression ability using non-zero weights after diffusion pruning, optimizing the connection between foreground and background to improve the quality of camouflage images.

In the following sections of this paper, the study will be explored in depth. The related work in Section II will review and analyze the technologies and methods involved in this study,

clarifying its position and value in the field of camouflage image generation. The OEM-DR in Section III will detail the proposed new method, including the design and implementation of the object enhancement strategy, and how non-zero weights after diffusion pruning enhance the model's feature expression. The experiments in Section IV will design and conduct a series of experiments to verify the effectiveness and performance of the proposed method, including comparative and ablation experiments with existing methods. Finally, the conclusion in Section V will summarize the main achievements and contributions of this study, discuss existing limitations, and suggest directions for future work.

II. RELATED WORK

A. Camouflage Image Generation

The goal of camouflage image generation is to create images of target objects that blend highly with the background and are difficult to detect. Early methods mainly relied on image processing techniques. For example, Chu et al. [12] hid the target by inserting similar textures and colors through image editing. Although these methods could achieve basic camouflage, the quality and diversity of the generated images were limited. With the development of deep learning, methods based on deep generative models have become mainstream. Zhang et al. [13] proposed a deep camouflage image generation model that learns complex camouflage patterns and generates more realistic images. Li et al. [14] developed a camouflage generation network without location information, which automatically learns the relationship between the target and the background, thereby improving the quality of the generated images. In the latest research, Zhao et al. [17] introduced the LAKE-RED method, which combines latent background knowledge retrieval with diffusion models to generate high-fidelity and diverse camouflage images. Yang et al. [18] provided a comprehensive review of diffusion models, covering their applications in camouflage image generation and other fields, offering valuable references for research. These advancements indicate that deep learning, especially diffusion models, holds great potential in camouflage image generation. Future research can further explore their combination with camouflage tasks to generate higher quality and more diverse camouflage images.

B. Data Augmentation

Data augmentation is crucial for enhancing the performance of deep learning models, especially when the number of samples is limited. It improves the model's generalization ability by generating more training samples. In the tasks of camouflage image detection and segmentation, data augmentation is particularly important because it is difficult to obtain camouflage images and their patterns are diverse. Fan et al. [1] enhanced the data by randomly inserting camouflage targets, which is a simple but effective way to improve detection performance. Le et al. [4] explored data augmentation in their Anabran network, such as random cropping and rotation, but mainly focused on geometric transformations and simple editing, which had limited enhancement of the diversity of

camouflage patterns. The development of diffusion models has brought new ideas to data augmentation. Dhariwal and Nichol [19] proposed an image synthesis method based on diffusion models, which generates high-quality images and performs excellently in multiple tasks, providing a new direction for camouflage image generation and data augmentation. Binkowski et al. [20] improved the MMD GAN training method, enhancing the quality and diversity of the generated images. Heusel et al. [21] introduced a two-time-scale update rule for GAN training, proving that it converges to a local Nash equilibrium, providing theoretical support for the training of generative models. These studies show that combining advanced generative models such as diffusion models with the task of camouflage image generation can generate richer and more diverse training data. Future research can further explore this combination to improve the performance of camouflage image detection and segmentation models.

III. OEM-DR MODEL

A. An Overview of the Proposed Framework

As shown in Fig. 1, the OEM-DR model starts from low-level image feature extraction and highlights important details in the image through the object enhancement module, which introduces sparsity to optimize the weight distribution, and then employs a pruning strategy to reduce the computational burden. The detail optimization module further enhances the image quality, which facilitates the fusion and generalization of the model to new data while retaining key information through the diffusion process. Finally, the model performance is tuned and optimized through the computation of the loss function. The specific details of these steps are elaborated in this section.

B. Object Enhancement Based on Sparse Pruning

In order to address the model's deficiency in object information learning, an object enhancement module is designed in this study. The purpose of this module is to deepen the model's understanding and learning of deep features of the target object through refined feature selection and enhancement. This study aims to further encode the object features through this enhancement module to achieve deeper understanding of foreground feature information in images. The working mechanism of this module is explained in detail in the following section, and its structure is detailed in Fig. 2.

In the forward propagation process of the model, the deep representation of foreground features f_g is extracted by a multilayer perceptron (MLP). the pre-trained background embedding matrix is subsequently transposed and the batch dimension is increased in order to obtain the background embedding vector C . The query vector q is obtained by applying a linear transformation layer to $f_g.C$ is used as the key k and the value v . The similarity matrix S_m is computed by using the einsum function for querying q and key k . matrix S_m , which is computed as follows:

$$S_m = \text{einsum}(q, k) \quad (1)$$

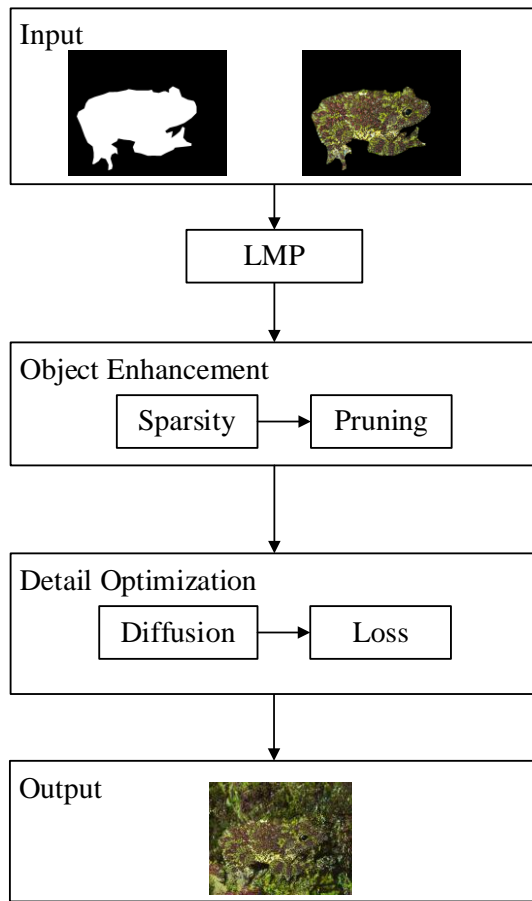


Fig. 1. OEM-DR model framework.

In this paper, we quantify sparsity by calculating the attentional weight of S_m and subsequently determining the pruning threshold. The L_1 norm l_{n1} and L_2 norm l_{n2} of the attention weight matrix A are computed with the following equations:

$$l_{n1} = \sum_{i=1}^n |a_i| \quad (2)$$

$$l_{n2} = \sqrt{\sum_{i=1}^n a_i^2} \quad (3)$$

Where a_i is the i -th element in the attention weight matrix A and n is the total number of elements.

The sparsity ratio R is defined as follows:

$$R = \frac{l_{n1}}{l_{n2} + \varepsilon} \quad (4)$$

Where ε is a negligible positive number added to prevent division by zero.

A Boolean mask P is generated using the sparse ratio R . Each element of this mask is determined by the condition:

$$P = \begin{cases} T, A > R \\ F, A \leq R \end{cases} \quad (5)$$

The initial pruning is accomplished by setting the elements of A for which P is F to zero through the product operation, as follows:

$$A_p = A \square P \quad (6)$$

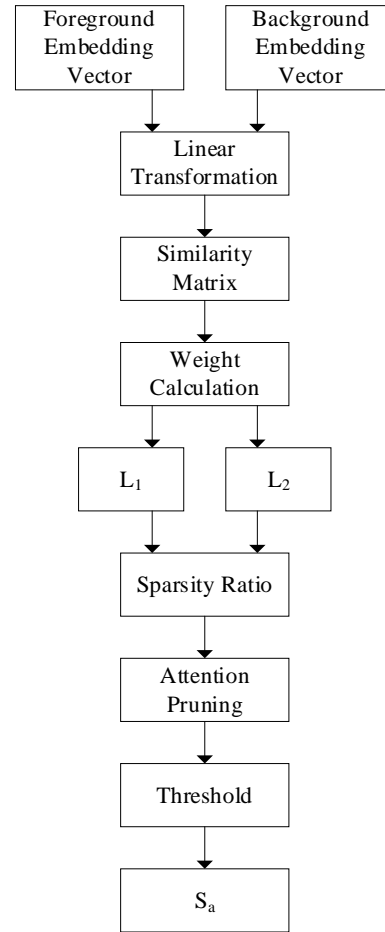


Fig. 2. Object enhancement module structure diagram.

Further, the L_1 norm A_{p1} of the pruned attention weight A_p is calculated using Eq. (2) and the pruning threshold T_p for each attention head is determined by the pruning ratio P_r as follows:

$$M_{AP} = \max(A_{p1}) \quad (7)$$

$$T_p = M_{AP} \times P_r \quad (8)$$

M_{AP} in Eq. (7) represents the maximum value of the computed L_1 norm on the last dimension of the weight matrix A_p .

The pruned weight matrix S_a is then obtained by generating a Boolean mask using the pruning threshold T_p through the operations of Eq. (5) and Eq. (6), and then the elements of A_p Ap below T_p are set to zero through the product operation in order to complete the pruning process.

C. Diffusion-based Detail Optimizer

In order to cope with the performance degradation caused by performing pruning, this study adds the DR module after pruning. the purpose of the DR module is to maintain or even improve the predictive performance of the model after pruning, and to enhance the generalization ability of the model to ensure the model's adaptability and accuracy to new data.

Referring to Fig. 3, the input to the DR module is the pruned weight matrix S_a . The influence of the non-zero weights retained after pruning is spread to the adjacent zero-weight regions through an iterative process, with a view to maintaining or even enhancing the generalization capability of the model while reducing the complexity of the model. The details of the operation are described below.

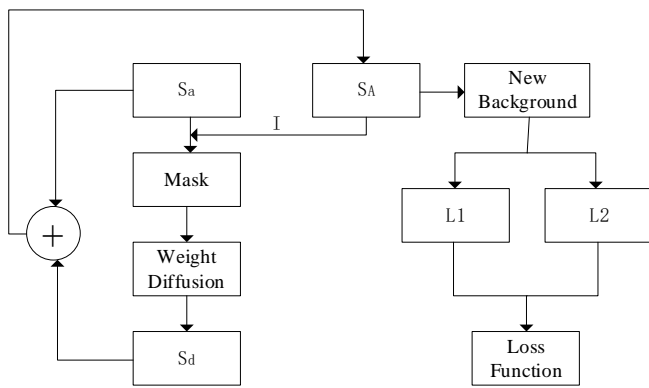


Fig. 3. Diffusion optimizer structure diagram.

First, a Boolean mask M is initialized to identify the location of the non-zero weights in S_a . This mask is obtained by comparing S_a with zero, where the non-zero element corresponds to a mask value of True.

Next, I iterations are executed, and in each iteration the following steps are performed:

First M is shifted one position to the right along the last dimension (feature dimension) to obtain the new mask M_n . This step simulates the process of diffusion of non-zero weights to the right along the feature dimension.

Next, the update of the weights is computed as follows:

$$S_d = S_a \times (1 - D \times (1 - M_n)) \quad (9)$$

D is a factor between 0 and 1 used to adjust the magnitude of the weight update.

The calculated update is then applied to S_a and the updated weights are:

$$S_A = S_d + S_a \quad (10)$$

Finally, a non-negative constraint is imposed to ensure that all weight values are non-negative. After I iterations, the post-diffusion weight matrix S_A is obtained.

In the second step, L_1 and L_2 regularization losses are introduced to process the new background P_n obtained after training with the features T_a of the target image as a way to achieve fine tuning of the model complexity. The new foreground image P_n is generated in the following way: in the background region identified by the mask, this paper replaces the original foreground f with the background feature S_A ; while in the foreground region, f is kept unchanged. Thus, the updated foreground image P_n is obtained. The L_1 and L_2 regularization losses operate as follows:

The L_1 loss measures the error by calculating the absolute value of the difference between the predicted value and the target value, and its mathematical expression is shown below:

$$L_1 = \lambda_1 \sum_{i=1}^n |w_i - w_i^{T_a}| \quad (11)$$

The L_2 loss measures the error by calculating the square of the difference between the predicted value and the target value, the mathematical expression of which is shown below:

$$L_2 = \lambda_2 \sum_{i=1}^n (w_i - w_i^{T_a})^2 \quad (12)$$

The w_i denotes the actual value of the i -th image, $w_i^{T_a}$ denotes the target value of the i -th image, and λ_1 and λ_2 denote the weights occupied by the two losses, respectively.

The L_1 and L_2 losses are combined into the total loss according to certain weights to realize the joint control of model complexity. The combined total loss function is defined as:

$$L = L_1 + L_2 \quad (13)$$

The main role of L_1 and L_2 losses is to prevent model overfitting and improve model generalization. Specifically, the L_1 loss helps with feature selection, while the L_2 loss helps with parameter stability.

IV. EXPERIMENT

A. Datasets

Following the research on covert object detection (COD) [3], this paper uses 4040 real images (3040 from the COD10K [1] dataset and 1000 from the CAMO [26] dataset) to train the model. To validate the generation performance, this study collects image-mask pairs from different domains and constructs a test dataset consisting of three subsets: covert objects (CO), salient objects (SO) and general objects (GO) [17]. In CO, there are 6473 image pairs from CAMO [4], COD10K [1] and NC4K [5]. We then randomly selected 6473 images from well-known salient target detection datasets (e.g., DUTS [6], DUT-OMRON [7], etc.) and segmentation datasets to evaluate the performance of the model on open domain data.

B. Evaluation Metrics and Parameter Setting

Following the good practices of AIGC and COD, the InceptionNet-based metrics FID [20] and KID [17] are chosen in this paper to measure the quality of the generated covert images. Once the image features have been extracted via InceptionNet, the distance between them is calculated to indicate the level of similarity between the model output and the target dataset. Unlike normal images, well-synthesized covert images should not contain easily recognizable objects, and extracting discriminative features is more challenging. A lower value of FID [2] indicates that the generated image is more similar to the real image in terms of visual features, which usually implies a better generation. KID [21] by kernel method is able to capture more sensitively the differences between the generated image and the real image, especially in terms of image detailed features. A lower value of KID [21] indicates that the quality of the generated image is closer to the real image.

In this paper, we use a latent diffusion model, pre-trained on a restoration task as an initialization. The model was designed to handle images and masks of size 512×512 and compressed to a potential space of $128 \times 128 \times 3$ using the pre-trained VQ-VAE. During the training process, the focus is on improving the model's understanding and learning of the deep features of the target object through object augmentation and regularized diffusion strategies, and does not fine-tune the autoencoder and

decoder components. The existing conditions are optimized and enhanced by the modules proposed in this paper. Parameter optimizations, such as initialization, data enhancement, and batch size, are set similar to the original paper. Finally, the model generates the artifact images and resizes them to align with the original input. This paper uses PyTorch for all experiments and a GeForce RTX 4060ti GPU for all experiments.

C. Comparison and Analysis of Model Results

To verify the effectiveness of the OEM-DR model, this paper selects the following nine models for comparison: The AB model [9] achieves seamless image region blending based on the Poisson equation, enabling the source image to naturally blend into the target background. The AdaIN model [12] rapidly accomplishes real-time arbitrary style transfer through adaptive instance normalization technology, converting the style while maintaining the image content unchanged. In the field of camouflage image generation, the CI model [11] imitates the camouflage mechanism of natural organisms to generate images that are difficult to detect; the DCI model [13] and LCGNet model [14] utilize deep learning to generate camouflage images that blend with complex backgrounds, with LCGNet being more flexible as it does not require specific location information; the LAKE-RED model [17] combines potential background knowledge retrieval with diffusion models to enhance the quality of camouflage images. In addition, the TFill model [15] achieves high-fidelity image completion by bridging global context interaction, while the RePaint-L model [16] performs image restoration based on denoising diffusion probabilistic models, generating content for missing areas. The LDM model [10] proposes a high-resolution image synthesis method based on latent diffusion models, promoting the development of high-quality image generation technology.

Table I shows the comparison of experimental results of various models on the dataset. Compared with the AB, CI, AdaIN, DCI, LCCNet models, which realize the camouflage effect by fusing the background with the foreground, EM-DR guides the background generation according to the features of the foreground, so that the generated image background is closer to the foreground and the foreground is more hidden.

TABLE I. COMPARISON OF EXPERIMENTAL RESULTS IN VARIOUS MODELS

Methods	Input	Camouflaged Objects		Salient Objects		General Objects		Overall		
		FID↓	KID↓	FID↓	KID↓	FID↓	KID↓	FID↓	KID↓	
Image Blending	AB ^[9]	F+B	117.11	0.0645	126.78	0.0614	133.89	0.0645	120.2	0.0623
	CI ^[11]	F+B	124.49	0.0662	136.3	0.738	137.19	0.0713	128.5	0.0693
	AdaIN ^[12]	F+B	125.16	0.0721	133.2	0.0702	136.93	0.0714	126.9	0.0703
	DCI ^[13]	F+B	130.21	0.0689	134.92	0.0665	137.99	0.069	130.5	0.0673
	LCGNet ^[14]	F+B	129.8	0.0504	136.24	0.0597	132.64	0.0548	129.9	0.055
Image Inpainting	TFill ^[15]	F	63.74	0.0336	96.91	0.0453	122.44	0.0747	80.39	0.0438
	LDM ^[10]	F	58.65	0.038	107.38	0.0524	129.04	0.0748	84.48	0.0488
	RePaint-L ^[16]	F	76.8	0.0459	114.96	0.0497	136.18	0.0686	96.14	0.0498
	LAKE-RED ^[17]	F	39.55	0.0212	88.7	0.0428	102.67	0.0625	64.27	0.0355
	Ours	F	37.44	0.0181	86.9	0.0387	100.48	0.0581	61.52	0.0311

Compared with TFill and LDM models, EM-DR strengthens the feature screening of foreground features, mainly in its ability to refine and optimize, which gives it an advantage in preserving the details of important foreground objects in the image. TFill and LDM, on the other hand, while having their advantages in texture synthesis and learning-based approaches, may require additional tuning or optimization to better handle foreground features in specific cases. Compared to the benchmark model LAKE-RED, EM-DR achieves significant improvements in all metrics, further demonstrating its ability to adequately enhance foreground features when generating camouflaged images, thus providing foreground camouflage effects.

Compared with the baseline model on the CO dataset, the FID metrics improved by 5.3% and the KID metrics improved by 14.6%, on the SO model, the FID metrics improved by 2% and the KID metrics improved by 9.5%, on the GO model, the FID metrics improved by 2.1% and the KID metrics improved by 7%, and on the whole the FID metrics improved by 4.3% and the KID metrics improved by 12.3%. by 4.3% for FID and 12.3% for KID. These improvements indicate that the EM-DR model shows better performance in the task of generating camouflage images, and the improvements in FID and KID metrics indicate that the images generated by the new model are closer to the real images in terms of visual and statistical properties.

D. Ablation Study

In order to verify the impact of the object enhancement module and diffusion optimizer mentioned in this paper on the model performance, ablation experiments are conducted on datasets containing CO, SO and GO, and the results are summarized in Table II.

TABLE II. ABLATION STUDY RESULTS

Methods		OEM	DR
CO	FID↓	37.67	37.77
	KID↓	0.0183	0.0187
SO	FID↓	85.85	87.92
	KID↓	0.0381	0.0391
GO	FID↓	102.07	102.57
	KID↓	0.0599	0.0599
Overall	FID↓	61.73	62.51
	KID↓	0.0313	0.0317

First, the OEM module is individually validated in this paper. Compared with the LAKE-RED model, OEM achieves significant improvement in both FID and KID, especially on the SO dataset. This indicates that OEM not only makes the camouflage images generated on hidden objects as well as ordinary objects close to the real image in terms of enhanced foreground features, but also generates camouflage images on salient objects even closer to the real image.

Secondly, the DR module is individually validated in this paper. Compared to the LAKE-RED model, DR also achieves significant improvement in both FID and KID but lacks in SO dataset compared to the OEM module.

In summary, the OEM and DR modules show excellent advantages in the camouflage image generation task, effectively improving the model performance.

E. Example of Analysis

In order to visualize the improvement of the OEM-DR model mentioned in this paper compared to the benchmark model LAKE-RED, several examples are selected for visualization.

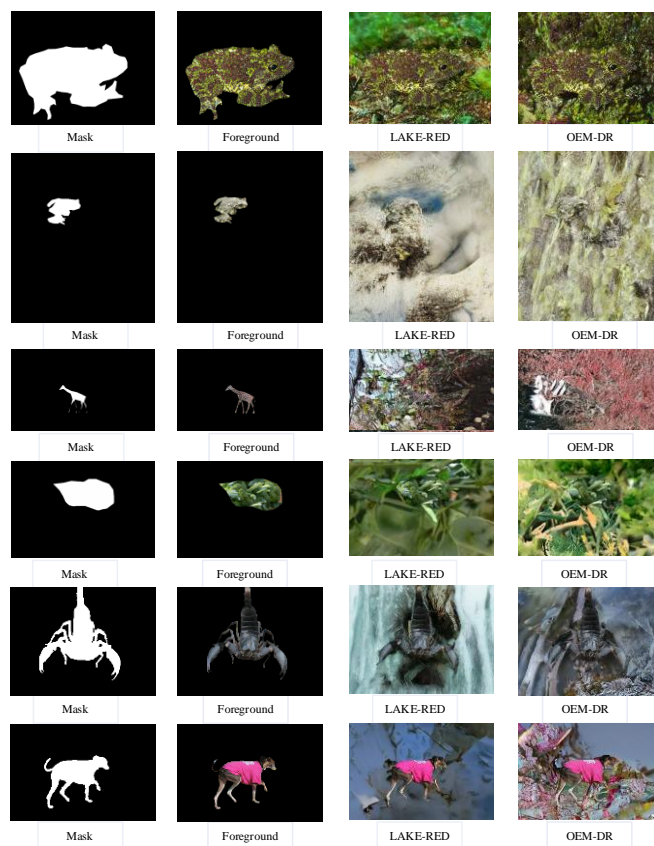


Fig. 4. OEM-DR Model visualization examples.

Fig. 4 visualizes the performance of the OEM-DR model proposed in this study compared with other benchmark models in generating camouflaged images. The figure groups the images according to different model sources: Fig. 4 (a) and 4 (b) are generated by the CO model, Fig. 4 (c) and 4 (d) by the SO model, while Fig. 4 (e) and 4 (f) are from the GO model. Taking the toad in Fig. 4 (a) as an example, the image generated by the OEM-DR model shows a more natural fusion between the background and the foreground, and the details of the background are optimized so that the foreground objects do not appear to be abrupt, which achieves a better camouflage effect visually. Similarly, in Fig. 4 (f), although the camouflaged objects are not completely hidden, the OEM-DR model still demonstrates its advantages in improving the overall image quality. The comprehensive comparison results show that the OEM-DR model has a significant advantage in generating images that contribute to effective camouflage, and is able to generate high-quality camouflage images that are more natural and rich in details.

Since both image generation quality and camouflage effectiveness need to be perceived by humans, this paper conducts a user study to obtain subjective judgments on the generation results. In this paper, 10 researchers in AI related

fields are invited to judge which of the images in Fig. 4 are not easily detectable. According to the judgment results except Fig. 4 (b) all others are considered to be less detectable in the foreground in the images generated by OEM-DR, which fully demonstrates the quality of OEM-DR in image generation and the effectiveness of the camouflage effect.

F. Discussion

In this study, the OEM-DR model significantly enhanced the performance of camouflage image generation through the object enhancement module and the diffusion optimizer. Compared with various existing models, OEM-DR excelled in terms of FID and KID metrics, particularly when dealing with salient objects. Although the model has the limitation of high computational resource demands, this is expected to be overcome in the future through algorithm optimization and hardware acceleration. The OEM-DR model provides a new perspective and methodology for the field of camouflage image generation, holding broad prospects for application.

V. CONCLUSION

In this paper, an innovative camouflage image generation method based on sparse pruning and weight diffusion is proposed. First, a new object enhancement module is designed, which effectively improves the extraction of key features in the foreground by the model, and thus significantly improves the authenticity of the camouflage image generation. Second, a detail optimization module based on the weight diffusion strategy is constructed, which improves the quality of the generated images by optimizing the background image generation process. In the experiments on CO, SO and GO datasets, this paper demonstrates excellent performance. Future research in camouflage visual perception will be continued to further propose feasible solutions.

However, it is worth noting that although this study has achieved certain results in the field of camouflage image generation, it faces the challenge of balancing computational complexity and efficiency during the implementation of the weight diffusion strategy. In future research work, this study will be committed to optimizing the adaptive weight diffusion strategy, dynamically adjusting the diffusion parameters according to the training progress of the model and the unique characteristics of the dataset. The exploration and optimization in this direction are aimed at promoting the continuous progress and development of camouflage image generation technology.

ACKNOWLEDGMENT

National Natural Science Foundation of China (No.62001133), Fund Project of Guangxi Key Laboratory of Wireless Broadband Communication and Signal Processing (GXKL06200114).

REFERENCES

[1] Fan D P, JI G P, SUN G, et al. Camouflaged Object Detection[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2020: 2774-2784.

[2] JI G P, ZHANG J, CAMPBELL D L, et al. Rethinking polyp segmentation from an out-of-distribution perspective[J]. Machine Intelligence Research, 2024, 21(4): 631-639.

[3] FAN D P, JI G P, CHENG M M, et al. Concealed Object Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(10): 6024-6042.

[4] LE T N, NGUYEN T V, NIE Z L, et al. Anabran network for camouflaged object segmentation[J]. Computer Vision and Image Understanding, 2019, 184(1): 45-56.

[5] LV Y Q, ZHANG J, DAI Y C, et al. Simultaneously Localize, Segment and Rank the Camouflaged Objects[C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2021: 11586-11596.

[6] WANG L J, LU H C, WANG Y F, et al. Learning to Detect Salient Objects with Image-Level Supervision[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI: IEEE, 2017: 3796-3805.

[7] YANG C, ZHANG L H, LU H C, et al. Saliency detection via graph-based manifold ranking[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR: IEEE, 2013: 3166-3173.

[8] WANG J Y, QI Y. Multi-task learning and joint refinement between camera localization and object detection[J]. Computational Visual Media, 2024, doi: 10.1007/s41095-022-0319-z.

[9] DI M J M, FACCILOLO G, MEINHARDT E. Poisson Image Editing[J]. Image Processing On Line, 2016, doi: 10.5201/ipol.2016.163.

[10] ROMBACH R, BLATTMANN A, LORENZ D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE, 2022: 10674-10685.

[11] CHU H K, HSU W H, MITRA N J, et al. Camouflage images[J]. ACM Transactions on Graph, 2010, 29(4): 1-8.

[12] HUANG X, BELONGIE S. Arbitrary style transfer in real-time with adaptive instance normalization[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 1510-1519.

[13] ZHANG Q, YIN G, NIE Y, et al. Deep camouflage images[C]//Proceedings of the AAAI Conference on Artificial Intelligence. New York, NY, USA: Association for the Advancement of Artificial Intelligence, 2020: 12845-12852.

[14] LI Y Y, ZHAI W, CAO Y, et al. Location-free camouflage generation network[J]. IEEE Transactions on Multimedia, 2022, 25(7): 5234-5247.

[15] ZHENG C X, SONG G X, CHAM T J, et al. Bridging global context interactions for high-fidelity image completion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE, 2022: 11502-11512.

[16] LUGMAYR A, DANELLJAN M, ROMERO A, et al. Repaint: Inpainting using denoising diffusion probabilistic models [C]// Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New Orleans, LA, USA: IEEE, 2022: 11451-11461.

[17] ZHAO P C, XU P, QIN P D, et al. LAKE-RED: Camouflaged images generation by latent background knowledge retrieval-augmented diffusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE, 2024: 4092-4101.

[18] YANG L, ZHANG Z, SONG Y, et al. Diffusion models: A comprehensive survey of methods and applications[J]. ACM Computing Surveys, 2023, 56(4): 1-39.

[19] DHARIWAL P, NICHOL A. Diffusion models beat gans on image synthesis[EB/OL]. (2021-05-11) [2024-10-23]. <https://arxiv.org/abs/2105.05233>.

[20] BINKOWSKI M, SUTHERLAND D J, ARBEL M, et al. Demystifying mmd gans[EB/OL]. (2018-01-04)[2024-10-23]. <https://arxiv.org/abs/1801.01401>

[21] HEUSEL M, RAMSAUER H, Unterthiner T, et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium[EB/OL]. (2017-06-266)[2024-10-23]. <https://arxiv.org/abs/1706.08500>

Adoption of Generative AI-Enhanced Profit Sharing Digital Systems in MSMEs: A Comprehensive Model Analysis

Mardiana Andarwati^{1*}, Galandaru Swalaganata², Sari Yuniarti³,

Fandi Y. Pamuji⁴, Edward R. Sitompul⁵, Kukuh Yudhistiro⁶, Puput Dani Prasetyo Adi⁷

Faculty of Information Technology, Universitas Merdeka Malang, Malang City, East Java, Indonesia^{1, 2, 4, 5, 6}

Faculty of Economics and Business, Universitas Merdeka Malang, Malang City, East Java, Indonesia³

National Research and Innovation Agency (BRIN), Bandung Indonesia⁷

Abstract—Adopting digital finance solutions is crucial for enhancing efficiency and competitiveness within the financial services industry, particularly for Micro, Small, and Medium Enterprises (MSMEs). This study examines the factors influencing the use and acceptance of a sharing-based digital system enhanced with a Generative AI website (E-Mudharabah), employing the Technology Acceptance Model (TAM) and the Unified Theory of Acceptance and Use of Technology (UTAUT). In this article, the Generative AI-enhanced profit-sharing digital systems called E-Mudharabah. It is a web-based management system facilitating capital management for financiers, consultants, and MSME actors. The research integrates key variables from both models, including Perceived Ease of Use, Perceived Usefulness, Performance Expectancy, Social Influence, Facilitating Conditions, Habit, and Technology Self-Efficacy, to assess their impact on Behavioral Intention and Actual Usage. The study utilizes a quantitative approach, gathering data through surveys and analyzing it using the Partial Least Squares Structural Equation Modeling (PLS-SEM) method. Results indicate significant positive effects of perceived usefulness, performance expectancy, and social influence on the behavioral intention to use E-Mudharabah. The findings underscore the role of user-friendly interfaces and societal acceptance in driving adoption. Perceived Usefulness was the most significant variable influencing Behavioral Intention and Actual Usage (p -value < 0.001). Additionally, Social Influence and Facilitating Conditions were shown to have substantial effects, highlighting the importance of user support and societal acceptance in technology adoption. The research also underscores the role of Technology Self-Efficacy in enhancing user confidence and engagement with the platform. These findings suggest that improving digital finance solutions' perceived benefits and ease of use while fostering a supportive environment can significantly boost their adoption rates.

Keywords—Digital finance; Generative AI; TAM; UTAUT; MSMEs

I. INTRODUCTION

As the spread of information technology is becoming more widespread, especially in the financial services industry, some parties need to adopt new information technologies that can improve their cost structure, and efficiency and improve their competitive position [1]. Increasing income and maintaining

productivity for every MSME business actor is important in distributing funds or capital [2], [3].

Digitalization is the application of steps that arise from innovation in an organization [4]. Digital adaptation is defined as the ability to utilize digital technology through the use of digital tools and online platforms that increase competitiveness [5]. An organization or company should transform and be able to adapt to innovate and compete in an almost digitized world [6].

Digital finance has shown tremendous potential in reaching previously underserved and underserved populations by offering tailored financial services and products [7]. Online banking is becoming one of the more efficient ways and adopting online banking will have a positive impact on bank performance in the future [8]. Flexibility, informality, and control styles support the development of strategies that enable MSMEs to face environmental demands based on innovative and sustainable solutions [9].

E-Mudharabah is a digitalization solution designed as a website-based management information system to make it easier for several parties, including financiers, consultants, and MSME actors to manage and regulate capital schemes [3]. Implementing such digital solutions is expected to enhance the efficiency and effectiveness of financial management within MSMEs.

According to Venkatesh et al. [10], adding UTAUT to the TAM model means adding variables and validation, which of course will also increase the findings to be tested from models that are stated in other contexts [11]. This allows for more complex predictions about technology adoption [12], especially adopting new or innovative technologies [13]. The findings of Venkatesh and Bala [14] by adding elements of UTAUT to the TAM make social and organizational factors increasingly important because there are elements related to the social, psychological, and organizational environment. In addition, according to Venkatesh et al. [10], Zhou et al. [13], the combination of TAM and UTAUT is used for models with adaptive conditions to technological changes and user behavior that changes at any time.

Despite extensive studies on TAM and UTAUT, few have explored their integration in the context of MSMEs adopting AI-enhanced financial tools like E-Mudharabah. This study seeks to bridge this gap by combining these models to offer deeper insights into user adoption behaviors, addressing limitations in previous frameworks that overlooked context-specific variables such as technology self-efficacy and facilitating conditions. These models function as techniques to estimate the likelihood of a population adopting remote technology by incorporating relevant additional variables [15]. This research was conducted to measure the factors that affect the acceptance and utilization of Mudharabah, which has been digitized into a website using a combined Technology Acceptance Model and Unified Theory of Acceptance and Use Technology.

The remainder of this paper is organized as follows: Section II presents the literature review on TAM and UTAUT models and their applications. Section III describes the research methodology, including data collection and analysis methods. Section IV discusses the results, highlighting the key findings. Finally, Section V concludes with implications, limitations, and avenues for future research.

II. LITERATURE REVIEW

Previous studies predominantly examined TAM or UTAUT in isolation, failing to capture their synergistic potential in addressing multifaceted adoption challenges. This study contributes by integrating these models, adding constructs such as Technology Self-Efficacy, and contextualizing the analysis within MSMEs, which are pivotal to economic growth yet underrepresented in such research.

A. Technology Acceptance Model (TAM)

The TAM model is a model used to identify the acceptance and use of a new technology and information system [16]. TAM was developed to improve understanding related to the user onboarding process such as providing new theoretical insights into the design and implementation of information systems successfully designed to evaluate new systems before their implementation [17].

Perceived Ease of Use (PEoU) is interpreted as a value that measures the extent to which a person's confidence in using the system will be free from physical and mental effort [17], in other words, a person does not accept difficulties, but the ease received by using a system [18].

Davis defines perceived usefulness (PU) as the level of trust a person has in using a system that improves an individual's job performance [17] and gets other benefits such as improving his or her job performance [18].

The research of Pitafi & Ali [19] provides an overview that actual use (AU) can be assessed according to the quality of a better system.

B. Unified Theory of Acceptance and Use of Technology (UTAUT)

UTAUT is a tool that can be used to assess the success rate of the introduction of new technology and help understand the

beneficial factors of a user population from internal or external that are the drivers of acceptance so that users can accept adopting new technology and using the system [20]. Ding et al. [21] interpret the use of the UTAUT model as an integrative model whose use is aimed at predicting the availability of an individual to adopt new technologies.

Performance Expectancy is defined as the individual level at which a person is confident that using a system or technology can help improve job performance [20] and can increase a person's efficiency or output [22].

Facilitating Condition is an individual factor that believes that an organization supports the use of the system through adequate infrastructure and technology [20].

Social influence is defined as a direct influence on behavior or can be referred to as the level to which a person feels that another important person can make him believe that he must use a new system [20]. According to Singh et al [22], Social effects or factors are user influences obtained through other people related to the use of the system.

In a study conducted by Venkatesh et al., [10], Habit is an additional construct variable in UTAUT which is described as various user habits that have an impact and influence on the use of a technology.

Behavior Intention is the intention and desire of an individual [23] that influences to use of a technology [24] Hidalgo-Crespo & Amaya-Rivas [25] explain Behavior intention as an effort to encourage a person based on certain habits.

C. Technology Self-Efficacy

Self-efficacy is an assessment of oneself regarding beliefs about individual abilities [26]. Self-Efficacy is an important construct that measures a person's confidence in the ability to display a particular behavior [27], through endurance and perseverance to overcome difficulties, the anxiety faced, and the level of success achieved afterward [28]. Technology Self-Efficacy in Saville & Foster research [29] is defined as a measurement of a person related to the level of confidence in the successful use of a technology.

D. Comparison of Acceptance Models

This sub-section discusses the comparison of case studies with various acceptance models over the last five years.

III. RESEARCH METHODOLOGY

A. Research Model and Hypothesis

The research model shown in Fig. 1 is a path research model, which is used to determine the relationship between variables [36]. In this study, a combination of TAM and UTAUT models was used to determine the factors that affect the use and acceptance of E-Mudharabah. The selection of TAM and UTAUT is based on a literature study from previous studies that recommend a model merger experiment as shown in Table I. Some of the variables that were not used in this study were based on a literature study on the results of previous studies where these variables were considered insignificant.

TABLE I. COMPARISON OF ACCEPTANCE MODELS IN THE LAST FIVE YEARS

Author and Research Model	Variable Construct	R ²	Insignificant Variable
[30] Liu et al., 2022, TAM	Behavior Intention	77.2%	Perceived Ease Of Use
[31] Förster, 2024, TAM	Use Behavior	0.6%	Behavior Intention
[32] Uzir et al., 2023, Modified TAM	Behavior Intention	68.3%	Perceived Financial Cost
[33] Altes et al., 2024, Modified TAM	Behavior Intention	-	Perceived Ease Of Use, Perceived Cost, Voluntariness, Experience
[34] Mukred et al., 2024, Modified TAM	Behavior Intention	37.1%	Perceived Ease Of Use
[35] Chen et al., 2024, UTAUT	Behavior Intention	82.5%	Effort Expectancy
[36] Yee et al., 2024, UTAUT	Behavior Intention	-	Social Influence
[37] Sultana et al., 2023, Modified UTAUT	Behavior Intention	-	Social Factor, Personal Innovativeness
[38] Bellet & Banet, 2023, Modified UTAUT	Intention To Use	89.4%	Anxiety, Price Value, Satisfaction
[39] Han et al., 2024, Modified UTAUT	Behavior Intention	17.4%	Facilitating Condition, Social Influence, Perceived Negative Outcomes, Trust
[40] Rejali et al., 2024, TAM-UTAUT	Behavior Intention	76.1%	Green Perceived Usefulness
[41] Edo et al., 2023, TAM-UTAUT	Behavior Intention	28.3%	Perceived Ease Of Use, Social Influence,
[42] Bajunaied et al., 2023, Modified TAM - UTAUT	Behavior Intention	49.3%	Social Influence, Privacy Inhibitors

The TAM variables used in this study were Perceived Ease of Use, Perceived Usefulness, and Actual Use. The UTAUT variables used were Performance Expectancy, Social Influence, Facilitating Conditions, Habit, and Behavioral Intention. The researcher also added that the Technology Self-Efficacy variable is an important construct that measures a person's confidence in the ability to display certain behaviors [27], through endurance and perseverance to overcome difficulties, anxiety faced, and the level of success achieved afterward [28].

In Fig. 1, the hypothesis that arises as a result of the model built is also explained. The following is a description of the hypothesis based on Fig. 1.

H1: Perceived Ease of Use (PEOU) factor has a positive effect on Behavior Intention (BI) of E-Mudharabah applications.

H2: Perceived Usefulness (PU) factor has a positive effect on Behavior Intention (BI) of E-Mudharabah applications.

H3: Performance Expectancy (PE) factor has a positive effect on Behavior Intention (BI) of E-Mudharabah applications.

H4: Social Influence (SI) factor has a positive effect on Behavior Intention (BI) of E-Mudharabah applications.

H5: Facilitating Conditions (FC) factor has a positive effect on Behavior Intention (BI) of E-Mudharabah applications.

H6: Habit (HB) factor has a positive effect on Behavior Intention (BI) of E-Mudharabah applications.

H7: Technology Self-Efficacy (TSE) factor has a positive effect on Behavior Intention (BI) of E-Mudharabah applications.

H8: Behavior Intention (BI) factor has a positive effect on the Behavior Intention (BI) of E-Mudharabah applications.

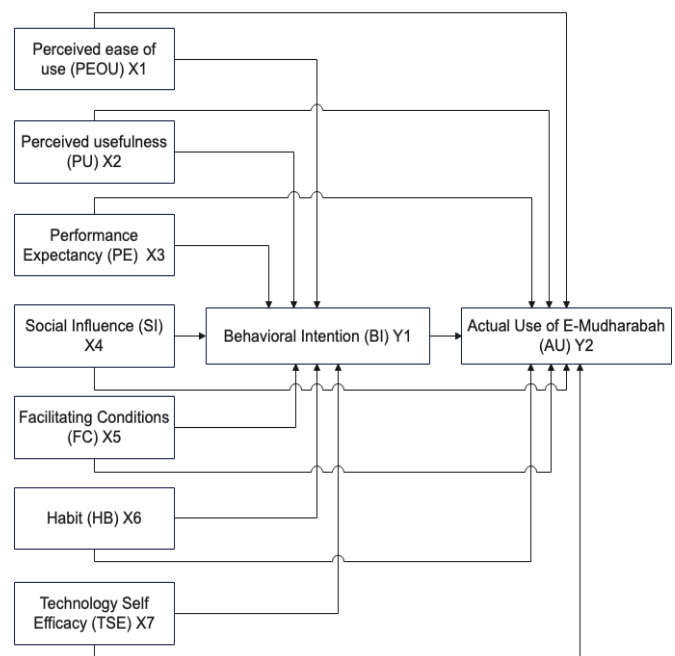


Fig. 1. Research model.

H9: Perceived Ease of Use (PEOU) factor has a positive effect on Actual Use (AU) of E-Mudharabah applications.

H10: Perceived Usefulness (PU) factor has a positive effect on Actual Use (AU) of E-Mudharabah applications.

H11: Performance Expectancy (PE) factor has a positive effect on the Actual Use (AU) of E-Mudharabah applications.

H12: Social Influence (SI) factor has a positive effect on the Actual Use (AU) of E-Mudharabah applications.

H13: Facilitating Conditions (FC) factor has a positive effect on the Actual Use (AU) of E-Mudharabah applications.

H14: Habit (HB) factor has a positive effect on the Actual Use (AU) of E-Mudharabah applications.

H15: Technology Self Efficacy (TSE) factor has a positive effect on the Actual Use (AU) of E-Mudharabah applications.

H16: Perceived Ease of Use (PEOU) factor has a positive effect on Actual Use (AU) of E-Mudharabah applications through Behavior Intention (BI)

H17: Perceived Usefulness (PU) factor has a positive effect on Actual Use (AU) of E-Mudharabah applications through Behavior Intention (BI).

H18: Performance Expectancy (PE) factor has a positive effect on Actual Use (AU) of E-Mudharabah applications through Behavior Intention (BI).

H19: Social Influence (SI) factor has a positive effect on Actual Use (AU) of E-Mudharabah applications through Behavior Intention (BI).

H20: Facilitating Conditions (FC) factor has a positive effect on Actual Use (AU) of E-Mudharabah applications through Behavior Intention (BI).

H21: Habit (HB) factor has a positive effect on Actual Use (AU) of E-Mudharabah applications through Behavior Intention (BI).

H22: Technology Self Efficacy (TSE) factor has a positive effect on the Actual Use (AU) of E-Mudharabah applications through Behavior Intention (BI).

B. Data Measurement

This study uses the Likert scale, which is a data measurement technique obtained through a survey to measure individual attitudes and opinions with five options of analytical responses, namely strongly disagree, disagree, undecided, agree, and strongly agree [43]. In this study, the variables are measured based on indicators as shown in Table II. The list of indicators was obtained and processed from literature studies as shown in Table I.

C. Data Collection

The target population is one of the MSMEs in the district in East Java which totals 93. They are a group of food and beverage entrepreneurs. The sample calculation technique uses the Slovin formula with a total sample obtained using the Slovin formula which is 75.

$$v = N / (1 + N\epsilon^2) \tag{1}$$

$$v = 93 / (1 + 93(0.05)^2)$$

$$v = 93 / 1.2325$$

$$v = 75$$

The sample of MSMEs that meet the criteria by filling out the entire questionnaire is 72 MSMEs, then the sample that does not provide a complete response will be eliminated [41].

D. Data Analysis

In this step, to test the variables and the relationships between the variables, the researcher uses Structural Equation Model (SEM) analysis. Briefly about SEM analysis is a validation test, reliability test, regression test, and hypothesis test. In this study, the statistical process uses the SMART PLS 4 application.

TABLE II. VARIABLES AND INDICATORS

No	Variable	Indicator	Code
1	Perceived Ease of Use (PEOU)	Easy to use	PEOU1
		Fast learning	PEOU2
		Clear interaction	PEOU3
		Interaction understood	PEOU4
		It doesn't require much effort.	PEOU5
2	Perceived Usefulness (PU)	Increase productivity	PU1
		Making work more efficient	PU2
		Useful for work	PU3
		Improve performance	PU4
3	Performance Expectancy (PE)	Increasing performance expectations	PE1
		Reach your goals faster.	PE2
		Increase job effectiveness	PE3
		Increase productivity	PE4
4	Social Influence (SI)	MSME managers encourage the use of	SI1
		Organizational influence	SI2
		Business owner support	SI3
5	Facilitating Condition (FC)	Resources available	FC1
		Enough knowledge	FC2
		Compatible with other technologies	FC3
		Help is always available.	FC4
6	Habit (HB)	Become a habit	HB1
		Used daily	HB2
		Automatic usage	HB3
		Routine habits	HB4
7	Technology Self-Efficacy (TSE)	Confident in ability	TSE1
		Can get work done without assistance	TSE2
		Can used with little information	TSE3
		Get the job done with confidence	TSE4
8	Behavioral Intention to Use (BI)	Intended use in employment	BI1
		Recommend to others	BI2
		Plan to continue using	BI3
		Interested in exploring features	BI4
9	Actual Use (AU)	Always use the system	AU1
		Uses most of the features	AU2
		Using in the job	AU3
		Relying on the system for tasks	AU4

IV. RESULT AND DISCUSSION

A. Data Demographics

Based on sub-section III.C, the total number is 75 respondents, but after a review of filling out the questionnaire, there are only 72 respondents who are complete and can be further analyzed.

Table III explains the age of the respondents where the average number of respondents is those aged 21 to 35 years. The last education is the most at the higher education level. This shows the form of mentoring and graduates from higher education in the Regency X area of East Java choose to do entrepreneurship in the food and beverage sector.

TABLE III. DEMOGRAPHICS CONDITION

	Samples (N=72)	%
Gender		
Male	27	37.5
Female	45	62.5
Age		
<20	12	16.7
21 – 35	34	47.2
>35	26	36.1
Level of Education		
Junior/Senior High School	11	15.3
Diploma	23	31.9
Bachelor	31	43.1
Other	7	9.7

B. Measurement Model Analysis (Outer Model)

Measurement analysis is carried out by paying attention to the validity and reliability values obtained through convergent validity and discriminant validity, and the reality values obtained through the reliability of constructs and indicators [23].

Convergent validity is considered to meet satisfactory criteria if the measurement items have high values in their respective constructs [40]. The Valid Criterion if the loading factor is greater than equal to 0.7 (≥ 0.7), and the measurement value of AVE (Average Variance Extracted) is greater than equal to 0.50 (≥ 0.5) [29]. Table IV shows the validity of all variables and their indicators, while Fig. 2 shows the outer loading of each variable.

The constructive/latent variable is stated to meet the convergent validity assumption if the AVE value is greater and higher than 0.5 [23]. So all latten or construct variables in this study shown in Table IV meet the Convergent Validity criteria.

The validity of Discrimination is attributed to the ability of measurement variables to distinguish between the objects being measured [18]. Validity discriminants are defined as diagonal relationships between variables [44]. The validity of discrimination is associated with the ability of measurement items to distinguish between the objects being measured [18].

This study uses the Fornell – Lacker Criterion and Cross Loading methods to determine the discriminatory validity value of all research variables with the results of the calculation of the Fornell – Lacker Criterion in Table IX and the results of Cross Loading in Table X. Both the table is located at the end of the article.

Convergent validity is considered to meet satisfactory criteria if the measurement items have high values in their respective constructs [40]. The Valid Criterion if the loading factor is greater than equal to 0.7 (≥ 0.7), and the measurement value of AVE (Average Variance Extracted) is greater than equal to 0.50 (≥ 0.5) [29]. Table IV shows the validity of all variables and their indicators, while Fig. 2 shows the outer loading of each variable.

The reliability of the research variables can be determined using composite variables [5], with reliable data criteria if the composite reliability value is greater than 0.70 (>0.70) and the Cronbach's alpha value is greater than 0.70 (>0.70) [29].

Cronbach's Alpha is used to evaluate the consistency of internal constraints [31], and satisfactory internal consistency if the value of Cronbach's Alpha and the Composite Reliability value of each variable exceed the value of 0.7 [39].

Thus, based on Table V with the measurement of data for each research variable, it shows that nine research variables meet the criteria of reliability with a Cronbach's Alpha value and a Composite Reliability value greater than the value of 0.7.

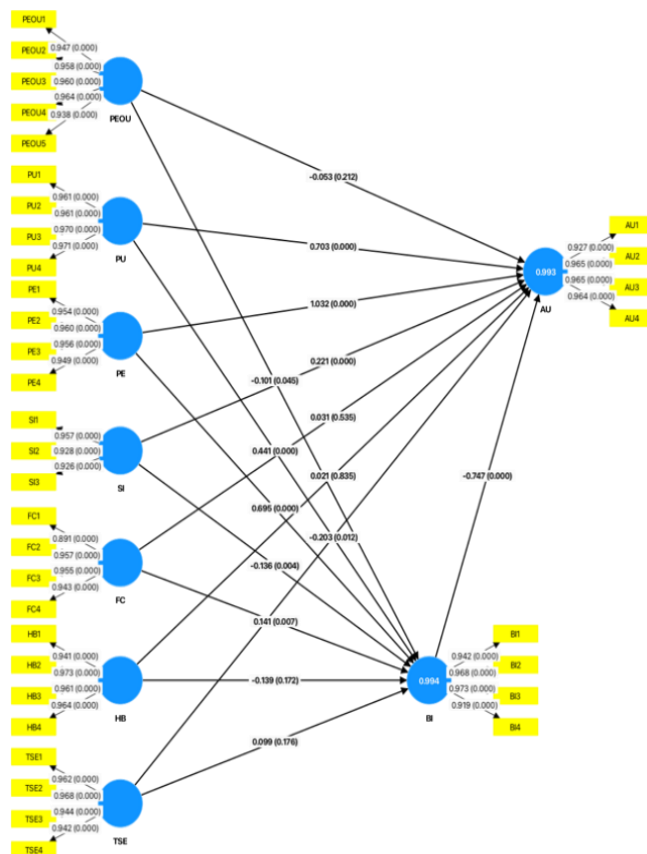


Fig. 2. Output line diagram.

TABLE IV. CONVERGENT VALIDITY

Variable	Indicator	Loading Factor	Valid/No	AVE
AU	AU1	0.927	Valid	0.913
	AU2	0.965	Valid	
	AU3	0.965	Valid	
	AU4	0.964	Valid	
BI	BI1	0.942	Valid	0.903
	BI2	0.968	Valid	
	BI3	0.973	Valid	
	BI4	0.919	Valid	
FC	FC1	0.891	Valid	0.877
	FC2	0.957	Valid	
	FC3	0.955	Valid	
	FC4	0.943	Valid	
HB	HB1	0.941	Valid	0.912
	HB2	0.973	Valid	
	HB3	0.961	Valid	
	HB4	0.964	Valid	
PE	PE1	0.954	Valid	0.912
	PE2	0.960	Valid	
	PE3	0.956	Valid	
	PE4	0.949	Valid	
PEOU	PEOU1	0.947	Valid	0.909
	PEOU2	0.958	Valid	
	PEOU3	0.960	Valid	
	PEOU4	0.964	Valid	
	PEOU5	0.938	Valid	
PU	PU1	0.961	Valid	0.933
	PU2	0.961	Valid	
	PU3	0.970	Valid	
	PU4	0.971	Valid	
SI	SI1	0.957	Valid	0.878
	SI2	0.928	Valid	
	SI3	0.926	Valid	
TSE	TSE1	0.962	Valid	0.910
	TSE2	0.968	Valid	
	TSE3	0.944	Valid	
	TSE3	0.942	Valid	

PeoU: Perceived Ease of Use; PU: Perceived Usefulness; PE: Performance Expectancy; SI: Social Influence; FC: Facilitating Conditions; HB: Habit; TSE: Technology Self Efficacy; BI: Behavior Intention; AU: Actual Use of E-Mudharabah

TABLE V. CRONBACH'S ALPHA AND COMPOSITE RELIABILITY

Variable	Cronbach's Alpha	Composite Reliability (rho_a)	Composite Reliability (rho_c)
AU	0.968	0.969	0.977
BI	0.964	0.964	0.975

FC	0.953	0.954	0.966
HB	0.971	0.972	0.979
PE	0.968	0.968	0.976
PEOU	0.975	0.975	0.980
PU	0.976	0.976	0.982
SI	0.931	0.934	0.956
TSE	0.967	0.968	0.976

PeoU: Perceived Ease of Use; PU: Perceived Usefulness; PE: Performance Expectancy; SI: Social Influence; FC: Facilitating Conditions; HB: Habit; TSE: Technology Self Efficacy; BI: Behavior Intention; AU: Actual Use of E-Mudharabah

C. Structural Model Analysis (Inner Model)

According to Hidalgo-Crespo & Amaya-Rivas in the study [25], Structural is used to describe the path coefficient, show the relationship of each research variable to the constructed variable, and determine the significant statistical value.

TABLE VI. R-SQUARE VALUE

Variable	R-Square Adjustment	Result
AU	0.993	Strong
BI	0.994	Strong

AU: Actual Use of E-Mudharabah; BI: Behavior Intention

The value of the R-Square Determinant is used to indicate the magnitude of the strength [23] and the magnitude of the influence of independent variables on the dependent variables which are divided into three namely [45]:

- 1) A value of determinant more than 0.67 (>0.67) is a Strong category.
- 2) Moderate category if the R-Square determinant value is between 0.33 – 0.67.
- 3) The category is weak if the R-Square determinant value is between 0.19 – 0.33.

Based on the results of the R-Square calculation in Table VI, the AU and BI construct variables are included in the strong category with values of 0.993 (99.3 %) and 0.994 (99.4 %).

This means that the AU dependent variable is influenced by the independent variable as a whole as much as 99.3% and the remaining 0.7% is influenced by other variables that were not tested in the study. Likewise, the BI dependent variable was influenced by the independent variable as a whole as much as 99.4% and the remaining 0.6% was influenced by other variables that were not tested in the study.

F-square or effect size is a measurement that assesses the relative impact between independent variables on dependent variables which are divided into several category classifications, namely strong categories with an f-square value of more than 0.35, medium categories with an f-square value of more than 0.15, and weak categories with an f-square value of less than 0.02 [45].

Based on the results from Table VII, several factors significantly influence both Behavior Intention and Actual Use of the E-Mudharabah system. Performance Expectancy (PE) emerged as the most influential factor on behavioral intention,

with an F-Square value of 2.689, indicating that users have high expectations that the E-Mudharabah platform will enhance their performance. Users believe that this system will help them achieve their goals more efficiently and improve their overall work effectiveness.

Perceived Usefulness (PU) also plays a crucial role in shaping users' behavioral intentions, with an F-Square value of 0.819. This suggests that users view the E-Mudharabah system as a valuable tool in their daily operations, which in turn, increases their likelihood of continued use. The perception that the system significantly benefits their work encourages users to integrate it into their routines.

TABLE VII. F-SQUARE VALUE

Path	F-Square	Result
PEOU → BI	0.120	Medium
PU → BI	0.819	Strong
PE → BI	2.689	Strong
SI → BI	0.173	Strong
FC → BI	0.163	Strong
HB → BI	0.050	Medium
TSE → BI	0.053	Medium
BI → AU	0.515	Strong
PEOU → AU	0.027	Medium
PU → AU	1.054	Strong
PE → AU	1.485	Strong
SI → AU	0.360	Strong
FC → AU	0.006	Weak
HB → AU	0.001	Weak
TSE → AU	0.195	Strong

PeoU: Perceived Ease of Use; PU: Perceived Usefulness; PE: Performance Expectancy; SI: Social Influence; FC: Facilitating Conditions; HB: Habit; TSE: Technology Self Efficacy; BI: Behavior Intention; AU: Actual Use of E-Mudharabah

Additionally, Perceived Ease of Use (PEOU) affects both Behavior Intention and Actual Use, with F-Square values of 0.120 and 0.027, respectively. This indicates that the easier users find the system to use, the more likely they are to adopt and sustain its use. Simplifying the user interface and ensuring that the system is accessible and user-friendly can significantly enhance user adoption.

Other factors such as Facilitating Conditions and Social Influence also significantly impact behavioral intention and actual usage. Social Influence, with an F-Square value of 0.173, underscores the importance of support from colleagues or superiors in encouraging the use of the system. Facilitating Conditions, which scored an F-Square value of 0.163 for Behavior Intention, suggests that the availability of resources and adequate technical support also contribute to users' willingness to use the system consistently.

From the results of this f-square test, it can be concluded that the pathway with the strongest influence on Behavior Intention (BI) is Performance Expectancy (PE) with the largest f-square value. Meanwhile, the pathways with the strongest influence on Actual Use (AU) are Performance Expectancy (PE) and Perceived Usefulness (PU). The pathways with the weakest influence on Actual Use (AU) are Facilitating Condition (FC) and Habit (HB). This shows that factors such as perceived usability and performance expectations are highly

influential in determining actual intention and use while supporting conditions and habits have a lower influence.

The hypothesis testing in this study uses the bootstrap technique with a significance value of 5% (0.05). Hypothesis acceptance is determined in P-Values [4], with the P-Value criterion being less than 0.05, so it is identified as a significant variable relationship to the latent/construct variable [33].

D. Discussion

After calculating the hypothesis in Table VIII, there are 22 research hypotheses with a total of 15 hypotheses that are accepted by influencing latent variables and seven research hypotheses that are rejected.

Based on the hypothesis testing results, several key insights emerge about the factors influencing MSMEs' usage of the Modified E-Mudharabah website (Micro, Small, and Medium Enterprises). Perceived Ease of Use (PEOU) plays a significant role in shaping Behavioral Intention (BI), suggesting that when MSME users find the E-Mudharabah system easy to use, they are more likely to intend to use it. However, PEOU does not directly affect Actual Use (AU); its influence on AU is mediated through BI. This highlights the importance of designing user-friendly interfaces to enhance user intentions, which translates into actual usage. For MSMEs, simplifying the user experience is crucial as it can reduce the time and effort required for them to adapt to new technology, allowing them to focus more on their core business activities.

TABLE VIII. HYPOTHESIS TESTING RESULT

	Hypothesis Path	T-Val	P-Val	Result
H1	PEOU → BI	1.949	0.026	Accept
H2	PU → BI	6.279	0.000	Accept
H3	PE → BI	9.691	0.000	Accept
H4	SI → BI	3.035	0.001	Accept
H5	FC → BI	2.596	0.005	Accept
H6	HB → BI	1.349	0.089	Reject
H7	TSE → BI	1.391	0.082	Reject
H8	BI → AU	4.645	0.000	Accept
H9	PEOU → AU	1.292	0.098	Reject
H10	PU → AU	7.735	0.000	Accept
H11	PE → AU	7.519	0.000	Accept
H12	SI → AU	3.650	0.000	Accept
H13	FC → AU	0.654	0.257	Reject
H14	HB → AU	0.219	0.413	Reject
H15	TSE → AU	2.522	0.006	Accept
H16	PEOU → BI → AU	1.757	0.040	Accept
H17	PU → BI → AU	4.424	0.000	Accept
H18	PE → BI → AU	4.453	0.000	Accept
H19	SI → BI → AU	2.537	0.006	Accept
H20	FC → BI → AU	1.994	0.023	Accept
H21	HB → BI → AU	1.438	0.075	Reject
H22	TSE → BI → AU	1.388	0.083	Reject

PeoU: Perceived Ease of Use; PU: Perceived Usefulness; PE: Performance Expectancy; SI: Social Influence; FC: Facilitating Conditions; HB: Habit; TSE: Technology Self Efficacy; BI: Behavior Intention; AU: Actual Use of E-Mudharabah

Perceived Usefulness (PU) is a crucial determinant, directly impacting both BI and AU. MSME users are more inclined to use the E-Mudharabah system if they perceive it as useful, confirming that practical benefits and functional advantages are strong motivators for adoption. Additionally, PU's indirect influence through BI underscores its comprehensive effect on user behavior. This finding emphasizes the need for continuous

improvements and updates that enhance the system's utility, ensuring it meets the specific needs and expectations of MSMEs effectively. By demonstrating tangible benefits such as increased efficiency, better financial management, and access to broader markets, the E-Mudharabah system can become indispensable for MSMEs.

TABLE IX. DISCRIMINANT VALIDITY FORNNEI -LACKER

	AU	BI	FC	HB	PE	PEOU	PU	SI	TSE
AU	0.955								
BI	0.975	0.950							
FC	0.915	0.918	0.937						
HB	0.980	0.982	0.907	0.960					
PE	0.986	0.987	0.909	0.976	0.955				
PEOU	0.885	0.863	0.957	0.863	0.876	0.954			
PU	0.970	0.983	0.916	0.984	0.959	0.860	0.966		
SI	0.974	0.942	0.922	0.950	0.961	0.909	0.933	0.937	
TSE	0.960	0.980	0.908	0.981	0.970	0.858	0.974	0.937	0.954

PEOU: Perceived Ease of Use; PU: Perceived Usefulness; PE: Performance Expectancy; SI: Social Influence; FC: Facilitating Conditions; HB: Habit; TSE: Technology Self Efficacy; BI: Behavior Intention; AU: Actual Use of E-Mudharabah

TABLE X. CROSS LOADING VALUE

	AU	BI	FC	HB	PE	PEOU	PU	SI	TSE
AU1	0.927	0.937	0.884	0.941	0.901	0.812	0.961	0.894	0.919
AU2	0.965	0.912	0.872	0.919	0.956	0.859	0.888	0.957	0.893
AU3	0.965	0.919	0.864	0.912	0.949	0.855	0.891	0.942	0.891
AU4	0.964	0.959	0.876	0.973	0.960	0.855	0.970	0.930	0.968
BI1	0.904	0.942	0.891	0.923	0.888	0.793	0.961	0.868	0.914
BI2	0.913	0.968	0.862	0.947	0.954	0.806	0.937	0.878	0.962
BI3	0.923	0.973	0.874	0.950	0.960	0.827	0.947	0.892	0.959
BI4	0.965	0.919	0.864	0.912	0.949	0.855	0.891	0.942	0.891
FC1	0.904	0.942	0.891	0.923	0.888	0.793	0.961	0.868	0.914
FC2	0.826	0.816	0.957	0.802	0.824	0.947	0.807	0.848	0.817
FC3	0.848	0.827	0.955	0.824	0.844	0.958	0.822	0.867	0.833
FC4	0.837	0.841	0.943	0.833	0.840	0.894	0.825	0.865	0.823
HB1	0.927	0.937	0.884	0.941	0.901	0.812	0.961	0.894	0.919
HB2	0.964	0.959	0.876	0.973	0.960	0.855	0.970	0.930	0.968
HB3	0.943	0.931	0.848	0.961	0.948	0.809	0.918	0.921	0.921
HB4	0.927	0.943	0.873	0.964	0.937	0.836	0.929	0.903	0.956
PE1	0.913	0.968	0.862	0.947	0.954	0.806	0.937	0.878	0.962
PE2	0.923	0.973	0.874	0.950	0.960	0.827	0.947	0.892	0.959
PE3	0.965	0.912	0.872	0.919	0.956	0.859	0.888	0.957	0.893
PE4	0.965	0.919	0.864	0.912	0.949	0.855	0.891	0.942	0.891
PEOU1	0.826	0.816	0.957	0.802	0.824	0.947	0.807	0.848	0.817
PEOU2	0.848	0.827	0.955	0.824	0.844	0.958	0.822	0.867	0.833
PEOU3	0.841	0.818	0.882	0.821	0.827	0.960	0.817	0.855	0.802
PEOU4	0.852	0.819	0.889	0.833	0.838	0.964	0.819	0.879	0.811
PEOU5	0.851	0.837	0.884	0.835	0.843	0.938	0.834	0.883	0.826
PU1	0.904	0.942	0.891	0.923	0.888	0.793	0.961	0.868	0.914
PU2	0.927	0.937	0.884	0.941	0.901	0.812	0.961	0.894	0.919
PU3	0.964	0.959	0.876	0.973	0.960	0.855	0.970	0.930	0.968
PU4	0.953	0.958	0.888	0.964	0.953	0.859	0.971	0.912	0.960
SI1	0.965	0.912	0.872	0.919	0.956	0.859	0.888	0.957	0.893
SI2	0.911	0.906	0.869	0.912	0.891	0.820	0.915	0.928	0.915
SI3	0.858	0.827	0.852	0.836	0.848	0.878	0.817	0.926	0.824
TSE1	0.913	0.968	0.862	0.947	0.954	0.806	0.937	0.878	0.962
TSE2	0.964	0.959	0.876	0.973	0.960	0.855	0.970	0.930	0.968
TSE3	0.894	0.907	0.877	0.904	0.888	0.818	0.916	0.892	0.944
TSE4	0.892	0.905	0.849	0.918	0.897	0.793	0.892	0.876	0.942

PEOU: Perceived Ease of Use; PU: Perceived Usefulness; PE: Performance Expectancy; SI: Social Influence; FC: Facilitating Conditions; HB: Habit; TSE: Technology Self Efficacy; BI: Behavior Intention; AU: Actual Use of E-Mudharabah

Performance Expectancy (PE) and Social Influence (SI) also emerge as significant factors in the context of MSMEs using the E-Mudharabah system. PE, which reflects users' belief that using the system will help them achieve desired business outcomes, significantly affects both BI and AU. Similarly, SI, the influence of peers and social networks, plays a vital role in shaping MSME user attitudes and behaviors towards the system. These findings suggest that promoting the system's effectiveness through success stories and leveraging social networks for endorsement can significantly boost user engagement and acceptance among MSMEs. Encouraging word-of-mouth and positive reviews from other MSMEs can create a supportive community that facilitates wider adoption.

Facilitating Conditions (FC), which refer to the availability of resources and support for using the system, significantly influence BI and indirectly affect AU through BI for MSMEs. This implies that providing adequate support, such as training programs, technical assistance, and user manuals, is essential for fostering user intentions, which in turn drives actual usage. However, FC does not directly impact AU, indicating that while supportive conditions are necessary, they alone are not sufficient to ensure usage without the mediation of BI. For MSMEs, ensuring that they have the necessary infrastructure and support to integrate the E-Mudharabah system into their operations is crucial for its successful adoption.

Interestingly, Habit (HB) and Technology Self-Efficacy (TSE) show distinct patterns of influence among MSME users. HB, or the extent to which users perform behaviors automatically due to learning, does not significantly impact BI or AU. This suggests that habitual behavior may not be a strong predictor in this context, and efforts should focus more on enhancing users' perceptions of ease and usefulness. On the other hand, TSE, which reflects users' confidence in their ability to use the system, directly influences AU but not BI. This indicates that MSME users' self-efficacy in using the technology is crucial for actual usage, even if it doesn't directly shape their intentions. Providing training and resources to boost the confidence of MSME users in their ability to effectively use the E-Mudharabah system can lead to higher actual usage.

Finally, Behavioral Intention (BI) itself is a significant predictor of AU. This confirms the pivotal role of BI in the adoption process, indicating that strategies aimed at enhancing BI—through improving PEOU, PU, PE, SI, and providing adequate FC—are likely to increase actual usage of the E-Mudharabah system among MSMEs. By understanding and addressing these factors, developers, and policymakers can optimize the system to better meet the needs of MSMEs, promoting widespread adoption and helping these enterprises to thrive in the digital economy.

Behavioral Intention (BI) serves as a critical mediator between several factors (Perceived Ease of Use, Perceived Usefulness, Performance Expectancy, Social Influence, Facilitating Conditions) and Actual Use (AU). This mediation highlights that while these factors are essential in influencing users' intentions, the actual usage of the e-Mudharabah system is predominantly driven by the intention to use it. This finding underscores the importance of enhancing users' intentions to

use the system as a pathway to achieving higher actual usage rates. By focusing on improving factors that drive behavioral intention, developers can indirectly increase the actual adoption and use of the system.

Perceived Usefulness (PU) has a strong direct impact on both Behavioral Intention (BI) and Actual Use (AU). Additionally, its indirect impact through BI further enhances its overall influence on AU. Users' perception of the system's usefulness is a pivotal factor in both their intention to use and their actual usage of the system. When users believe that the e-Mudharabah system will significantly benefit their work and improve their performance, they are more likely to adopt and utilize it.

Social Influence (SI) significantly affects both Behavioral Intention (BI) and Actual Use (AU), indicating that the opinions and behaviors of peers and social networks play a crucial role in technology adoption. This result highlights the power of social validation in driving the adoption of new technologies. Users are more likely to use the e-Mudharabah system if they see their peers and social networks endorsing and using it.

Habit (HB) and Technology Self-Efficacy (TSE) do not significantly impact Behavioral Intention (BI), although TSE does have a direct effect on Actual Use (AU). This suggests that habitual behavior and confidence in using the technology are not primary drivers of intention or actual use in this context. These findings indicate that simply relying on users' habitual behavior or their confidence in using technology may not be sufficient to drive the adoption and usage of the e-Mudharabah system. Other factors, such as perceived ease of use, usefulness, and social influence, play more critical roles.

In this study, the intention and desire of MSMEs to use the E-Mudharabah information system is not influenced by the habit and level of trust in a technology, but is positively influenced by the convenience of the E-Mudharabah information system, the benefits received by using the E-Mudharabah information system, the usefulness of the E-Mudharabah information system, the adequate conditions to use an information system and the influence of someone who has used an information system E-Mudharabah.

In contrast to the research by Putri et al [23] which discusses the acceptance of financial technology, it shows that the ease of adoption of technology does not affect a person's desire to use the information technology. Table IV explains that the convenience, usefulness and benefits, social influence, and condition of facilities in MSME places can arouse a person's intention to use and operate the E-Mudharabah information system, while individual habits and abilities cannot be an influence to make the intention to use the E-Mudharabah information system.

V. CONCLUSION

The study underscores the substantial impact of various factors on the intention of MSME (Micro, Small, and Medium Enterprises) actors to utilize the E-Mudharabah information system. It reveals that the intention to adopt this technology is significantly influenced by ease of use, perceived usefulness, and the tangible benefits offered by the system. Social

influence—peer pressure or encouragement from the social environment—and the state of MSME facilities also play a pivotal role. These factors collectively account for an impressive 99.4% of the variance in the intention to adopt the E-Mudharabah information system, indicating their overwhelming importance. However, the study also notes that habitual familiarity with the system and the technical ability of individuals to use it do not significantly drive their intention to adopt the technology.

Moreover, when moving from intention to actual adoption and utilization, the system demonstrates a similarly strong positive impact. The transition from intention to real-life usage is influenced by the benefits perceived by MSME actors, social support, and their readiness, with a high explanatory value of 99.3%. This finding emphasizes that while intention is critical, the practicality and value of the system in addressing specific business needs also significantly drive adoption.

However, the research acknowledges a limitation in its current methodology, specifically the lack of discriminant validity. This issue implies that while the findings are robust, they may not fully capture the nuanced distinctions among variables or populations. To enhance the reliability and applicability of future studies, researchers should expand the target sample size and diversify the population of MSMEs under investigation. This approach will help refine the measurement tools and ensure the data better represents the broader MSME landscape.

While the study demonstrates the significant impact of perceived usefulness and performance expectancy, it does not account for longitudinal adoption trends. Future studies should adopt longitudinal designs to evaluate sustained usage and investigate additional factors such as cultural influences and economic conditions.

ACKNOWLEDGMENT

This research was supported by the Ministry of Education, Culture, Research and Technology of Indonesia. We thank our colleagues from Universitas Merdeka Malang, Indonesia who provided insight and expertise that greatly assisted the research.


REFERENCES

- [1] K. Bauer and S. E. Hein, "The effect of heterogeneous risk on the early adoption of Internet banking technologies," *J Bank Financ*, vol. 30, no. 6, pp. 1713–1725, 2006, doi: 10.1016/j.jbankfin.2005.09.004.
- [2] M. Andarwati, G. Swalaganata, F. Y. Pamuji, and N. D. Hendrawan, "Rancang Bangun Sistem Informasi Manajemen e-Mudharabah Berbasis Website," in *Seminar Nasional Sistem Informasi (SENASIF)*, 2023, pp. 3776–3787.
- [3] M. Andarwati, G. Swalaganata, F. Y. Pamuji, and N. D. Hendrawan, "Application Of The RAD (Rapid Application Development) Method To Develop A Website-Based E- Mudharabah Savings And Loans System," vol. 10, pp. 1–9, Mar. 2024, doi: 10.9790/0050-10060109.
- [4] D. Karpova and A. Proskurina, "The Need for Sociotechnical Turn in the Study of Society Digitalization," vol. 3, p. 71, 2021.
- [5] U. W. Nuryanto, Basrowi, I. Quraysin, and I. Pratiwi, "Magnitude of digital adaptability role: Stakeholder engagement and costless signaling in enhancing sustainable MSME performance," *Heliyon*, vol. 10, no. 13, pp. e33484–e33484, 2024, doi: 10.1016/j.heliyon.2024.e33484.

- [6] Y. Mou, "The impact of digital finance on technological innovation across enterprise life cycles in China," *Heliyon*, vol. 10, no. 14, pp. e33965–e33965, 2024, doi: 10.1016/j.heliyon.2024.e33965.
- [7] J. Yang, Y. Wu, and B. Huang, "Digital finance and financial literacy: Evidence from Chinese households," *J Bank Financ*, vol. 156, no. September, 2023, doi: 10.1016/j.jbankfin.2023.107005.
- [8] R. Hernández-Murillo, G. Llobet, and R. Fuentes, "Strategic online banking adoption," *J Bank Financ*, vol. 34, no. 7, pp. 1650–1663, 2010, doi: 10.1016/j.jbankfin.2010.03.011.
- [9] A. Escobar-Castillo *et al.*, "Factors that impact the innovation capability in MSMEs: Case of Colombia's Atlantico Department," *Procedia Comput Sci*, vol. 224, no. 2021, pp. 490–494, 2023, doi: 10.1016/j.procs.2023.09.070.
- [10] V. Venkatesh, J. Y. L. Thong, and X. Xu, "Consumer Acceptance And Use Of Information Technology: Extending The Unified Theory Of Acceptance And Use Of Technology," *MIS Quarterly*, vol. 36, no. 1, pp. 157–178, 2012.
- [11] M. J. Mortenson and R. Vidgen, "A computational literature review of the technology acceptance model," *Int J Inf Manage*, vol. 36, no. 6, pp. 1248–1259, 2016.
- [12] R. P. Bagozzi, "The legacy of the technology acceptance model and a proposal for a paradigm shift," *J Assoc Inf Syst*, vol. 8, no. 4, p. 3, 2007.
- [13] T. Zhou, Y. Lu, and B. Wang, "Integrating TTF and UTAUT to explain mobile banking user adoption," *Comput Human Behav*, vol. 26, no. 4, pp. 760–767, 2010.
- [14] V. Venkatesh and H. Bala, "Technology acceptance model 3 and a research agenda on interventions," *Decision sciences*, vol. 39, no. 2, pp. 273–315, 2008.
- [15] M. Rouidi, A. E. Elouadi, A. Hamdoune, K. Choujtani, and A. Chati, "TAM-UTAUT and the acceptance of remote healthcare technologies by healthcare professionals: A systematic review," *Inform Med Unlocked*, vol. 32, no. March, p. 101008, 2022, doi: 10.1016/j.imu.2022.101008.
- [16] T. J. Habibie, R. Yasirandi, and D. Oktaria, "The analysis of Pangandaran fisherman's actual usage level of GPS based on TAM model," *Procedia Comput Sci*, vol. 197, no. 2021, pp. 34–41, 2021, doi: 10.1016/j.procs.2021.12.115.
- [17] F. D. Davis, "A technology acceptance model for empirically testing new end-user information systems: Theory and results," *Management*, vol. Ph.D., no. January 1985, p. 291, 1985, doi: oclc/56932490.
- [18] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Q*, vol. 13, no. 3, pp. 319–339, 1989, doi: 10.2307/249008.
- [19] A. H. Pitafi and A. Ali, "An empirical investigation on actual usage of educational app: Based on quality dimensions and mobile self-efficacy," *Heliyon*, vol. 9, no. 9, pp. e19284–e19284, 2023, doi: 10.1016/j.heliyon.2023.e19284.
- [20] V. Venkatesh, M. G. Moris, G. B. Davis, and F. D. Davis, "User Acceptance Of Information Technology: Toward A Unified View," *Inorg Chem Commun*, vol. 67, no. 3, pp. 95–98, 2003, doi: 10.1016/j.inoche.2016.03.015.
- [21] Y. Ding, R. Guo, M. Bilal, and V. G. Duffy, "Exploring the influence of anthropomorphic appearance on usage intention on online medical service robots (OMSRs): A neurophysiological study," *Heliyon*, vol. 10, no. 5, pp. e26582–e26582, 2024, doi: 10.1016/j.heliyon.2024.e26582.
- [22] S. Singh, V. Kumar, M. Paliwal, S. V. P. Singh, and S. Mahlawat, "Explaining the linkage between antecedents' factors of adopting online classes and perceived learning outcome using extended UTAUT model," *Data Inf Manag*, vol. 7, no. 4, 2023, doi: 10.1016/j.dim.2023.100052.
- [23] G. A. Putri, A. K. Widagdo, and D. Setiawan, "Analysis of financial technology acceptance of peer to peer lending (P2P lending) using extended technology acceptance model (TAM)," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 9, no. 1, p. 100027, 2023, doi: 10.1016/j.joitmc.2023.100027.
- [24] I. Y. Alyoussef, "Acceptance of a flipped classroom to improve university students' learning: An empirical study on the TAM model and the unified theory of acceptance and use of technology (UTAUT)," *Heliyon*, vol. 8, no. 12, p. e12529, 2022, doi: 10.1016/j.heliyon.2022.e12529.

- [25] J. Hidalgo-Crespo and J. L. Amaya-Rivas, "Citizens' pro-environmental behaviors for waste reduction using an extended theory of planned behavior in Guayas province," *Clean Eng Technol*, vol. 21, no. June, p. 100765, 2024, doi: 10.1016/j.clet.2024.100765.
- [26] E. T. Straub, "Understanding Technology Adoption: Theory and Future Directions for Informal Learning," *Rev Educ Res*, vol. 79, no. 2, pp. 625–649, 2009, [Online]. Available: Theory and Future Directions for Informal Learning.pdf
- [27] D. R. Compeau and C. A. Higgins, "Computer Self-Efficacy: Measure And Initial Development Of A Test," *MIS Quarterly*, vol. 19, no. 2, pp. 189–211, 2017, [Online]. Available: <https://www.astm.org/Standards/E2368.htm>
- [28] A. L. Zeldin and F. Pajares, "Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers," *Am Educ Res J*, vol. 37, no. 1, pp. 215–246, 2000, doi: 10.3102/00028312037001215.
- [29] J. D. Saville and L. L. Foster, "Does technology self-efficacy influence the effect of training presentation mode on training self-efficacy?," *Computers in Human Behavior Reports*, vol. 4, p. 100124, 2021, doi: 10.1016/j.chbr.2021.100124.
- [30] Y. Liu, J. Henseler, and Y. Liu, "What makes tourists adopt smart hospitality? An inquiry beyond the technology acceptance model," *Digital Business*, vol. 2, no. 2, 2022, doi: 10.1016/j.digbus.2022.100042.
- [31] K. Förster, "Extending the technology acceptance model and empirically testing the conceptualised consumer goods acceptance model," *Heliyon*, vol. 10, no. 6, 2024, doi: 10.1016/j.heliyon.2024.e27823.
- [32] M. U. H. Uzir *et al.*, "Applied artificial intelligence: Acceptance-intention-purchase and satisfaction on smartwatch usage in a Ghanaian context," *Heliyon*, vol. 9, no. 8, pp. e18666–e18666, 2023, doi: 10.1016/j.heliyon.2023.e18666.
- [33] G. C. Altes, A. K. S. Ong, and J. D. German, "Determining factors affecting Filipino consumers' behavioral intention to use cloud storage services: An extended technology acceptance model integrating valence framework," *Heliyon*, vol. 10, no. 4, pp. e26447–e26447, 2024, doi: 10.1016/j.heliyon.2024.e26447.
- [34] M. Mukred, U. A. Mokhtar, B. Hawash, H. AlSalman, and M. Zohaib, "The adoption and use of learning analytics tools to improve decision making in higher learning institutions: An extension of technology acceptance model," *Heliyon*, vol. 10, no. 4, pp. e26315–e26315, 2024, doi: 10.1016/j.heliyon.2024.e26315.
- [35] Y. Chen, S. K. Khan, N. Shiwakoti, P. Stasinopoulos, and K. Aghabayk, "Integrating perceived safety and socio-demographic factors in UTAUT model to explore Australians' intention to use fully automated vehicles," *Research in Transportation Business and Management*, vol. 56, no. November 2023, p. 101147, 2024, doi: 10.1016/j.rtbm.2024.101147.
- [36] L. C. Yee, C. Kwok Yip, C. C. Seng, and L. K. Kei, "Integrating the adapted UTAUT model with moral obligation, trust and perceived risk to predict ChatGPT adoption for assessment support: A survey with students," *Computers and Education: Artificial Intelligence*, vol. 6, no. May, p. 100246, 2024, doi: 10.1016/j.caeai.2024.100246.
- [37] N. Sultana, R. S. Chowdhury, and A. Haque, "Gravitating towards Fintech: A study on Undergraduates using extended UTAUT model," *Heliyon*, vol. 9, no. 10, pp. e20731–e20731, 2023, doi: 10.1016/j.heliyon.2023.e20731.
- [38] T. Bellet and A. Banet, "UTAUT4-AV: An extension of the UTAUT model to study intention to use automated shuttles and the societal acceptance of different types of automated vehicles," *Transp Res Part F Traffic Psychol Behav*, vol. 99, no. April, pp. 239–261, 2023, doi: 10.1016/j.trf.2023.10.007.
- [39] J. Han, T. Welch, U. Voß, T. Vernoux, R. Bhosale, and A. Bishopp, "Factors influencing nurses' acceptance of patient safety reporting systems based on the unified theory of acceptance and use of technology (UTAUT)," *iScience*, p. 109936, 2024, doi: 10.1016/j.imu.2024.101554.
- [40] S. Rejali, K. Aghabayk, A. Mohammadi, and N. Shiwakoti, "Evaluating public a priori acceptance of autonomous modular transit using an extended unified theory of acceptance and use of technology model," *J Public Trans*, vol. 26, no. January, p. 100081, 2024, doi: 10.1016/j.jpuptr.2024.100081.
- [41] O. C. Edo, D. Ang, E. E. Etu, I. Tenebe, S. Edo, and O. A. Diekola, "Why do healthcare workers adopt digital health technologies - A cross-sectional study integrating the TAM and UTAUT model in a developing economy," *International Journal of Information Management Data Insights*, vol. 3, no. 2, p. 100186, 2023, doi: 10.1016/j.ijime.2023.100186.
- [42] K. Bajunaied, N. Hussin, and S. Kamarudin, "Behavioral intention to adopt FinTech services: An extension of unified theory of acceptance and use of technology," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 9, no. 1, p. 100010, 2023, doi: 10.1016/j.joitmc.2023.100010.
- [43] K. A. Batterton and K. N. Hale, "The Likert Scale What It Is and How To Use It," *Phalanx*, vol. 50, no. 2, pp. 32–39, 2017.
- [44] L. D. F. Fornell C., "Evaluating structural equation models with unobservable variables and measurement error," *Journal of Marketing Research This*, vol. 18, no. 1, pp. 39–50, 2016.
- [45] R. F. Robby and T. Mauritsius, "Level of Student Satisfaction With New Binusmaya: Measuring and Analyzing Using the End User Computing Satisfaction (Eucs) Framework," *J Theor Appl Inf Technol*, vol. 101, no. 15, pp. 6144–6155, 2023.

DeepLabV3+ Based Mask R-CNN for Crack Detection and Segmentation in Concrete Structures

Yuewei Liu 

School of Civil Engineering and Architecture, Wenzhou Polytechnic, Zhejiang, 325000, China

Abstract—In order to solve the problem of concrete structure crack detection and segmentation and improve the efficiency of detection and segmentation, this paper proposes a crack detection and segmentation method for concrete structure based on DeepLabV3+ and Mask R-CNN algorithm. Firstly, a crack detection and segmentation scheme is designed by analysing the crack detection and segmentation problem of concrete structure. Secondly, a crack detection method based on Mask R-CNN algorithm is proposed for the crack detection problem of concrete structure. Then, a crack segmentation method based on DeepLabV3+ algorithm is proposed for the crack segmentation problem of concrete structure. Finally, bridge crack image data is used for the crack detection and segmentation of concrete structure. Finally, the concrete structure crack detection and segmentation method is validated and analysed using bridge crack image data. The results show that the Mask R-CNN model has better performance in the localisation and identification of cracks, and the DeepLabV3+ model has higher accuracy and contour extraction integrity in solving the crack segmentation problem.

Keywords—DeepLabV3+; Mask R-CNN; concrete structure; crack detection and segmentation; deep learning algorithm

I. INTRODUCTION

With the rapid development of the society, the state's investment in infrastructure such as roads, bridges and buildings is increasing [1]. Most of the above infrastructures are composed of concrete. And the concrete structure may produce cracks for various reasons, which has a great impact on its subsequent use as well as safety [2]. The causes of cracks in concrete facilities include the following: 1) thermal expansion and contraction of concrete produces a large number of cracks; 2) frequent vehicle trips put a great deal of pressure on the highway nucleus bridges, resulting in cracks; 3) too high or too low a standard of concrete ratios cause cracks in the structure; and 4) improper construction causes the concrete shrinkage nucleus to crack prematurely. If the concrete cracks are not repaired in time may lead to highway fracture, bridge nuclear construction facilities collapse, endangering traffic travel [3]. Traditional concrete crack detection and maintenance using manual methods, not only consume manpower and material resources, but also inaccurate detection results, high risk, easy to accident [3]. In order to facilitate the detection of concrete crack nuclei to reduce the risk of detection, artificial intelligence technology is used, not only to improve the traditional image processing technology problems, but also to detect the good results [4]. Concrete structure crack detection and segmentation research is mainly divided into concrete crack detection, crack segmentation and other issues research. The current concrete crack detection methods are divided into contact detection method, non-contact

detection method, and image crack detection method [5-7]. Contact detection method is generally manual detection method, is through visual inspection or measurement attack to detect and record the crack size and location, this method is simple, but has great limitations [5]. Non-contact detection methods are generally used to detect cracks with the help of existing infrared, radar and other equipment, the use of which not only does not have a bad effect on the concrete being detected, but also solves the problem of losses due to human error in judgment [6]. Image crack detection method mainly uses image segmentation technology to detect concrete cracks, which includes segmentation algorithms such as thresholding, region growing, edge detection, etc. This method has the advantages of high detection efficiency and low cost [7]. Xiao et al. [8] proposed the histogram bimodal method to segment the image. Xi et al. [9] proposed the OTUS algorithm to detect cracks, which not only can effectively detect small and irregular cracks, but also the detection accuracy reaches 85%. Li and Yang [10] used osmotic hair to detect the concrete cracks, and got a better result. Moezi et al. [11] combined the seepage theory with the adaptive Canny operator to improve the effect of the detection algorithm. Although the traditional data image detection method is low cost and simple to operate, it is only adapted to simple environments, and the detection segmentation results are not ideal. With the development of artificial intelligence technology, deep learning algorithms are introduced into the image crack detection segmentation problem. Cha et al. [12] used deep learning technology CNN to identify cracks on building surfaces, and achieved 98% accuracy. Zhang et al. [13] used Mask-RCNN combined with FPN feature pyramid network module kernel Resnet model to improve the extraction accuracy of crack disease features, and showed a better recognition accuracy. Li et al [14] for the problem of extracting cracks on concrete surfaces, and proposed an improved lightweight global convolutional network image segmentation model for pavement cracks. Kim and Cho [15] uses the PSPNet semantic segmentation platform for model detection of highway bridge cracks, which shows good detection results.

Although these methods have made significant progress in improving the accuracy and efficiency of crack detection, there are still some shortcomings [16]: (1) the segmentation models suffer from the presence of pooling layers, which leads to the loss of some positional information, restricting their ability to accurately identify fine cracks; (2) some of the structures sacrifice the feature resolution in successive pooling operations or convolutional steps, which makes the prediction task limited and affecting the performance of image segmentation; (3) various types of detection models still need to be further verified for their performance and stability in practical applications.

Concrete crack detection and segmentation is crucial to ensure the safety of concrete structures. For the current concrete crack detection and segmentation problems, this paper proposes Mask R-CNN [18] based on DeepLabV3+ [17] for concrete structure crack detection and segmentation, and the specific contributions of the paper are as follows: (1) analyse the concrete structure crack detection and segmentation problems, and put forward the research scheme; (2) around the concrete structure crack detection problems, put forward the concrete structure crack detection model based on the Mask R-CNN algorithm; (3) A crack segmentation model based on DeepLabV3+ algorithm is proposed for the concrete structure crack segmentation problem; (4) The proposed method is analysed and validated using concrete crack data, and the results show that the model method improves the detection efficiency and accuracy of concrete.

II. CRACK DETECTION AND SEGMENTATION IN CONCRETE STRUCTURES

A. Analysis of the Problem

As one of the common defects in concrete structures, cracks have a direct impact on the safety and durability of bridges. The detailed analysis of cracks in concrete bridges (Fig. 1) can provide a comprehensive understanding of the health status of bridge structures and provide a scientific basis for timely maintenance and repair of bridges [19].



Fig. 1. Schematic diagram of concrete bridge cracks

Concrete crack characteristics mainly include: 1) crack morphology and distribution, which mainly reflects the bridge structure force and deformation process; 2) crack size and shape, the size of which is directly related to the degree of damage to the structure, and its distribution pattern can reflect the structural uneven force, deformation inconsistency and so on, as shown in Fig. 2.

For the task of crack detection in concrete bridges, the concern is the morphology of the cracks. Currently, according to the morphology, concrete cracks are classified into linear cracks and web-like cracks (Fig. 3). Deep learning algorithm-based crack detection method for concrete bridges needs to be targeted at different scales, can effectively detect cracks of various sizes and shapes, and can cope with different lighting conditions and environmental changes, the specific analysis is shown in Fig. 4.

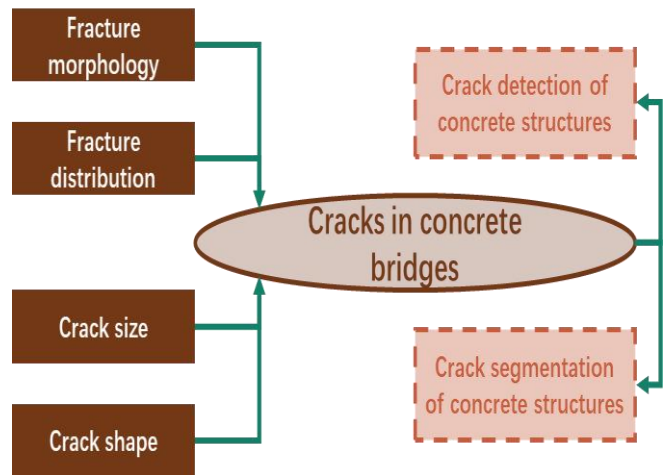


Fig. 2. Characteristics analysis of cracks in concrete bridges.



(a) Straight cracks.



(b) Reticulated cracks.

Fig. 3. Concrete bridge crack patterns.

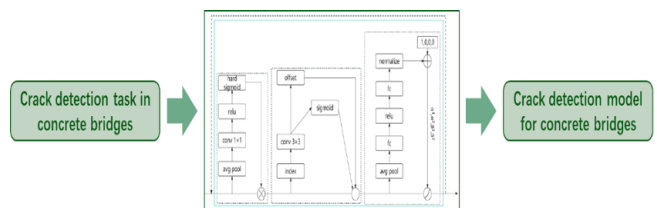


Fig. 4. Concrete bridge crack detection problem analysis.

In the task of concrete bridge crack detection, only crack target detection is not enough, and the introduction of image segmentation model based on deep learning is crucial. The concrete bridge crack segmentation study is not only able to accurately locate each crack, but also able to accurately segment the outline of the crack, which is analysed as shown in Fig. 5.

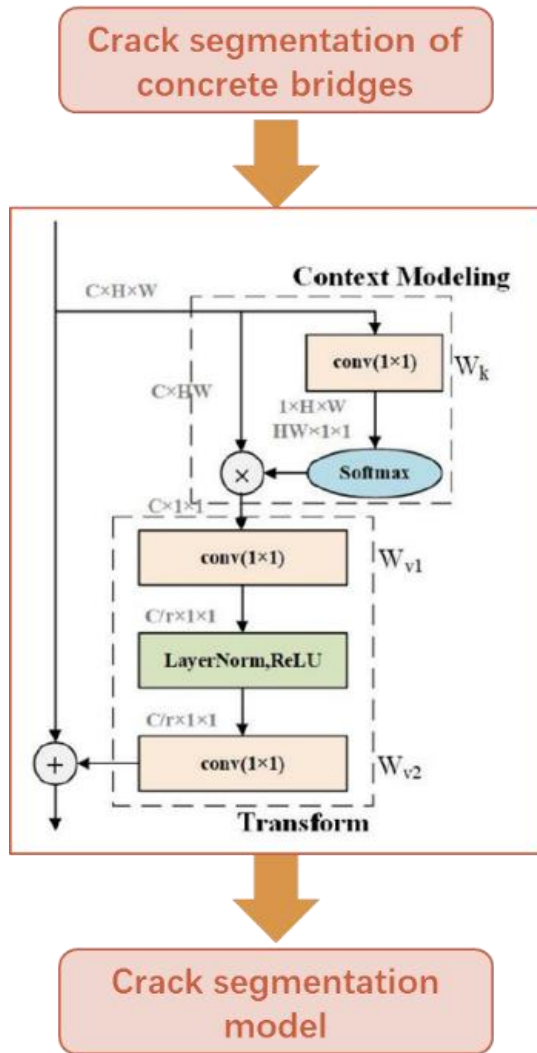


Fig. 5. Analysing crack segmentation problems in concrete bridges.

B. Design of Crack Detection and Segmentation Programmes for Concrete Structures

According to the analysis of concrete structure crack detection and segmentation problem, this paper adopts deep learning algorithm to construct concrete structure crack detection and segmentation model, the specific design scheme is shown in Fig. 6. Concrete structure crack detection and segmentation scheme design from two modules to study the concrete structure crack analysis problem, the first module is the concrete structure crack detection model construction, using deep learning algorithms (Mask R-CNN) to locate the overall location of the crack; the second module is the concrete structure crack segmentation model construction, using deep learning algorithms (DeepLabV3+) to segment the crack contour of the cracks.

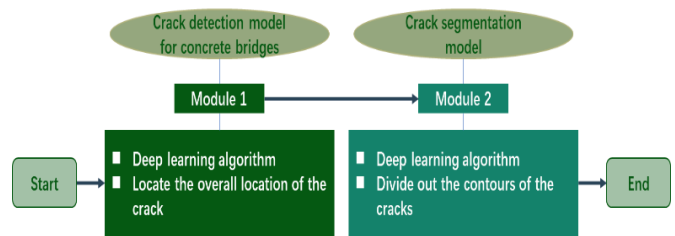


Fig. 6. Programme design.

III. CRACK AND SEGMENTATION OF CONCRETE STRUCTURES DETECTION

A. DeepLabV3+ Algorithm

The DeepLabV3+ algorithm is the latest version of the Deeplab series [20], and the specific structure is shown in Fig. 7. The DeepLabV3+ network model utilises a combination of the spatial pyramid pooling module and the decoder-encoder structure of the deep neural network to achieve fine segmentation of the target boundary. The spatial pyramid pooling module detects input features at multiple rates and multiple effective fields of view through filtering or pooling operations to encode multi-scale contextual information, and the network structure is shown in Fig. 7; the encoder-decoder structure captures clearer target boundaries by gradually reverting to spatial information, and the network diagram is shown in Fig. 8.

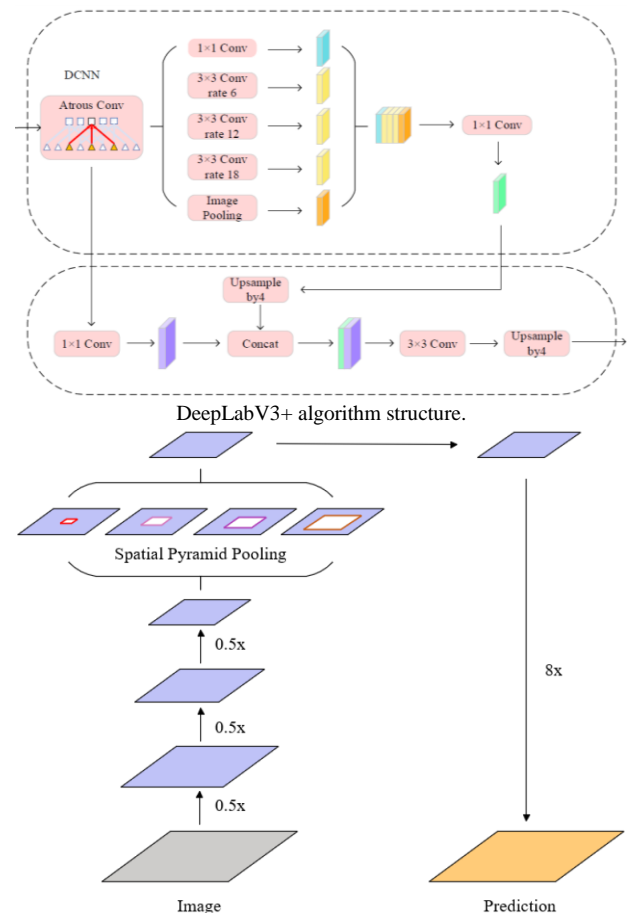


Fig. 7. Encoder-decoder structure network.

The DeepLabV3+ network model structure mainly consists of two parts, Encoder and Decoder. The encoder part of the DeepLabV3+ network uses the DeepLabV3 network as a whole for feature extraction of feature maps of arbitrary resolution output from the deep neural network. The encoder works as follows: first a Conv+BN+ReLU is used, then a convolution, plus global average pooling is used to obtain the scale features, and finally the features are upsampled using Conv+BN+ReLU+ bilinear interpolation to keep the feature maps of the same size. The decoder part is spliced using the outputs of the encoder and DCNN parts, and finally a convolution and upsampling is used for the output.

The DeepLabV3+ network model convolution layer effectively reduces the number of parameters in the model, reduces the risk of overfitting, and improves the generalisation of the model [21]. The convolution process (shown in Fig. 8) is as follows:

$$a_{i,j} = F \left(\sum_{m=0}^{M-1} \sum_{n=0}^{N-1} w_{m,n} x_{i+m,j+n} \right) \quad (1)$$

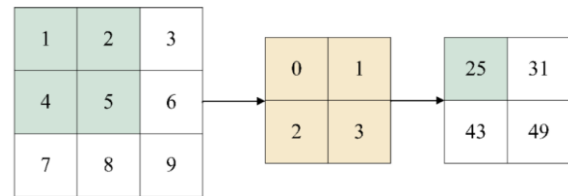


Fig. 8. Convolution process.

Where, $a_{i,j}$ is the feature map element; $x_{i,j}$ is the i -th row and j -th column element in the convolution feature map; $w_{m,n}$ is the m -th row and n -th column ground weight in the convolution; F is the activation function.

B. Mask R-CNN Algorithm

Mask R-CNN algorithm is a deep learning algorithm [22] for target detection and instance segmentation. Fig. 9 extends Faster-RCNN (Convolutional Neural Network based target detection algorithm) by adding a segmentation branch to predict the exact boundary and mask of the target.

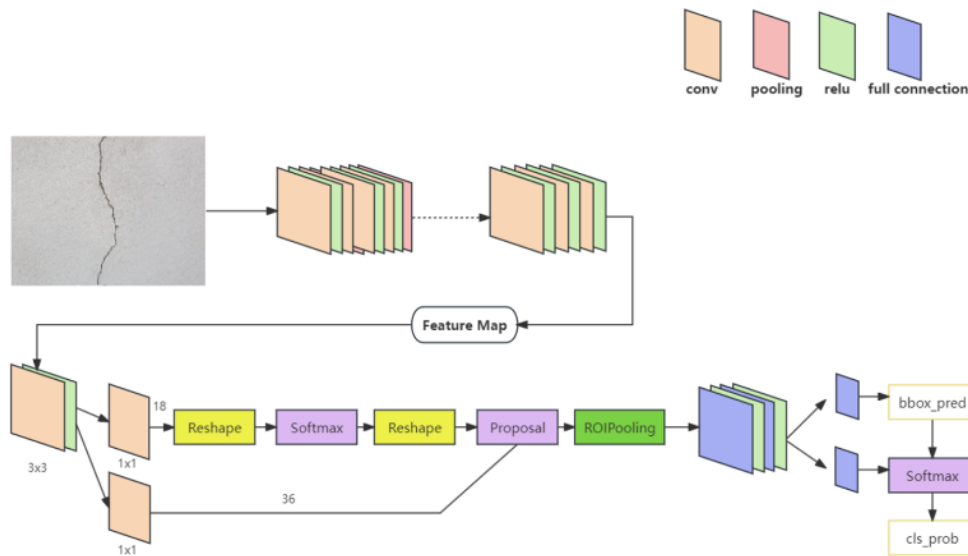


Fig. 9. Mask R-CNN algorithm.

The target detection process of Faster R-CNN is divided into several key steps (shown in Fig. 10):

- (Feature extraction by Convolutional Neural Network (CNN) to obtain high-level semantic information from the input image;
- Introducing a region proposal network (RPN) to generate a large number of candidate target regions, i.e., anchoring boxes, on the convolutional feature map;
- Non-maximal suppression (NMS) is used to eliminate highly overlapping redundant anchor frames, ensuring that each target region is represented by only one candidate frame;
- In the fully connected layer, the Faster R-CNN performs classification and bounding box regression of targets.

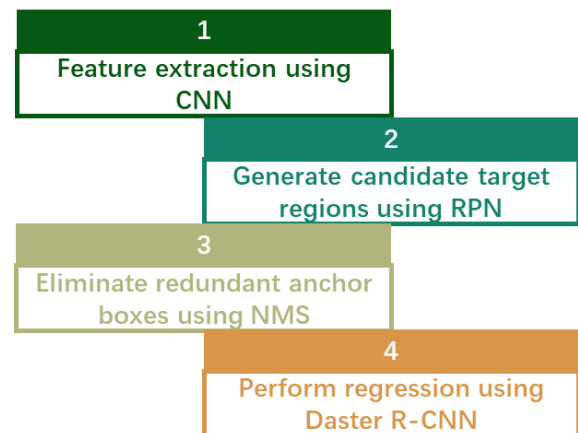


Fig. 10. Steps of Mask R-CNN algorithm.

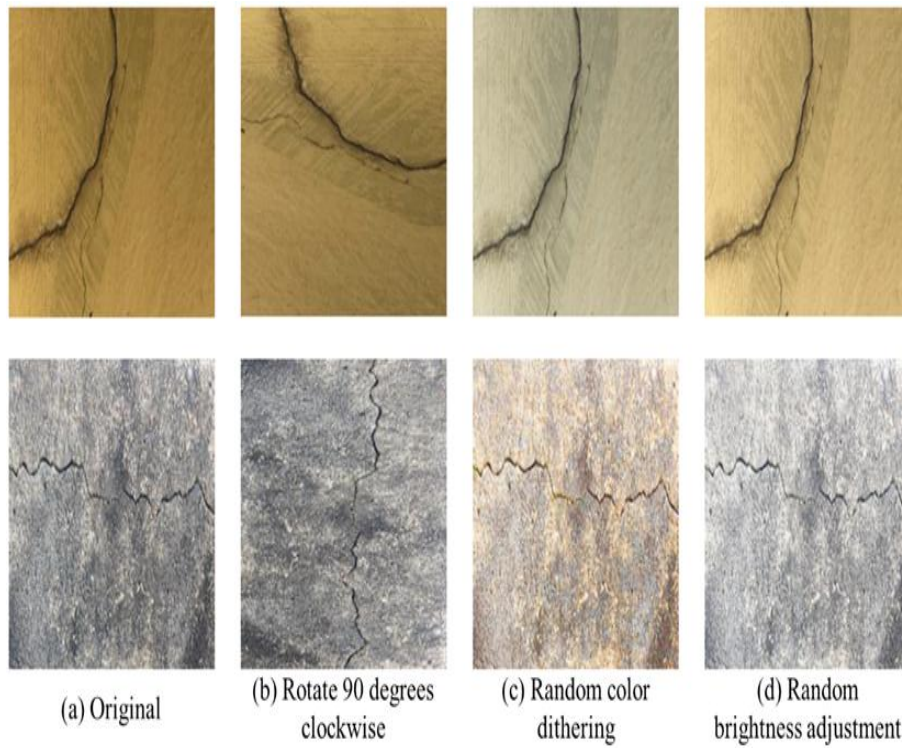


Fig. 13. Image enhancement processing diagram.

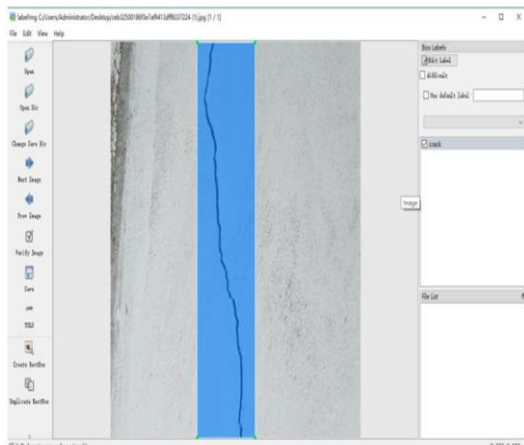


Fig. 14. Crack image annotation interface.

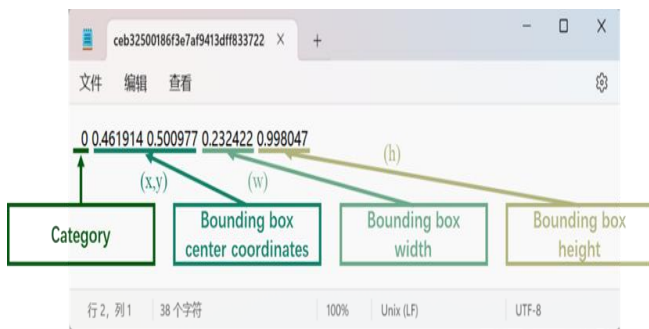


Fig. 15. Structure of crack image annotation labels.

B. Experimental Environment and Training Parameter Settings

In order to match the computational complexity of the algorithm, this paper configures the experimental environment as shown in Table I.

TABLE I. EXPERIMENTAL ENVIRONMENT SETTINGS

Environment Configuration	Parameterisation
CPU	AMD Ryzen7 5800H
RAM	16GB
GPU's	RTX 3060 Laptop
memory	6GB
fig. pattern	Pytorch

The parameter settings of the depth algorithm used in this paper are shown in Table II.

TABLE II. PARAMETER SETTINGS OF DEEP ALGORITHM

Parameter name	Parameterisation
Batch_size	2
epoch	156
optimisation algorithm	AdamW
learning rate	0.00285
Input image resolution	640 x 640

C. Evaluation Indicators

In order to effectively evaluate the effectiveness of the algorithm in this paper, the check accuracy rate (Precision), the check full rate (Recall) and the mean average precision (mean Average Precision, mAP) are used as the evaluation indexes, and the specific calculation formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{2}$$

$$R = \frac{TP}{TP + FN} \tag{3}$$

$$mAP = \frac{\sum_{i=1}^C AP_i}{C} \tag{4}$$

where P is the checking accuracy rate; R is the checking completeness rate; mAP is the Average Precision (AP) of the mean; TP denotes the true category, i.e., the model correctly predicts the positive category samples to be positive; FP denotes the pseudo-positive category, i.e., the model incorrectly predicts the negative category samples to be positive; FN denotes the pseudo-negative category, i.e., the model incorrectly predicts the positive category samples to be negative; AP_i denotes the Average Precision (AP) value for the i -th category; and C denotes the total number of categories.

D. Analysis of Experimental Results

1) Analysis of concrete crack detection results: In order to verify the effectiveness of the concrete crack detection algorithm proposed in this paper, SSD [26], Faster R-CNN [27], and YOLOv5 [28] are used in this paper to conduct comparative experiments with Mask R-CNN, and the results of the experiments are shown in Table III and Fig. 16.

As can be seen from Table III, the concrete crack detection model Mask R-CNN proposed in this paper has a high detection accuracy, with mAP reaching 94.9%, Precision reaching 90.5% and Recall reaching 89.5%.

TABLE III. EXPERIMENTAL RESULTS OF DIFFERENT DETECTION MODELS

Arithmetic	mAP	Precision	Recall
SSD	0.876	0.832	0.748
Faster R-CNN	0.933	0.885	0.855
YOLOv5	0.928	0.879	0.851
Mask R-CNN	0.949	0.905	0.895

From Fig. 16, it can be seen that SSD, Faster R-CNN and YOLOv5 models have duplicated detection frames and low detection completeness, and the Mask R-CNN model proposed in this paper has a better performance in the localisation and identification of cracks compared to the target detection model, especially in the crack detection completeness, which is better than the other models and can satisfy the daily crack detection

in concrete bridges. The Mask R-CNN model proposed in this paper.

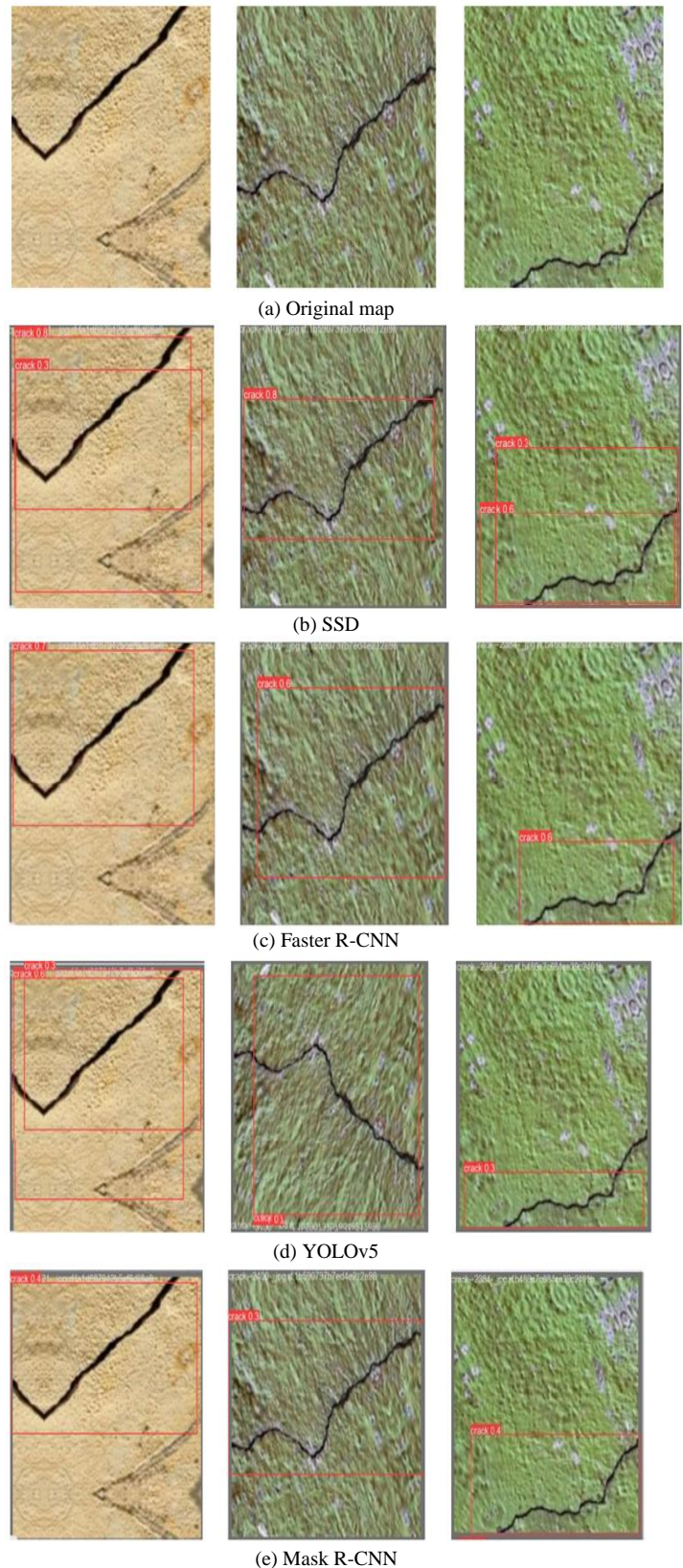


Fig. 16. Comparison of detection results.

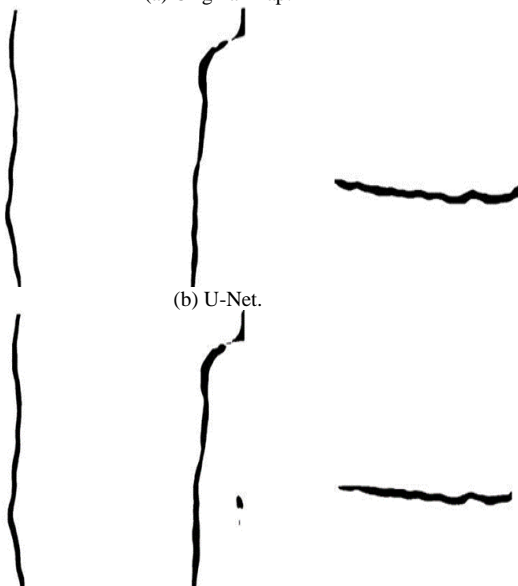
2) *Analysis of concrete crack segmentation results:* In order to verify the effectiveness of the bridge crack profile segmentation algorithm proposed in this paper, U-Net [29], Mask R-CNN, and YOLOv5-seg [30] are used in this paper to conduct comparative experiments with DeepLabV3+, and the results of the experiments are shown in Table IV and Fig. 17. From Table IV, it can be seen that the DeepLabV3+ model has the highest segmentation effect accuracy compared with other segmentation models, with mAP, Precision, and Recall reaching 48.9%, 67.4%, and 56.7%, respectively. From the segmentation effect diagram (Fig. 17), it can be seen that the bridge crack detection model proposed in this paper achieves better results in the evaluation indexes. In order to more intuitively show the effect performance of different models in bridge crack contour extraction, the same bridge crack image is segmented by using different models, and the results of contour extraction verify that the algorithms proposed in this paper have high accuracy and completeness of contour extraction.

TABLE IV. EXPERIMENTAL RESULTS OF DIFFERENT CRACK SEGMENTATION MODELS FOR CONCRETE STRUCTURES

Arithmetic	mAP	Precision	Recall
U-Net	0.466	0.616	0.503
Mask R-CNN	0.476	0.643	0.520
YOLOv5-seg	0.485	0.637	0.528
DeepLabV3+	0.489	0.674	0.567



(a) Original map.



(b) U-Net.

(c) Mask R-CNN.



Fig. 17. Comparison of segmentation results.

V. CONCLUSION AND OUTLOOK

Aiming at the current concrete structural crack detection and segmentation methods, which have problems such as low recognition ability and inaccurate extraction of the contour of cracks, this paper fuses DeepLabV3+ and Mask R-CNN algorithms, and proposes a concrete structural crack detection and segmentation method based on DeepLabV3+ and Mask R-CNN algorithms. A concrete structure crack detection and segmentation method based on DeepLabV3+ and Mask R-CNN model is proposed by analysing the problem of concrete structure crack detection and segmentation, designing a solution method, and combining DeepLabV3+ and Mask R-CNN algorithm. Using the bridge crack image data for validation and analysis, the Mask R-CNN model has a better performance in the localisation and identification of cracks compared to the target detection model, especially in the integrity of crack detection than other models; the bridge crack detection model achieves better results in the evaluation indexes, and has a high accuracy and integrity of contour extraction in the task of crack contour extraction for concrete bridges. The next step is to improve the training speed of DeepLabV3+ and Mask R-CNN algorithms, and further validate the effectiveness of the algorithms from crack detection problems in different fields.

The proposed method effectively solves key issues in crack detection, especially in locating fine cracks and achieving better segmentation. Despite the strong performance, we note two primary limitations: 1), The current implementation requires significant computational resources, and the training process is slow. 2), Although tested on bridge crack data, the method's performance in other types of concrete structures or environmental conditions has not yet been fully validated.

Prospects for future research directions through this study can be focused on the following aspects: Firstly, Future research could focus on optimizing the training process to make the models faster and more efficient without sacrificing accuracy. Secondly, Additional studies are needed to test the method on

more diverse datasets and environments to ensure the model's robustness across different types of concrete structures. Finally, investigating the use of lightweight neural networks could reduce the computational load and make real-time crack detection more feasible. These future directions aim to further enhance the method's practical applicability and broaden its use in various civil engineering contexts.

REFERENCES

- [1] Kumar C , Sinha A K .Automated Crack Detection and a Web Tool Using Image Processing Techniques in Concrete Structures[J]. Nondestructive Testing, 2023(11):59.
- [2] Choi Y , Park H W , Mi Y S S .Crack Detection and Analysis of Concrete Structures Based on Neural Network and Clustering[J].sensors, 2024, 24(6).
- [3] Li H , Zhang H , Zhu H ,Gao K, Liang H, Yang J. Automatic crack detection on concrete and asphalt surfaces using semantic segmentation network with hierarchical Transformer[J].Engineering Structures, 2024, 307.
- [4] Kapadia H K , Patel P V , Patel J B .Convolutional Neural Network Based Improved Crack Detection In Concrete Cubes[J]. Computing and Digital Systems, 2023.
- [5] Kim Y H .Defect Detection and Characterization in Concrete Based on FEM and Ultrasonic Techniques[J].Materials, 2022, 15.
- [6] Ghannadiasl A , Ghaemifard S .Crack detection of the cantilever beam using new triple hybrid algorithms based on Particle Swarm Optimization[J]. Frontiers in Structural and Civil Engineering:English Edition, 2022, 16(9):14.
- [7] Yoshinaka F , Nakamura T , Uesugi U K .Characterisation of internal fatigue crack initiation in Ti-6Al-4V alloy via synchrotron radiation X-ray computed tomography[J].Fatigue & Fracture of Engineering Materials and Structures, 2023, 46(6):2338 -2347.
- [8] Xiao X Y, Zhang X Y, Du X F. A review of automatic bridge crack detection methods based on image processing[J]. Electronic Testing,2019,(19):52-53+55.
- [9] Xi F, Liu H, Huang Z, Talab A, Mahgoub A .Detection crack in image using Otsu method and multiple filtering in image processing techniques[J]. for Light and Electronoptic, 2016, 127(3): 1030-1033.
- [10] Li Y X ,Yang Q S. Research on concrete surface crack detection based on U2net neural network[J]. Journal of Qinghai University,2024,42(03):77-85.
- [11] Moezi S A , Zakeri E , Zare A , Nedaei M .On the application of modified cuckoo optimisation algorithm to the crack detection problem of cantilever Euler -Bernoulli beam[J].Computers & Structures, 2015, 157(09):42-50.
- [12] Cha Y J, Choi W, Büyüköztürk O. Deep learning-based crack damage detection using convolutional neural networks[J]. Computer-Aided Civil and Infrastructure Engineering, 2017, 32(5): 361-378.
- [13] Zhang S X, Zhang H C, Li X Z, Hu J. Research on multi-objective identification of pavement crack damage based on machine vision[J]. Highway Traffic Science and Technology,2021,38(03):30-39.
- [14] Li G, Gao Z Y, Zhang X C, Zhao H X, Liu Z. Application of improved global convolutional network in pavement crack detection[J]. Advances in Lasers and Optoelectronics,2020,57(08):111-119.
- [15] Kim B, Cho S. Image-based concrete crack assessment using mask and region-based convolutional neural network[J]. Structural Control and Health Monitoring, 2019, 26(8): e2381.
- [16] Kirthiga R , Elavenil S .A survey on crack detection in concrete surface using image processing and machine learning[J].Journal of Building Pathology and Rehabilitation, 2024, 9(1).
- [17] Lu H .An Identification Method for Mixed Coal Vitrinite Components Based on An Improved DeepLabv3+ Network[J].Energies, 2024, 17.
- [18] Wang F, Chen X J Automatic recognition method for substation instrument panel display based on Fast R-CNN and DeepLabV3+[J]. Journal of Engineering Design, 2024, 31 (06): 750-756
- [19] Kim B C , Son B .Crack Detection in Concrete Images using a Dilatation and Crack Detection Algorithm based on Image Processing[J]. Society for Advanced Composite Structures, 2022.
- [20] Xue X , Luo Q , Bu S S S .Citrus Tree Canopy Segmentation of Orchard Spraying Robot Based on RGB-D Image and the Improved DeepLabv3+[J].Agronomy, 2023., 13(8).
- [21] Su Z.P., Li J.W., Jiang J.W., Lu Y.L., Zhu M.. Semantic segmentation method for remote sensing images based on improved DeepLabV3+[J]. Advances in Lasers and Optoelectronics, 2023, 60(6):349-356.
- [22] Ber Z C, Lu Y C, Zhu Y X, Ma X H, Duan E Z. A caged dead duck recognition method based on improved Mask R-CNN[J]. Journal of Agricultural Machinery,2024,55(07):305-314.
- [23] Zhao L. Intelligent detection method of in-service bridge diseases based on visual data fusion and machine learning algorithm[J]. Computing Technology and Automation,2023,42(04):47-52.
- [24] Teng S , Liu A , Chen B , Wang J, Wu Z, Fu J. Unsupervised learning method for underwater concrete crack image enhancement and augmentation based on cross domain translation strategy[J].Engineering Applications of Artificial Intelligence, 2024, 136.
- [25] Hu M Y, Xia X, Yang C X, Cao J J, Chai X J. Design and implementation of semi-supervised image annotation system based on deep learning[J]. Journal of China Agricultural University,2021,26(05):153-162.
- [26] Ziying M , Shaolin H , Ye X K .Fine Crack Detection Algorithm Based on Improved SSD[J].):43-47.
- [27] Jenipher V N , Radhika S .Lung tumor cell classification with lightweight mobileNetV2 and attention-based SCAM enhanced faster R-CNN[J].Evolving Systems, 2024, 15(4):1381-1398.
- [28] Bai T .Multiple Object Tracking Based on YOLOv5 and Optimized DeepSORT Algorithm[J].Journal of Physics: Conference Series, 2023, 2547(1).
- [29] Gertsvolf D , Horvat M , Aslam D ,et al. A U-net convolutional neural network deep learning model application for identification of energy loss in infrared thermographic images[J].Applied Energy, 2024, 360.
- [30] Hao J F, Li Y T, Lai B W. Multi-model tensile detection segmentation system based on YOLOv5-seg[J]. Modern Computer, 2023, 29(16):1-7.

Multi-Objective Optimization of Construction Project Management Based on NSGA-II Algorithm Improvement

Yong Yang*, Jinrui Men

School of Energy and Building Engineering, Shandong Huayu University of Technology, Dezhou, 253000, China

Abstract—In the building industry, one of the key components to ensuring a project's successful completion is multi-objective project management. However, due to its own limitations, the traditional multi-objective management approach for projects is no longer able to meet the requirements of building construction and urgently needs to be improved. This is because the construction industry is becoming more competitive and construction standards are improving. Traditional methods for multi-objective optimization typically involve simply summing multiple objectives with weights, overlooking the interdependencies among these objectives. These methods often get trapped in local optimal solutions and rely heavily on predefined models and parameters, limiting their adaptability to sudden changes during the construction process. Therefore, a multi-objective management approach based on multi-objective genetic algorithm for construction projects is proposed. It enables in-depth analysis and comprehensive optimization of the complex relationships between objectives, leading to more informed decisions. By facilitating rapid iteration and adaptation, it enables timely adjustments and optimizations to ensure that project goals remain consistent in complex and dynamic environments. In the experimental validation, the NSGA-II algorithm achieved a significant accuracy of 0.642 and success rate of 0.504 on the VOT dataset, both of which improved by about 1.0% and 0.6% compared to the comparison algorithm. Experimental results on the TrackingNet dataset revealed that the algorithm achieved an accuracy of 0.791 and a success rate of 0.763, while it still maintained an accuracy of 0.542 and a success rate of 0.763 in the face of occlusion. The enhanced multi-objective genetic algorithm had higher accuracy and success rates. This demonstrates the efficiency and excellence of the multi-objective management optimization approach suggested in this study for building projects. The research results have some application value in the multi-objective optimization of engineering projects.

Keywords—NSGA-II algorithm; improvement strategy; construction engineering; project management; multi-objective optimization

I. INTRODUCTION

With the development of the economy and the continuous acceleration of urbanization, the construction industry has gradually become an important part of the national economy. When carrying out construction, project management is extremely important, which involves the economic benefits of the project, quality, safety, environmental protection and other aspects of the requirements [1]. Current construction project management often adopts the traditional method of single objective or simple weighting for optimization, which makes it

difficult to effectively balance the trade-offs between different objectives, and lacks the flexibility to respond to dynamic changes in demand. This limitation not only affects the overall efficiency and sustainability of the project, but also renders the project inadequate in the face of emergencies. Therefore, it is particularly important to propose a multi-objective optimization (MOO) method with strong adaptability and high computational efficiency in a dynamic construction environment. Consequently, it offers enhanced scientific and comprehensive decision support for construction project management, thereby promoting the coordinated development of resource utilization, economic benefits, and environmental protection in the construction industry. This area of research is of particular interest to experts and scholars in the industry. The research structure is divided into five sections. Section I outlines the importance of MOO in construction project management, as well as the problems and challenges in the current research. Section II is to construct and introduce the construction and solution of MOO model of construction engineering project, and discuss the objective function model of construction time, cost, quality and environment management. Section III shows the empirical test of the algorithm, including experimental design, data set selection and performance evaluation. Section IV discusses the experimental results and analyzes the performance and advantages of the algorithm in different application scenarios. Finally, Section V is the conclusion, which summarizes the main findings, practical application value, and future research direction.

In related research, Hamidreza et al. created a novel machine learning (ML) model to address the issue that current labor estimation algorithms often only take particular construction project types or specific influencing factors into account. At the work package level, the model could forecast the labor resource use time series. The results of the study indicated that the model could be used to predict the utilization of labor resources in construction projects, which could help in resource allocation. It was also able to prioritize the available resources to improve the overall performance of the project [2]. To dynamically manage the preliminary costs of a construction project, Zhouxin et al. proposed an artificial intelligence-driven analytical model for estimating and controlling construction costs. ML had the potential to enhance cost estimation during the construction process' procedural stage. The study's findings demonstrated that the ML model could be used to optimize workflow for cost savings and the useful outcomes of data-driven management [3]. Ghorqi et al. proposed a

*Corresponding Author

computational model for a multi-objective (MO) whale optimization algorithm (WOA) based on the Pareto profile to minimize the overall project delay time and associated costs. A comparative study revealed that the MO WOA-supported scheduling technique and solution approach were implemented, and this led to notable gains [4]. The process of cost optimization in construction projects requires maximizing value through effective resource management, cost control and achieving project goals within budget. Therefore, MAJDI Bisharah et al. identified the variable aspects that had significant impact on project cost (PC) through feature selection method. The outcomes revealed that it could effectively allocate resources to achieve project success and improve profitability [5]. In the construction project management mentioned above, the model improves resource management under a single objective, but it primarily focuses on cost control and lacks consideration of other key factors such as quality and environmental impact. The primary motivation of the research is to incorporate additional objectives, such as project quality and environmental impact, to construct a more comprehensive MOO model. The objective is to achieve labor optimization and the balance between time and cost.

Van et al. constructed an analytical model based on both exploratory factor analysis and validation factor analysis. The study's findings showed that the model helped create the theory behind the variables affecting how effective the materials management process is. The key findings about the influencing factors could be used to measure the effectiveness of the materials management process [6]. Simon et al. classified elements into major underlying determinants that caused the majority of the performance deviations, so reducing the scope of construction cost and time management to a few concrete, important areas of attention. This improved the management of the variables influencing time and cost overruns in public construction projects. This supported and improved the rapid decision making required in a variable environment such as construction [7]. Aiming at the problem of efficiency improvement in construction project management, Si et al. put forward the application method of self-organizing digital concept in data mining and intelligent planning. Therefore, it could achieve the improvement of efficiency in construction management practice stage. The study showed that the use of intelligent self-organizing data mining systems in this process could optimize the design and construction complexity, and evaluate the effectiveness of the model by integrating digital twin-driven intelligent construction and basic theoretical method. The results confirmed that the self-organizing model had a direct impact on time prediction planning [8]. Lawal et al. addressed the sustainability challenges facing the Nigerian construction industry by proposing ways to achieve better economic, social and environmental sustainability outcomes through resource optimization and reduced rework practices. Complex relationships between variables were assessed using a structural equation model based on covariance. Path analysis revealed a significant positive correlation between resource optimization and rework reduction and the social and environmental outcomes of construction firms [9]. Due to the identified conflict between efficiency and environmental impacts in off-site construction, Zheng et al. proposed a MOO framework. The framework incorporated a grouping technique

for hybrid flow prefabricated production scheduling, aiming to minimize carbon emissions and reduce late/early departure penalties. The proposed approach reduced carbon emissions by an average of 37.5% through real case studies, while late/early penalties were reduced by more than 10% [10]. A simulation-based method was presented by Sensenses et al. to maximize the cost-time trade-off for project planning issues in the face of uncertainty. Several project schedule collapse scenarios that produced equally plausible realizations were taken into consideration during this approach. The suggested approach has a considerable deal of potential to improve project management, according to experimental data [11]. While the construction project management approach outlined above provides a basic management theory, it lacks a discussion of the joint optimization of multiple objectives. In addition, insufficient consideration of factors affecting project quality and the environment leads to a one-sidedness in optimization results. It also provides no empirical verification for navigating dynamic and complex project environments. The research motivation is to develop a MOO framework that incorporates the relationship between material management and other objectives, along with a more flexible and adaptive algorithm to address dynamic changes in the construction environment. Research models include weighted sum, particle swarm optimization, and MO genetic algorithms. The weighted sum method can be subjective in weight selection and may not capture nonlinear relationships between objectives. Particle swarm optimization can suffer from premature convergence in complex problems, while MO genetic algorithms often have long convergence times and lack flexibility. In contrast, NSGA-II is chosen for this study because of its ability to effectively discriminate between high and low quality solutions. By utilizing crowding distance (CD), it maintains solution diversity and delivers high-quality results quickly. Additionally, it adapts well to both linear and nonlinear relationships among objectives, making it suitable for various MOO challenges, including construction projects.

In summary, most of the existing studies only consider the coordination and optimization of two objectives in the construction project. However, the existing research on MOO mostly focuses on the coordination of single or limited objectives, such as duration and cost, quality and environment. Many studies tend to ignore the complex relationships between them, which leads to the lack of applicability and comprehensiveness of optimization results in practical decision-making. Many models in the existing literature make the simplifying assumption that the environment and parameters remain static. This approach fails to adequately address the dynamic adjustment requirements that arise during project execution. Despite the prevalence of the MOO algorithm, it exhibits suboptimal efficiency, particularly in complex MO scenarios where the computational complexity is substantially elevated. In view of the above gaps, this study proposes a MO management method based on the improved NSGA-II algorithm. This method aims to comprehensively optimize project duration (PD), cost, quality, and environment, and to solve the problem of insufficient treatment of multiple interactive objectives in the current literature. The enhancement of the NSGA-II algorithm has been demonstrated to improve the adaptability of the algorithm in dynamic, changing

environments. This enhancement enables the algorithm to respond to changes in the project in real time, thereby ensuring the effectiveness of the optimal solution. An algorithm efficiency optimization strategy is proposed to improve computational efficiency and reduce execution time. The research innovation is mainly reflected in the improvement of NSGA-II, including the pre-computation of the dominant relationship between individuals to reduce the computational complexity of non-dominant sorting. The sorting strategy has been demonstrated to expedite the identification of congestion distance. Rather than employing the conventional adjacent calculation, local information is utilized to expedite the calculation process and enhance the diversity and velocity of selecting high-quality solutions. The shared fitness mechanism is adopted to make the lower level individuals get higher priority in the selection. The main contribution of this research is to improve the responsiveness and computational efficiency of the model in project management in the face of complex dynamic environments, realize real-time adjustment and optimization of the actual construction, provide important reference value for the sustainable development of the construction industry, and fill the management efficiency gap of the current MOO technology.

II. EP MULTI-OBJECTIVE MODEL: CONSTRUCTION AND SOLUTION

A. MOO Model Construction for EPs

In construction projects, time, cost, quality, and environment are key interrelated management objectives that influence outcomes. A well-structured schedule is essential for smooth progress, which has a direct impact on operating costs and customer satisfaction. Effective cost control addresses both DCs (such as materials and labor) and IDCs (such as management fees). MOO helps managers reduce costs while maintaining quality and schedules. High-quality buildings reduce maintenance costs and safety risks, improve customer

satisfaction, and require rigorous quality control to meet relevant codes and standards. In addition, managing the environmental impacts of construction - such as dust, wastewater, and noise pollution - enhances a company's corporate social responsibility and market competitiveness, especially in the context of green building certification. Reasonable arrangement of PD is an important prerequisite for MO management of EPs, which is also one of the main objectives of management. To guarantee that the project can be finished in the allotted time with both quality and quantity, management of PD necessitates dynamic adjustment of the construction time of each step in accordance with the plan. The general OF model of PD management is shown in Eq. (1) [12].

$$B = \max_{L \in L_m} \left(\sum_{(i,j) \in L_m} t_{ij} \right) \quad (1)$$

In Eq. (1), B denotes the PD. L_m denotes the set of feasible paths. (i, j) denotes the project process. t_{ij} denotes the time required for a single process. Cost management of a project is a dynamic control process. The structure of its components and its relationship with the duration are shown in Fig. 1.

In Fig. 1(a), the PC of the project is mainly composed of two parts: direct cost (DC) and indirect cost (IDC). The DC project construction is directly related to the project, including land use, machinery, construction equipment, etc. IDC are not directly related to the construction of the project but must be spent on the part of the project, mainly for personnel wages, management fees and so on. The purpose of PC management is mainly to carry out reasonable planning for all costs in the whole process of project construction. Furthermore, the project schedule is guaranteed to control the costs required for the project as far as possible, to maximize the economic benefits. The OF model of PC management is shown in Eq. (2) [13].

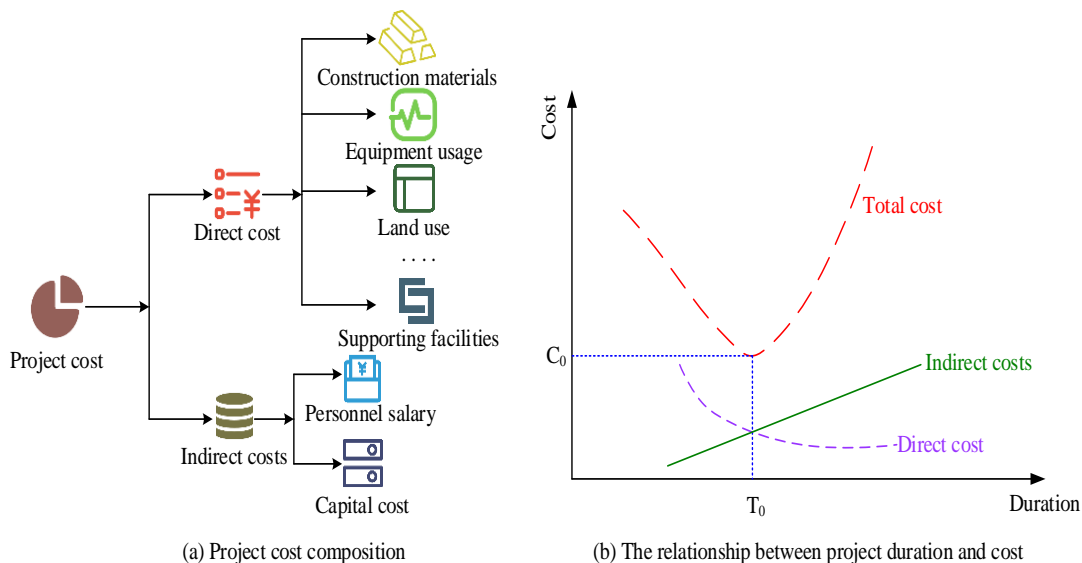


Fig. 1. Schematic diagram of construction project duration cost optimization model.

$$C = \sum_{(i,j) \in L_m} (C_{ij}^d + C_{ij}^i) \quad (2)$$

In Eq. (2), C represents the total PC. C_{ij}^d denotes the DC required to complete the process. C_{ij}^i denotes the IDC required to complete the process. The DC of the project is declining in Fig. 1(b) as the project's duration rises. This is because the shorter the project duration, the more labor and machines are used per unit of time, which increases the DC of the project. However, a reduction in PD also reduces the cost of labor, project management, etc., thus reducing the IDC of the project. The total cost (TC) of a project decreases and then increases as the PD increases. When the PD increases to a certain level, there will be a minimum value (MinV) of the TC of the project. This MinV can be solved using the OF optimization model of PD-cost to obtain the optimal solution between PC and duration, as shown in Eq. (3) [14].

$$C = \sum_{(i,j) \in L^m} \left[c_{ij}^d + a_{ij} (t_{ij}^n - t_{ij})^2 \right] + \sum_{(i,j) \in L^m} b * t_{ij} \quad (3)$$

In Eq. (3), a_{ij} denotes the marginal cost factor of the process. t_{ij}^n represents the planning time needed to finish the procedure. The real time needed to finish the process is indicated by the letter t_{ij} . b denotes the project overhead cost required for a single day. The established quality standards in each stage of the building process serve as the foundation for the project's quality management. This enables the quality of the project (QOP) to meet the requirements of the contract, the industry and other parties. The main influencing factors of project quality and its relationship with PD are shown in Fig. 2.

In Fig. 2(a), the factors affecting the QOP are multifaceted, mainly including the technical ability of the personnel, the quality of construction materials, construction technology, etc., which affect the quality of each process of the project. When managing the QOP, it is necessary to find out the deficiencies in these factors according to the actual situation of the project and take corresponding measures to improve [15]. Unlike calculating the duration and cost of a project, the quality of a

project is highly subjective. It is difficult to calculate directly. It is necessary to quantify it in an attempt to establish the OF model. The study set different weights according to the impact of different factors on the overall QOP to establish the OF model of project quality, as shown in Eq. (4).

$$D = \sum_{(i,j) \in L_m} w_{ij} q_{ij} \quad (4)$$

In Eq. (4), D denotes the project quality. w_{ij} denotes the impact weight of project quality. q_{ij} denotes the actual quality level. To build the objective function model of project quality, it is essential to quantify and weight each factor affecting the overall quality, which has a significant impact on the model performance. Key influencing factors include the technical ability of construction personnel, material quality, construction standards and technologies, project management and supervision processes, and the external environment. To quantify these factors and set their weights, the expert scoring method can be applied, followed by the weighted average method to calculate the weight for each factor. In Fig. 2(b), there is a roughly positive correlation between project quality and its required duration. The longer the duration of the project, the more construction time for each process, and theoretically the overall QOP will be better. However, the duration of a project obviously cannot be infinitely long, so the QOP can only fluctuate within a given duration. As the PD continues to increase, the improvement of project quality is also more and more limited, gradually converging to a fixed value. The OF optimization model of PD-quality is shown in Eq. (5).

$$\beta_{ij} = \frac{q_{ij}^n - q_{ij}}{(t_{ij}^n - t_{ij})^2} \quad (5)$$

In Eq. (5), q_{ij}^n denotes the level of quality to be achieved.

The environmental management of the project is one of the key elements of project management nowadays, this is because all EPs will inevitably pollute the surrounding environment during the construction process and must be controlled [16]. The types of environmental pollution from projects and their relationship with PD are shown in Fig. 3.

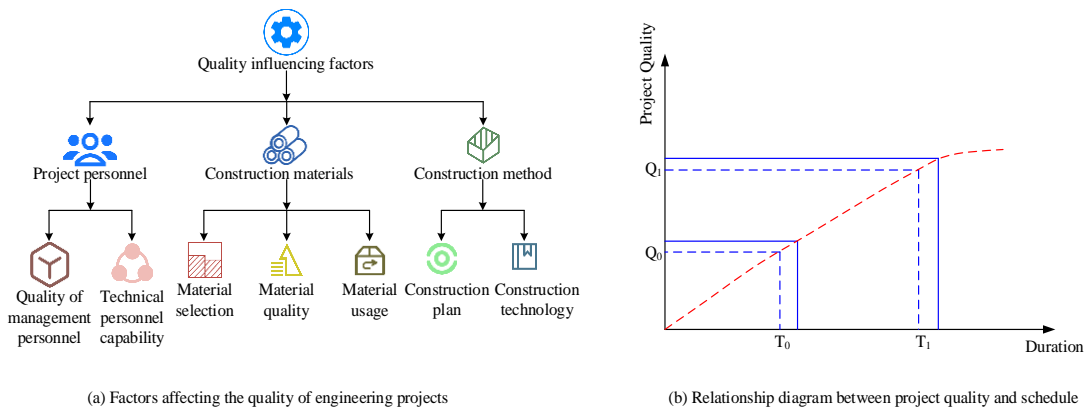


Fig. 2. Schematic diagram of project duration quality optimization model construction.

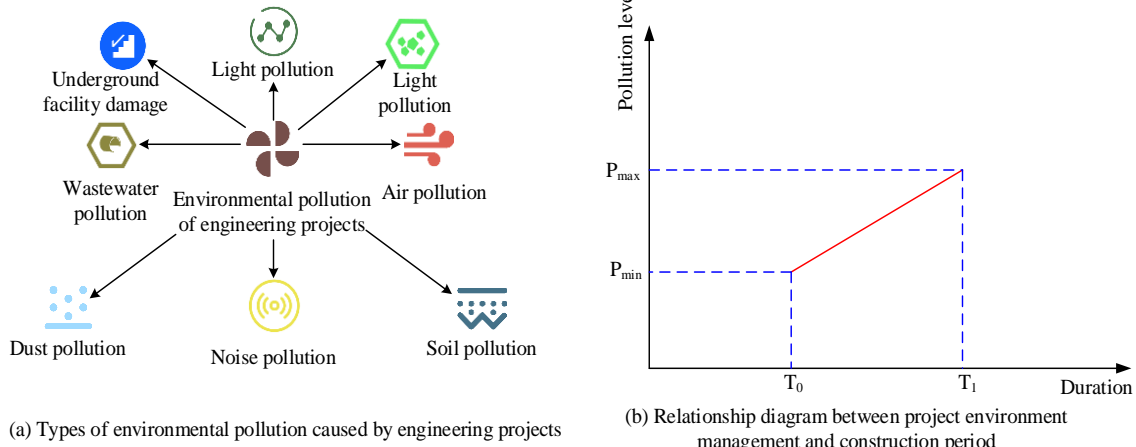


Fig. 3. Schematic diagram of construction period environment optimization model construction.

In Fig. 3(a), in the construction process of EPs, the pollution to the environment mainly includes dust pollution, light pollution, garbage pollution, water pollution, noise pollution and so on. All these pollution are unavoidable for engineering construction. The environmental management of the project is to minimize the degree of pollution of the building construction on the surrounding environment under the condition of meeting other requirements. The study uses factor analysis to assess the impact of different types of pollution on the environment. Factor analysis is a multivariate statistical technique used to identify latent variables that influence observable variables, help understand relationships among factors, and reduce data dimensionality for improved modeling. First, several measures of environmental impact are selected as inputs to the analysis using accurate and valid data. The correlation coefficient between these variables is calculated to assess their interrelationships. Principal component analysis is then performed to extract the underlying factors, followed by factor rotation to ensure that each factor is significantly loaded by only a few variables. The factor load matrix is analyzed to determine which variables most influence each factor. Finally, an objective function model for environmental management is established, as presented in Eq. (6).

$$f(x) = \min E = (L \times W) \times \sum_{ij \in L} (t_{ij} \times e_{ij}) + K \quad (6)$$

In Eq. (6), E indicates the degree of environmental pollution. L denotes the distance between the project site and the residential area. W denotes economic indicators. e_{ij} denotes the evaluation value of pollution factors. K denotes the proportion of resource consumption. In Fig. 3(b), PD and environment are interacting with each other. The PD will affect the environment, and at the same time, the environmental factors will also affect the planning of the PD and the actual completion time of the process. Fig. 3(b) shows a positive correlation between project environmental management and PD. As PD increases, the need for environmental management increases, as does its complexity. Longer construction periods lead to more activities, greater environmental impacts, and higher management costs to comply with environmental standards. The environmental impacts of prolonged

construction may vary at different stages, and the diversification and complexity of activities require more stringent management measures to address emerging environmental risks. Since it is not possible to quantitatively calculate the various types of pollutants, the study obtains the OF optimization model of PD-environment from the time that each process lasts, as shown in Eq. (7) [17].

$$e_{ij} = e_{ij}^n - \varpi_{ij} (t_{ij}^n - t_{ij}) + e_{ij}^s \quad (7)$$

In Eq. (7), e_{ij} denotes the actual degree of environmental pollution from the project. e_{ij}^n denotes the degree of environmental pollution during the planning period. ϖ_{ij} represents the rate of change of environmental pollution. e_{ij}^s represents the estimable part of environmental pollution during the planning period. Eq. (8) illustrates how a comprehensive optimization model of PD-cost-quality-environment may be created by combining the aforementioned OF optimization models between PD and PC, quality and environment.

$$\left\{ \begin{array}{l} \min T = \max \left(\sum_{(i,j) \in L^m} t_{ij} \right) \\ \min C = \sum_{(i,j) \in L^m} \left[c_{ij}^n + a_{ij} (t_{ij}^n - t_{ij})^2 \right] + \sum_{(i,j) \in L^m} b^* t_{ij} \\ \max Q = \sum_{(i,j) \in L^m} \lambda_{ij} \left[q_{ij}^n - \beta_{ij} (t_{ij}^n - t_{ij}) \right] \\ \min E = \sum_{(i,j) \in L^m} \left[e_{ij}^n - \varpi_{ij} (t_{ij}^n - t_{ij}) + e_{ij}^s \right] \end{array} \right. \quad (8)$$

In Eq. (8), β_{ij} denotes the marginal coefficient of project quality. λ_{ij} represents the influence coefficient of the process on the overall QOP. By solving the MOF optimization model of the project, the duration of the project can be reduced as much as possible within the standard requirements. This can result in

lower cost, higher quality and less pollution to the environment. In real-world projects, pollution is assessed in four dimensions: actual pollution levels, pollution levels during the design period, the rate of change in pollution levels, and projected pollution levels for the design period. The actual level of pollution is measured by on-site monitoring of pollutant concentrations, including particulate matter, noise levels, and water quality. Pollution during the design phase is assessed using environmental impact assessment reports or predictive models that identify potential sources of pollution at various stages of construction. Statistical models using historical data from similar projects help predict pollution levels. The rate of change in pollution is quantified by analyzing time series data, comparing pre- and post-construction levels of pollutants, and calculating annual rates of change. During the design phase, pollution can be managed through established environmental goals and regulations, with ongoing monitoring to assess actual emissions against allowable standards.

B. Project MOO Model Solution Based on NSGA-IIA

The NSGA-IIA is a type of genetic algorithm, which simulates the optimization method of biogenetic mechanism in nature, and is used to solve the MOF optimization model. Therefore, the research utilization will use this algorithm to solve the MOF model of the process project. Fig. 4 illustrates the algorithm's flow.

In Fig. 4, the first step of the NSGA-IIA is to generate an initialization population. The initialization population consists of a set of individuals. Each individual represents a feasible solution to the model. A non-dominant sorting is performed on

these individuals, and then crossover and mutation are used to generate a new generation of a sub-population. This sub-population represents a subset of the solution space. The second step is to merge these two populations, the current population and the newly generated subpopulation, to form a new population. Non-dominated sorting (NDS) of the merged population identifies multiple Pareto fronts. The purpose of NDS is to classify individuals into different ranks based on their dominance relationships. Individuals known as the first rank are optimal because no other individual dominates them. Individuals in the second rank are dominated by individuals in the first rank and so on. Diversity among individuals in the same rank is assessed by calculating the CD. For each goal, individuals are ranked according to the value of that goal. The distance between individuals is calculated and the crowding of the bordering individuals is recorded as infinite. The CD for intermediate people is the total of the differences between two adjacent persons on that target. The right people counts will be chosen to make up the next generation of the population, taking into account the CD and the outcomes of NDS [18-19]. Loop the above steps and stop running the algorithm when a predetermined number of iterations is reached. It also outputs the final result as the optimal solution of the MOF model. Due to the need for fast NDS and CD calculation, the NSGA-IIA has a high computational complexity and requires a long running time, especially when there are more optimization functions in the objective model [20]. Therefore, as shown in Fig. 5, improving the algorithmic process is required to lower the computational complexity and running time and optimize the MOF of EPs utilizing the NSGA-II method.

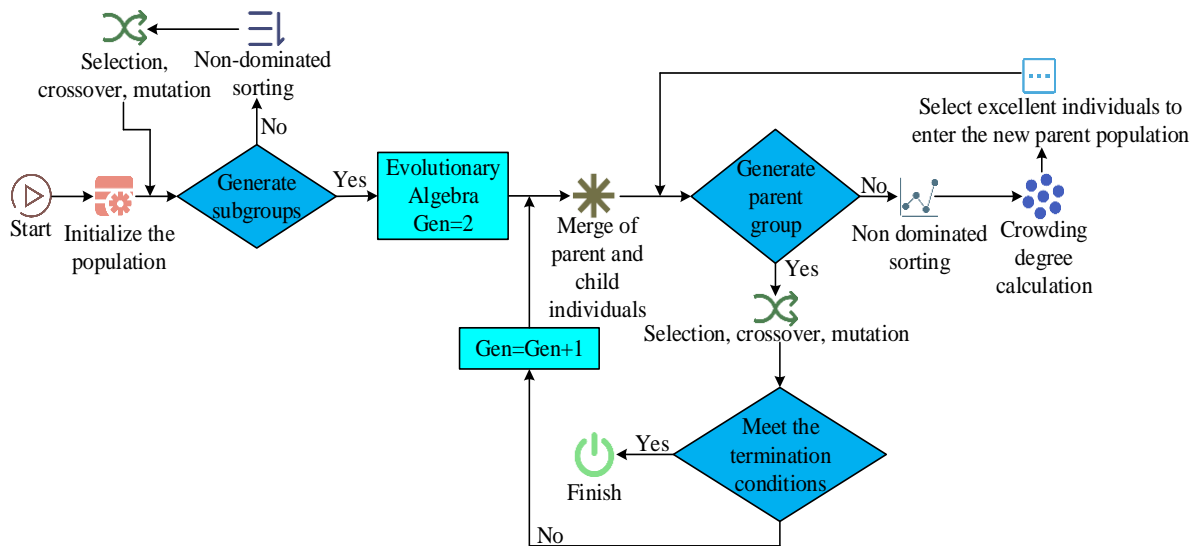


Fig. 4. Schematic diagram of SGA-II algorithm flow.

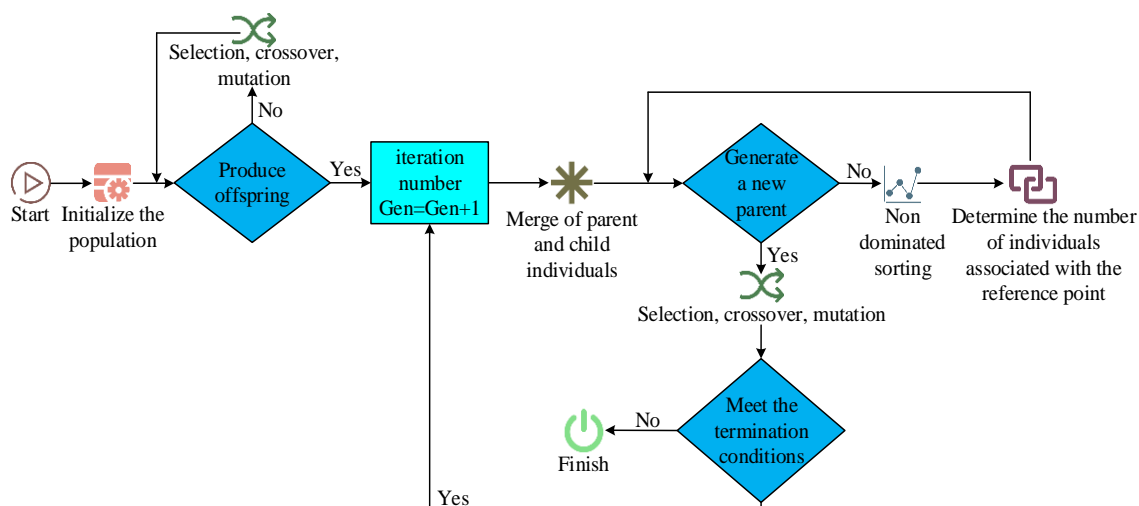


Fig. 5. Schematic diagram of the improved NSGA-IIA flow.

In Fig. 5, the improved NSGA-IIA is first optimized for fast NDS and CD calculation. The computational complexity of NDS can be reduced by pre-computing the dominance relationship between individuals. A more efficient sorting strategy is used to find the CD quickly. Secondly, NDS is used to rank the new population. Hierarchical management of populations is done through shared fitness. Among them, the value of shared fitness decreases accordingly with the increase of the rank, thus guaranteeing the priority of the lower ranked individuals in the selection process. The third step is to utilize the method of reference points, which are evenly arranged on the standard hyperplane. The individuals in the population are also uniformly divided so as to calculate the number of reference points, as shown in Eq. (9).

$$P = \binom{M + H - 1}{H} \quad (9)$$

In Eq. (9), P denotes the number of reference points. M denotes the dimension. H denotes the number of copies. Then the OF in the optimization model needs to be quantized to facilitate the association of the set reference points, as shown in Eq. (10).

$$f'_i(x) = f_i(x) - Z_i^{\min} \quad (10)$$

In Eq. (10), $f'_i(x)$ denotes the quantized function. $f_i(x)$ denotes the original OF. Z_i^{\min} denotes the MinV of the

function. After the quantization of the function is completed, it is also necessary to find the extreme point of the function, as shown in Eq. (11).

$$ASF(X, W) = \text{MAX}_{i=1:m} (f'_i(x) / W_i) \quad (11)$$

In Eq. (11), $ASF(X, W)$ denotes the ASF function. Based on the extreme points of the function, the corresponding function values can be obtained. The function values of each individual are unified into a plane of the same dimension as the established reference point, as shown in Eq. (12).

$$f_i^n(x) = \frac{f'_i(x)}{a_i} = \frac{f_i(x) - Z_i^{\min}}{a_i} \quad (12)$$

In Eq. (12), a_i represents the distance of each function value to the plane coordinate axis. The position at which each function value is placed is linked to the reference point, and the person represented by the function value with the smallest distance between them is found and linked to the reference point. Finally, the selected individuals are iterated until the optimal solution of the target model is achieved. The core of the improved NSGA-IIA for solving the MOF optimization model of an EP is the cross-variation process of the population individuals, as shown in Fig. 6.

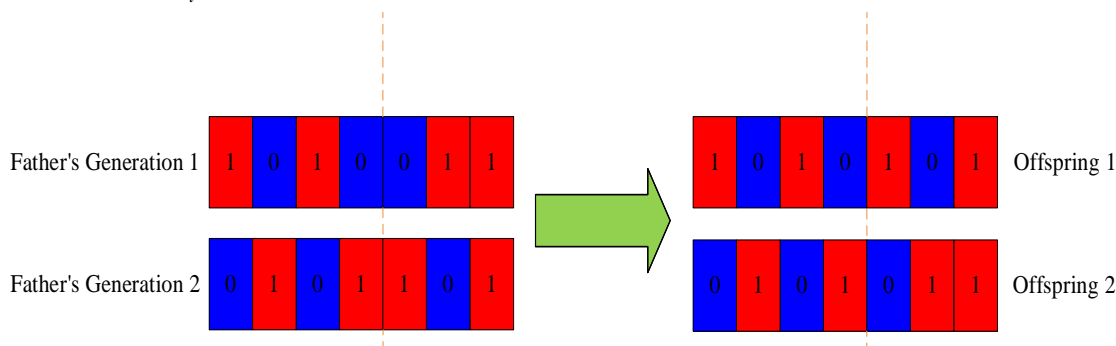


Fig. 6. Schematic diagram of BX crossover operator.

In Fig. 6, the improved NSGA-IIA utilizes the SBX for the simulation of the single-point crossover operator, which ensures that useful information can be obtained from the newly generated populations, as shown in Eq. (13).

$$\begin{cases} x_{1j}(t) = 0.5 * \left[(1 - \gamma_j) x_{2j}(t) + (1 + \gamma_j) x_{1j}(t) \right] \\ x_{2j}(t) = 0.5 * \left[(1 + \gamma_j) x_{2j}(t) + (1 - \gamma_j) x_{1j}(t) \right] \end{cases} \quad (13)$$

In Eq. (13), x_{1j} and x_{2j} denote individuals of the parent population. γ_j denotes the degree of mixing of individual characteristics of the parent population. The formula for the degree of mixing is shown in Eq. (14).

$$\gamma_j = \begin{cases} (2u_j)^{\frac{1}{\eta+1}} & , u_j \leq 0.5 \\ \left[\frac{1}{2(1-u_j)} \right]^{\frac{1}{\eta+1}} & , u_j > 0.5 \end{cases} \quad (14)$$

In Eq. (14), u_j denotes a constant between 0 and 1. η denotes the distribution index. Due to the minimal population variety at the beginning of the algorithm, the search process may settle on locally optimal solutions. It is investigated that the diversity of the population can be enhanced by dynamically increasing the crossover probability and mutation probability. This raises the likelihood that a globally optimal solution will be found by enabling the algorithm to search a larger solution space. As the number of iterations increases, the mutation probability is gradually reduced, which can effectively reduce the occurrence of "catastrophic mutation". Catastrophic variation refers to the fact that in the later stages of the search, a large number of sudden changes may destroy the good solutions that have been found. By controlling the diversity of the algorithm at a later stage, the algorithm is stabilized. This facilitates convergence to high quality potential solutions as shown in Eq. (15).

$$\begin{cases} p'_c = p_c \times \left(1 - \frac{\text{gen}}{\text{max gen}} \right) \\ p'_m = p_c \times \left(1 - \frac{\text{gen}}{\text{max gen}} \right) \end{cases} \quad (15)$$

In Eq. (15), p_c denotes the adaptive crossover probability. gen denotes the current iteration number. max gen the maximum number of iterations. While the MOO model presented in this study focuses on construction projects, it has broader applicability in other fields. For example, in manufacturing, key objectives such as schedule, cost, and quality must be balanced, with production adjustments made according to market demand. Similarly, public infrastructure projects must navigate complex codes and standards while addressing schedule, cost, quality, and environmental concerns. Using this MOO approach, project managers can better manage stakeholder relationships and effectively meet multiple project objectives.

III. ANALYSIS OF NSGA-II AND ITS IMPROVED ALGORITHMS

A. Performance Analysis of NSGA-IIA

To verify the effect of the proposed method, the implementation environment in the study is mainly focused on the actual construction scene of the construction project, including project planning, resource allocation, and construction management. The improved NSGA-II algorithm will be applied to a typical construction project that must meet multiple objectives such as time reduction, cost control, quality assurance, and environmental protection within a limited time and budget. By monitoring and analyzing real-time data from all phases of the project, such as construction progress, resource consumption, and environmental impact, the algorithms will dynamically adjust optimization strategies to cope with unforeseen changes and challenges. Furthermore, the study will address the distinct requirements of various construction projects (e.g., residential, commercial, public facilities, etc.) during the implementation phase. These unique implementation environment factors will subsequently encourage the refinement and optimization of the algorithm, thereby ensuring the efficacy and relevance of the final decision. To verify the success rate and accuracy of MO genetic algorithm in solving item function model, a series of systematic evaluation measures and tools are adopted. In terms of algorithm implementation, based on Python programming language, NumPy and SciPy libraries are used for numerical computation, and the results are visualized with Matplotlib library. To verify the success rate (SR) and accuracy of the MO genetic algorithm in solving the item function model, the study will conduct comparative analysis experiments on two different datasets. The super-parameters of the model are set as follows: The population size is 100, and the larger population size can provide a better diversity of solutions, avoid premature convergence, and help find the global optimal solution. The mutation probability is set to 0.1, which is mainly used to keep moderate changes and prevent the algorithm from falling into the local optimal solution. The crossover probability is set to 0.9, which is mainly used to speed up the convergence rate of solutions, and can effectively produce high-quality descendant solutions in MOO. The number of iterations is set to 500 to ensure that the algorithm converges in the right time and obtains the best solution. The experiment uses the VOT and TrackingNet datasets. The VOT dataset is a benchmark for evaluating various visual target tracking algorithms, while TrackingNet focuses on efficient performance evaluation, providing various scenarios, video sequences, tagging boxes, and metrics for measuring algorithm accuracy and robustness. The validation measures adopted in this study include key performance indicators such as accuracy, success rate, and center position error. Moreover, the effect of the improved algorithm is comprehensively evaluated through comparison and analysis with existing MOO algorithms. The analysis compared the SiamRPN, SiamFC and SiamRPN18++ algorithms. SiamRPN demonstrated high accuracy through its regional proposal network, effectively selecting from multiple candidate regions, strong positioning capabilities, and a MOO process. The SiamFC algorithm processes target features with consistent convolution operations and retrieves them by calculating feature similarity. SiamRPN18++, on the other hand, uses a

deeper network architecture to enhance feature extraction, improving adaptability and accuracy for multiple tracking targets. Firstly, the study conducts overlap rate and center

position error analysis on VOT dataset. The results are shown in Fig. 7.

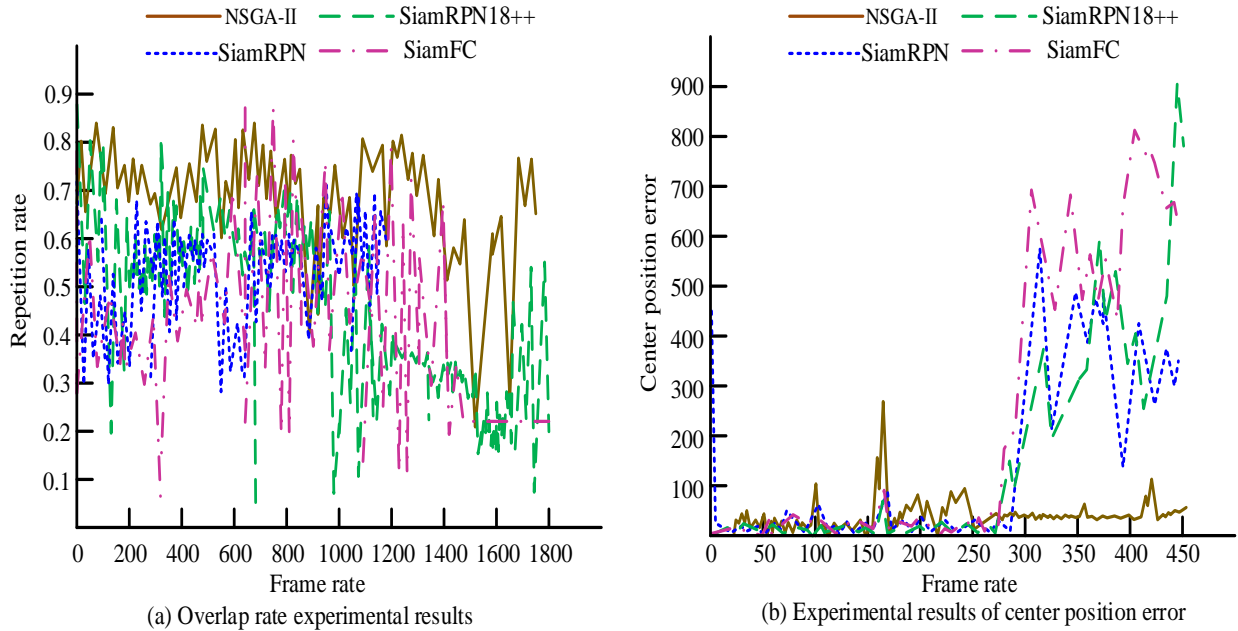


Fig. 7. Comparison analysis curve of overlap rate and center position (I).

In Fig. 7, the NSGA-II has the optimal overlap rate of 0.75, which is the best. Compared with other algorithms, the average overlap rate increases by 0.156%, which effectively optimizes the efficiency of solving the target features. This lays the foundation for the improvement of the solving performance in the case of multiple objectives. The NSGA-II performs well in the tracking error of the sequence center position with minimal error. Compared with other basic algorithms, NSGA-II algorithm is able to reduce the center position error by 5.84, 30.63, and 43.42 pixel points, respectively. When the number of objective model functions is too many, NSGA-II algorithm

significantly reduces the center position error. The proposed algorithm makes full use of the evaluation mechanism of non-dominant sorting and congestion distance to maintain the diversity of solutions and obtain better solutions in a short time. In target tracking tasks, the algorithm can more accurately capture the location information of the target, thereby improving the positioning accuracy and reducing the error. To better show the performance of NSGA-II, the study conducts comparative experiments on the VOT dataset. The SR and accuracy change graphs are shown in Fig. 8.

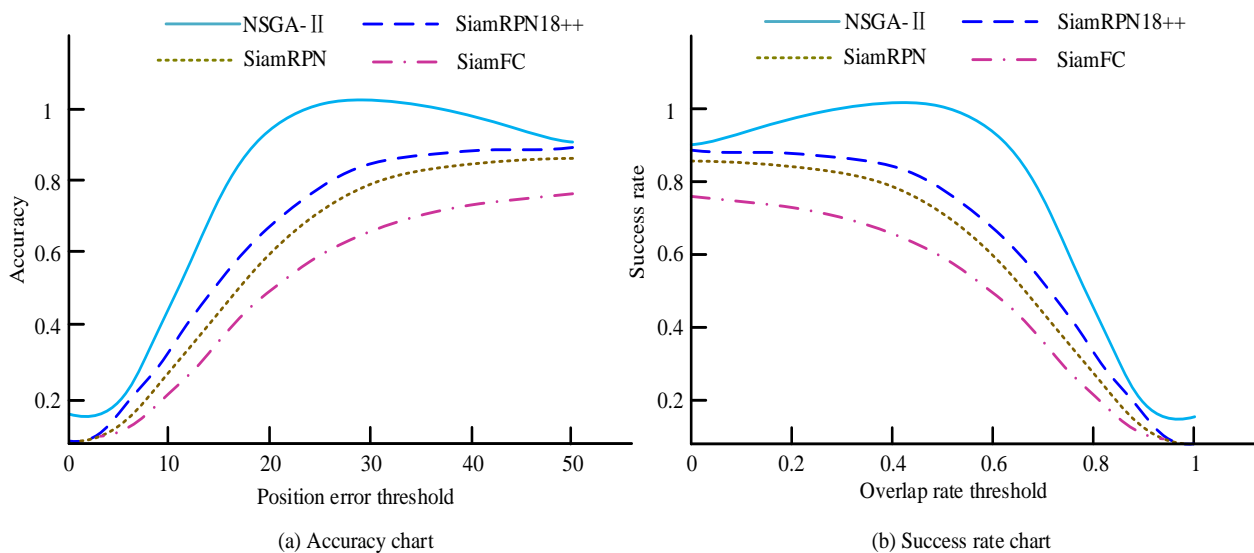


Fig. 8. Success rate and accuracy chart on the VOT dataset (I).

Fig. 8(a) represents the accuracy results of different algorithms in the validation of VOT dataset. The results display that the NSGA-IIA proposed in the study has an accuracies of 0.642. Its overall improvement in accuracy over SiamRPN18++, SiamRPN, and SiamFC algorithms is about 1.0%. Fig. 8(b) represents the SR results of different algorithms in the validation of VOT dataset. The findings show that, in comparison to the other three algorithms, the NSGA-IIA suggested in the study has a SR of 0.504, indicating an overall improvement in accuracy of roughly 0.6%. The results indicate that NSGA-II has higher localization accuracy in tracking the target and can capture the actual position of the target more efficiently with less error. In specific scenarios, it can maintain its reliability in a variable environment. Considering the relatively limited size of the VOT dataset, the effectiveness of

the algorithm needs to be examined more comprehensively. The obvious improvement in the results in Fig. 8 shows the effectiveness of the algorithm in dealing with complex multi-target problems, which can better detect and identify target features and reduce the number of false tracks. This enhancement in performance provides a robust technical foundation for dynamic management and decision-making in construction projects, thereby enabling timely responses to unexpected situations and strategic adjustments in a changing construction environment. Therefore, the research conducted further experiments on the TrackingNet dataset to verify the success rate and accuracy of NSGA-II algorithm under different environmental conditions, including the availability and price fluctuations of construction materials, the availability and cost of labor, as shown in Fig. 9.

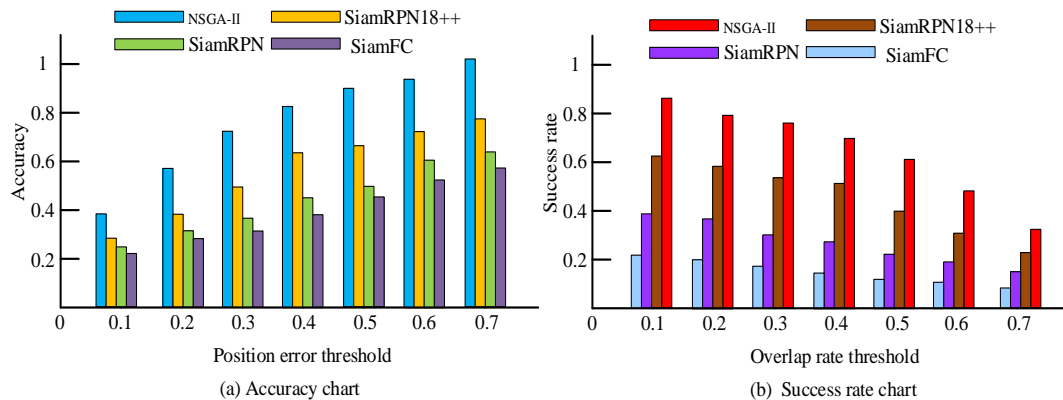


Fig. 9. Success rate and accuracy graph on the TrackingNet dataset (I).

Fig. 9(a) represents the accuracy results of different algorithms on TrackingNet dataset. The results display that the accuracy of the proposed algorithm has 0.791, which is an overall improvement of about 1.2% compared to the accuracy of SiamRPN18++, SiamRPN, and SiamFC algorithms. In the performance detection under occlusion attribute, the accuracy of the proposed algorithm has 0.542, which is an overall improvement of 0.4% compared to the comparison algorithms. Fig. 9(b) represents the SR results of different algorithms in TrackingNet dataset. The results display that the SR of the proposed algorithm of the study has 0.763, which is a 1.8% improvement compared to the SR of the comparison algorithms.

Similarly, in the performance detection under occlusion attribute, the proposed algorithm has a SR of 0.763, which is an improvement of 3.3% compared to the SR of the comparison algorithm. The results indicate that NSGA-II still exhibits relatively strong robustness in dealing with complex and challenging tracking situations, proving its applicability and reliability in real application scenarios.

B. Performance Analysis of the Improved NSGA-IIA

To verify the overlap rate and center position error of the improved NSGA-IIA, the study will conduct a comparative analysis experiment on the VOT dataset, as shown in Fig. 10.

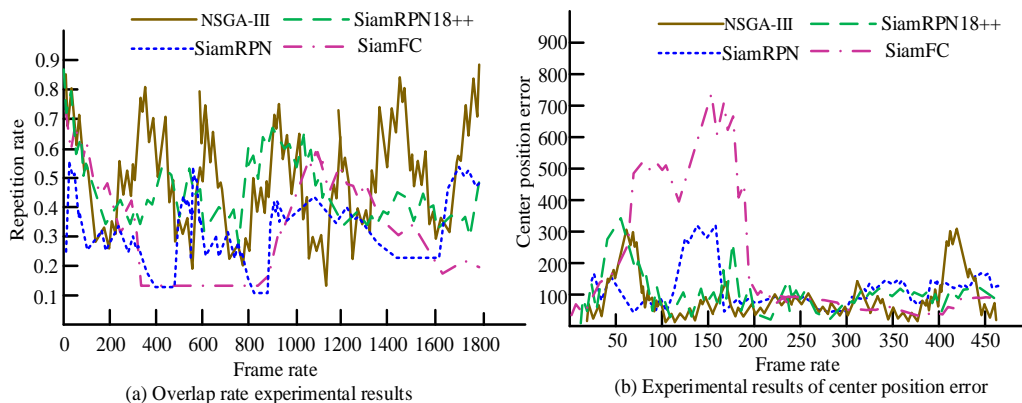


Fig. 10. Comparison analysis curve of overlap rate and center position (II).

In Fig. 10, the overlap rate of the improved algorithm is optimal. Compared with other algorithms, the average overlap rate of the algorithm is improved by 0.126. This indicates that the improved algorithm can obtain more accurate optimal solutions of the objective model, which improves the accuracy and robustness of the algorithm in solving the MOF optimization model. The improved algorithm is ranked first with the minimum error in the center position of the sequence.

Compared with other algorithms, its average center position error is reduced by 23.450 pixel points. This indicates that the improved method has a smaller center position error and is able to solve the best solution of the MOF model more accurately. The study uses the VOT dataset for a comparative experiment to confirm the enhanced algorithm's tracking performance, as illustrated in Fig. 11.

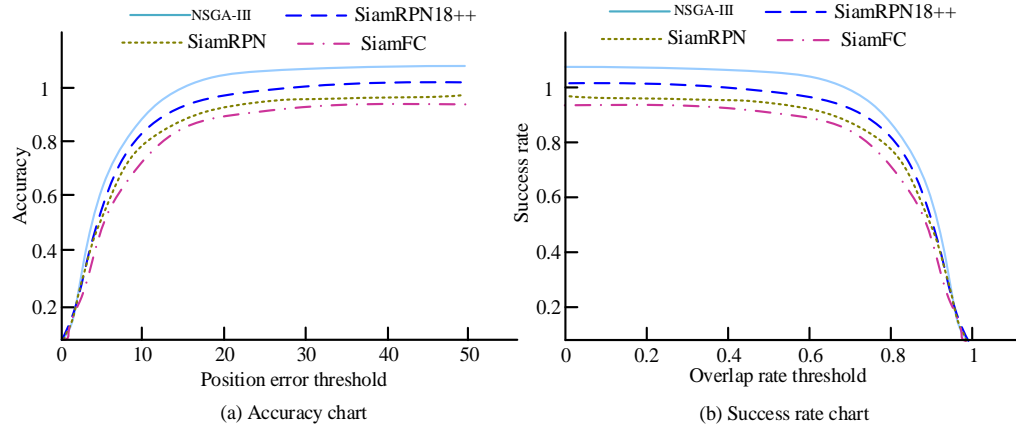


Fig. 11. Success rate and accuracy graph on the VOT dataset (II).

In Fig. 11, the SR and accuracy of the improved algorithm are 0.690 and 0.845, respectively. Compared with other algorithms, the average improvement is 1.5% and 1.2%. This indicates that the algorithm performs optimally among the tested algorithms and fully proves its effectiveness. In the performance test under the occlusion attribute environment, the accuracy and SR of the proposed algorithm are improved compared with the comparison algorithms. Among them, the research algorithm improves about 2.1% in accuracy, and the final accuracy is 0.918. It improves about 2.7% in SR, and the final SR is 0.691. The enhancement in the success rate signifies

the algorithm's capacity to effectively address occlusion and variations in the target, leveraging its adaptive characteristics and dynamic adjustment mechanism to ensure uninterrupted target tracking. In the context of complex construction projects, this advantage enables project managers to make faster and more accurate decisions in a highly dynamic construction environment. Consequently, this improves the overall project management efficiency and effectiveness. Fig. 12 displays the enhanced algorithm's validation findings using the TrackingNet dataset.

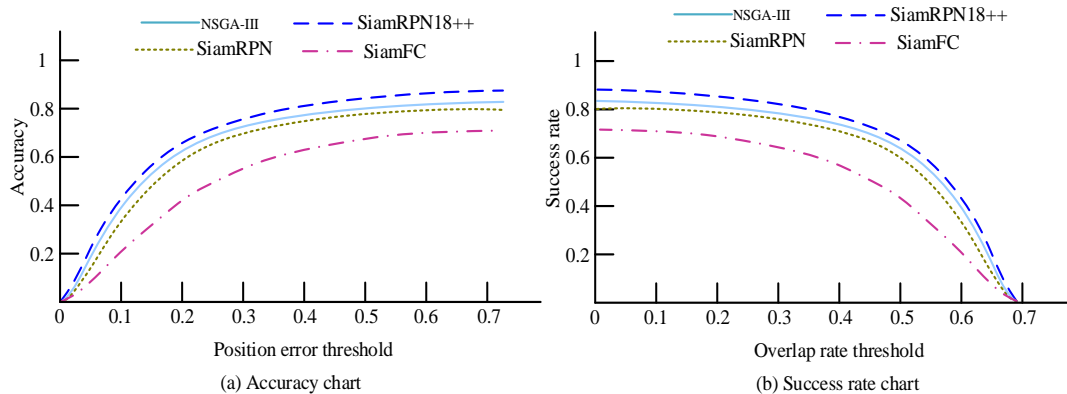


Fig. 12. Success rate and accuracy graph on the TrackingNet dataset (II).

In Fig. 12, the overall SR and accuracy of the improved algorithm under the TrackingNet dataset are 0.490 and 0.572, respectively, which are improved by 1.2% and 0.6% on average compared to other algorithms. By improving the algorithm, it shows good performance in the test of the dataset, which verifies the effectiveness of the improvement measures. In the presence of occlusion, the algorithm achieves an SR of 0.454%, which is 4.2% higher on average than other algorithms. It also achieves an accuracy of 0.536%, which is 1.5% higher than

other algorithms on average. This further confirms the effectiveness of the algorithm in performing MOO. The specific construction project is a sub-division project of a comprehensive commercial plaza. The total budget cost of the project is 5 million yuan, and the planned construction period is 120 days. The project management team is challenged to coordinate multiple objectives, including optimizing construction cost, ensuring construction quality, and reducing

environmental impact while meeting the deadline. The results of MOO in construction projects are shown in Table I.

TABLE I. RESULTS OF MULTI-OBJECTIVE OPTIMIZATION IN CONSTRUCTION PROJECTS

Goal	NSGA-II	Improved NSGA-II	Change
Duration (days)	120	110	-10
Total cost (ten thousand yuan)	500	480	-20
Quality Score (0-1)	0.75	0.85	0.1
Environmental impact score (score)	30	25	-5

Table I shows that based on the improved NSGA-II algorithm, the PD is reduced by 10 days from 120 days to 110 days. The TC is reduced from 5 million yuan to 4.8 million yuan, saving 200,000 yuan. The quality score of the project increases by 0.10 from 0.75 to 0.85. The environmental impact score is reduced by five points from 30 to 25. The findings indicate that the model effectively realizes the multiple objectives of reducing the construction period, minimizing costs, enhancing quality, and decreasing environmental impact. It also provides robust algorithmic support and a scientific foundation for the management of construction projects.

IV. RESULTS AND DISCUSSION

This study used the improved NSGA-II algorithm to analyze the MOO model for construction projects, focusing on the effective coordination of PD, cost, quality, and environmental impact. The simulation results showed a reduction in PD from 120 days to 110 days, indicating improved construction efficiency and less time wasted, facilitating earlier project delivery. On the cost front, total spend decreased from 5 million yuan to 4.8 million yuan, highlighting the significant savings achieved through optimized resource allocation while maintaining quality requirements. The quality score improved from 0.75 to 0.85, ensuring better overall project quality while reducing duration and cost. In addition, the environmental impact score decreased from 30 to 25 points, demonstrating algorithm's commitment to sustainable practices. Unlike traditional optimization methods, the improved NSGA-II could manage multiple objectives simultaneously, allowing project managers to adjust the weights of objectives based on situational needs to achieve optimal balance in complex environments. This flexibility was of paramount importance for contemporary construction projects, which frequently encountered rapid changes and unforeseen risks. The enhanced algorithm facilitated real-time updates to objectives and constraints, enabling timely strategy adjustments to ensure project effectiveness. In the same type of research, compared with the research in literature [2], this study used particle swarm optimization algorithm, and the average cost reduction was only 150,000 yuan. The study in reference [3] used a MO WOA, and the quality score was improved by 0.05. The research in literature [4] showed that the algorithm could achieve environmental quality improvement while reducing the score by only about 3 points. Compared with other relevant studies, the results of this study showed that the improved NSGA-II algorithm achieved more significant time reduction, cost saving,

quality improvement, and environmental impact reduction in the MOO of civil engineering projects.

V. CONCLUSION

The research built a MOO model that incorporated the project schedule, cost, quality, and environment in an attempt to address the issue of MO management optimization of construction projects. Moreover, the improved NSGA-IIA was utilized for solving and validation. Through experiments on the VOT and TrackingNet datasets, the NSGA-IIA performed well on the VOT dataset. The accuracy reached 0.642 and the SR was 0.504, which were 1.0% and 0.6% better than the comparison algorithm, respectively. On the TrackingNet dataset, the accuracy was 0.791 and the SR was 0.763. Moreover, the accuracy under occlusion was 0.542 and the SR was 0.763, demonstrating the robustness of the algorithm in complex environments. The improved algorithm exhibited a high accuracy and SR of 0.690 and 0.845 in the VOT dataset, which provided a more reliable solution for selecting the target model. This research contributed to the knowledge system by expanding MOO theory and proposing a model based on the improved NSGA-II algorithm, focusing on the duration, cost, quality, and environmental impact of construction projects. This model enriched the theoretical framework of MOO and offers new insights for researchers in architecture. For the AEC industry, the model enhanced project management efficiency, promotes sustainable development, and informed industry policies and standards. By facilitating better coordination of multiple objectives, the approach emphasizes the importance of balancing economic benefits with environmental protection, ultimately providing sustainable solutions that help organizations achieve their green building goals. However, the research is not without its limitations. The computational efficiency and stability must be improved in the face of extreme dynamic changes and high-dimensional targets. The presence of decision bias, stemming from incomplete or inaccurate data, is a notable concern. Consequently, future research endeavors will prioritize the further optimization of the algorithm's performance, the exploration of more adaptive mechanisms, and the development of parallel computing methods to enhance the prediction and analysis of data.

FUNDINGS

The research is supported by Dezhou development and Reform Commission, Dezhou Engineering Research Center of Intelligent Construction.

REFERENCES

- [1] Mehdi Tavakolan, Farzad Chokan, Mostafa Dadashi Haji. Simultaneous project portfolio selection and scheduling from contractor perspective. *International Journal of Construction Management*, 2024, 24(3): 298-313.
- [2] Hamidreza G, Ahmed H. Estimating labor resource requirements in construction projects using machine learning. *Construction Innovation*, 2024, 24 (4): 1048-1065.
- [3] Zhouxin Yi, Xiu Luo. Construction Cost Estimation Model and Dynamic Management Control Analysis Based on Artificial Intelligence. *Iranian Journal of Science and Technology, Transaction of Civil Engineering*, 2024, 48(1): 577-588.
- [4] Ghoroghi, Mahyar, Ghoddousi, Parviz, Makui, Ahmad, et al. Integration of resource supply management and scheduling of construction projects using multi-objective whale optimization algorithm and NSGA-II. *Soft*

- computing: A fusion of foundations, methodologies and applications, 2024, 28(5): 3793-3811.
- [5] Bader Aldeen Almahameed, Majdi Bisharah. Applying Machine Learning and Particle Swarm Optimization for predictive modeling and cost optimization in construction project management. *Asian Journal of Civil Engineering*, 2023, 25(2): 1281-1294.
- [6] Van P B, Peansupap V. Confirmatory analysis on factors influencing the material management effectiveness construction projects. *Engineering Construction and Architectural Management*, 2024, 31 (6): 2536-2562.
- [7] Simon W, Munch S L, Kranker J L. Using principal component analysis to identify latent factors affecting cost and time overrun in public construction projects. *Engineering, Construction and Architectural Management*, 2024, 31 (6): 2415-2436.
- [8] Si J, Wan C, Yang C K. Self-Organizing Optimization of Construction Project Management Based on Building Information Modeling and Digital Technology. *Iranian Journal of Science and Technology, Transaction of Civil Engineering*, 2023, 47(6):4135-4143.
- [9] Lawal Y A, Abdul-Azeez I F, Olateju O I. Sustainable Project Management Practices and the Performance of Construction Companies. *Management Dynamics in the Knowledge Economy*, 2024, 12(3):302-320.
- [10] Zheng Ruiyan, Li Zhongfu, Li Long. Group technology empowering optimization of mixed-flow precast production in off-site construction. *Environmental Science and Pollution Research*, 2024, 31(8): 11781-11800.
- [11] Sena Senses, Mustafa Kumral. Trade-off between time and cost in project planning: a simulation-based optimization approach. 2024,100(2): 127-143.
- [12] Tran, Duc Hoc. Optimizing time-cost in generalized construction projects using multiple-objective social group optimization and multi-criteria decision-making methods. *Engineering construction & architectural management*, 2020, 27(9): 2287-2313.
- [13] Ghannad, Pedram, Lee, Yong-Cheol, Friedland, Carol J. Mult objective Optimization of Post disaster Reconstruction Processes for Ensuring Long-Term Socioeconomic Benefits. *Journal of management in engineering*, 2020, 36(4): 4020038.1-4020038.15.
- [14] Zhang, Lihui, Dai, Guyu, Zou, Xin. Robustness-based multi-objective optimization for repetitive projects under work continuity uncertainty. *Engineering construction & architectural management*, 2020, 27(10): 3095-3113.
- [15] Aladdin Alwisy, Ahmed Bouferguene, Mohamed Al-Hussein. Framework for target cost modelling in construction projects. *The international journal of construction management*, 2020, 20(2): 89-104.
- [16] Zohrehvandi, Shakib, Vanhoucke, Mario, Soltani, Roya. A reconfigurable model for implementation in the closing phase of a wind turbines project construction. 2020, 27(2): 502-524.
- [17] Emre Cevikcan, Yildiz Kose. Optimization of profitability and liquidity for residential projects under debt and equity financing. *Built environment project and asset management*, 2021, 11(2): 369-391.
- [18] Wellendorf A, Tichelmann P, Uhl J. Performance Analysis of a Dynamic Test Bench Based on a Linear Direct Drive. *Archives of Advanced Engineering Science*, 2023, 1(1):55-62.
- [19] Purohit J, Dave R. Leveraging Deep Learning Techniques to Obtain Efficacious Segmentation Results. *Archives of Advanced Engineering Science*, 2023, 1(1): 11-26.
- [20] Aryavalli S N G, Kumar G H. Futuristic Vigilance: Empowering Chipko Movement with Cyber-Savvy IoT to Safeguard Forests. *Archives of Advanced Engineering Science*, 2023, 1(8): 1-16.

Performance Evaluation of Efficient and Accurate Text Detection and Recognition in Natural Scenes Images Using EAST and OCR Fusion

Vishnu Kant Soni^{1*}, Vivek Shukla², S. R. Tandan³, Amit Pimpalkar⁴, Neetesh Kumar Nema⁵, Muskan Naik⁶

Department of Computer Science and Engineering, Dr. C. V. Raman University, Bilaspur, C.G. India^{1, 2, 5}

Department of Computer Science, Government R. V. R. S. Kanya Mahavidyalaya, Kawardha, C.G. India³

Department of Computer Science and Engineering (AIML)-Shri Ramdeobaba College of Engineering and Management, Ramdeobaba University, Nagpur, India⁴

Department of Computer Science and Engineering, Lakhmi Chand Institute of Technology, Bilaspur, C.G. India⁶

Abstract—Scene texts refer to arbitrary text found in images captured by cameras in real-world settings. The tasks of text detection and recognition are critical components of computer vision, with applications spanning scene understanding, information retrieval, robotics, and autonomous driving. Despite significant advancements in deep learning methods, achieving accurate text detection and recognition in complex images remains a formidable challenge for robust real-world applications. Several factors contribute to these challenges. First, the diversity of text shapes, fonts, colors, and styles complicates detection efforts. Second, the myriad combinations of characters, often with unstable attributes, make complete detection difficult, especially when background interruptions obscure character strokes and shapes. Finally, effective coordination of multiple sub-tasks in end-to-end learning is essential for success. This research aimed to tackle these challenges by enhancing text discriminative representation. This study focused on two interconnected problems: Scene Text Recognition (STR), which involves recognizing text from scene images, and Scene Text Detection (STD), which entails simultaneously detecting and recognizing multiple texts within those images. This research focuses on implementing and evaluating the Efficient and Accurate Scene Text Detector (EAST) algorithm for text detection and recognition in natural scene images. The study aims to compare the performance of three prominent Optical Character Recognition (OCR) techniques—TesseractOCR, PaddleOCR, and EasyOCR. The EAST model was applied to a series of sample test images, and the results were visually represented with bounding boxes highlighting the detected text regions. The inference times for each image were recorded, highlighting the algorithm's efficiency, with average times of 0.446, 0.439, and 0.440 seconds for the respective test images. These results indicate that the EAST algorithm is accurate and operates in real-time, making it suitable for applications requiring immediate text recognition.

Keywords—Scene text recognition; optical character recognition; deep learning; feature extraction; scene text detection

I. INTRODUCTION

Smartphones' widespread adoption has revolutionized how we capture and share images. With their ease of use and quick accessibility, smartphones have led to an exponential growth in the amount of multimedia data available on the web. From

advertisements and holiday pictures to business cards and newspaper articles, these devices have made digitizing content a common practice. However, this abundance of data has also presented new challenges [1-2].

Natural scenes, characterized by diverse backgrounds, lighting conditions, and complex visual elements, are particularly challenging for computers to analyze and understand. Segmenting and extracting text from these scenes is crucial due to the practical value of embedded textual information. Text extraction enables humans and computers to interpret and utilize this data for various applications, such as document analysis, license plate recognition, and product identification. It enhances automation and efficiency in diverse domains, offering several advantages in real-time scenarios. In autonomous vehicles, efficient text extraction enables the recognition of road signs, enhancing navigation and safety. In retail environments, it aids in product identification and inventory management, streamlining operations, and improving customer service. Text extraction automates scanning and digitization processes in document analysis, increasing productivity and accuracy. Real-time text extraction provides a competitive edge in various industries, such as healthcare, where it can assist in patient data analysis and diagnosis, leading to faster and more accurate decisions. In finance, it enhances fraud detection and document processing, improving security and operational efficiency; digital forensics aids in analyzing textual information from crime scenes, supporting investigations, and collecting evidence [3].

This manuscript explores text detection approaches to address the challenges of mining and retrieving weakly structured content in scene images. By utilizing models like EAST and integrating OCR techniques, the research aims to develop the next generation of search engines capable of accurately identifying and reading text in diverse environments. Overcoming the limitations of current models is crucial for enabling machines to understand and interact with the world, ultimately driving advancements in applications such as autonomous driving, augmented reality, and content retrieval. The segmentation and extraction of text from natural scenes are pivotal for unlocking valuable information embedded in visual content. By enabling real-time text

*Corresponding Author

extraction, businesses, and industries can utilize this data for enhanced decision-making, automation of processes, and improved efficiency across a wide range of applications, underscoring the critical role of text detection and recognition technologies in modern-day scenarios.

The following are the novelties of the research:

1) *Real-time performance evaluation:* The research highlights the EAST algorithm's efficiency, demonstrating low inference times for text recognition, making it suitable for real-time applications.

2) *Integration of multiple OCR techniques:* The study uniquely combines TesseractOCR, PaddleOCR, and EasyOCR with the EAST algorithm, providing a comprehensive comparison of their performance in STD.

3) *Visual validation of results:* Using bounding boxes to represent detected text visually enhances the understanding of the algorithm's accuracy and effectiveness.

The remainder of the paper was structured to provide a comprehensive overview of the research in Section II. Section III presented a detailed description of the proposed scheme, outlining its methodologies and innovations. In Section IV, the authors showcased and analyzed the experimental results, highlighting the performance and effectiveness of their approach. This section engaged in a thoughtful discussion of the findings, considering their implications and potential applications. Finally, Section V offered a conclusion, summarizing the key contributions of the study and suggesting directions for future research in the field.

II. RELATED WORK

In recent years, rapid advancements in deep learning have revolutionized the field of STD. Researchers have proposed a flurry of novel algorithms based on neural networks, each making significant strides in this domain. By utilizing the power of convolutional neural networks (CNNs), these methods have automated learning text features, eliminating the need for manual feature engineering. This breakthrough has propelled STD technology to new heights [4]. Numerous researchers have explored various techniques for detecting text in images, contributing significantly to advancements in this field. Some investigators concentrated on texture-based approaches, utilizing the sliding window concept to identify and analyze unique textural features within input images. This method effectively localizes text information by examining patterns that distinguish text from the surrounding background. Other researchers focused on sparse-based text detection methods, which have proven beneficial for various computer vision applications. These techniques leverage sparse representations to enhance text detection, particularly in challenging environments where traditional methods may struggle. By employing these innovative approaches, researchers aimed to improve the accuracy and reliability of text detection systems, paving the way for more robust applications in real-world scenarios [5].

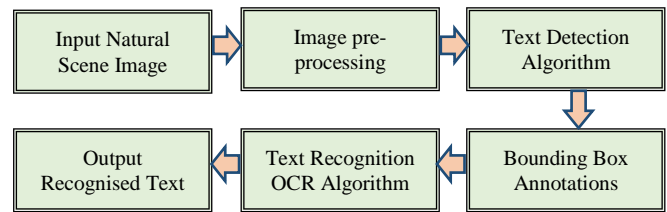


Fig. 1. Pipeline of text detection and extraction.

The pipeline, as illustrated in Fig. 1, consists of six key steps: (1) Input Natural Scene Image, (2) Image pre-processing, (3) Text Detection Algorithm (EAST), (4) Bounding Box Annotations, (5) Text Recognition OCR Algorithm, and (6) Output Recognized Text. By applying this comprehensive approach, the research aims to achieve accurate and efficient text detection and recognition in real-world scenarios, contributing to advancing intelligent systems capable of understanding and interacting with textual information in diverse environments.

Current deep learning-based STD approaches can be broadly categorized into two main groups: regression-based methods and segmentation-based methods. Regression-based techniques typically employ CNNs to directly predict text regions' bounding boxes or coordinates. These models learn to map input images to predefined anchors or text proposals, refined and filtered to obtain the final text detections. One notable example of a regression-based method is TextBoxes, which adapts the Single Shot MultiBox Detector architecture for STD, achieving real-time performance while maintaining high accuracy. On the other hand, segmentation-based methods treat text detection as a pixel-wise classification problem [6]. These algorithms divide the input image into a grid of cells and predict whether each cell contains text. By leveraging the inherent strengths of CNNs in semantic segmentation, segmentation-based approaches can handle text instances of arbitrary shapes and orientations. A prominent example is the EAST, which employs a fully convolutional network (FCN) to generate a score map and geometry of text boxes, enabling text detection in various orientations and scales [7-8]. Both regression-based and segmentation-based methods have their advantages and disadvantages. Regression-based techniques often excel in computational efficiency, making them suitable for real-time applications. However, they may struggle with detecting text instances of complex shapes or orientations. Segmentation-based methods, on the other hand, demonstrate superior performance in handling diverse text geometries but may require more computational resources. Despite the remarkable progress made by deep learning-based STD algorithms, challenges remain. Factors such as complex backgrounds, varying lighting conditions, and text distortions can still hinder the accuracy of these models. Ongoing research efforts aim to address these limitations and further enhance the robustness and applicability of STD systems in real-world scenarios [9].

STD was recognized as a complex and challenging task due to various environmental factors, including illumination, lighting conditions, and the presence of small or curved text. Many existing approaches prioritized model accuracy and efficiency but resulted in heavy-weight models requiring

substantial processing resources. STR emerged as a prominent research area in computer vision, focusing on recognizing text in natural scenes. Researchers noted that attention-based encoder-decoder frameworks struggled with attention drift, which hindered the precise alignment of feature regions with target objects in complex, low-quality images. Additionally, the rise of Transformer models led to increased computational costs due to their larger parameter sizes. X. Luan et al. [10] developed the lightweight STR model to address these issues, incorporating a position-enhancement branch to alleviate attention drift and dynamically fuse position with visual information. Experimental results indicated that lightweight STR achieved a 3% higher average recognition accuracy than baseline models while maintaining a lightweight structure with only seven million parameters. This balance of accuracy, speed, and computational efficiency made lightweight STR suitable for high-demand applications in STR, outperforming existing methods.

Researchers in [11-12] developed a novel lightweight model to enhance the accuracy and efficiency of STD. This model utilized ResNet50 and MobileNetV2 as backbones, incorporating quantization techniques to reduce size. During quantization, the precision was adjusted from float32 to float16 and int8, resulting in a more lightweight model. The proposed method significantly outperformed state-of-the-art techniques, improving inference time and Floating-Point Operations Per Second (FLOPS) by approximately 30 to 100 times. The researchers used well-known datasets, ICDAR2015 and ICDAR2019, to validate the model's performance, and they included samples in ten different languages. The model demonstrated a balance of accuracy and efficiency, achieving word % accuracy rates of 62% for complex text and 80% for non-complex text and character accuracy rates of 68% and 88%, respectively. R. Harizi et al. [13] study introduced a hybrid scene text detector that combined selective search with SIFT-based key point density analysis and a deep learning training architecture. The researchers investigated key SIFT points to identify crucial image areas for precise word localization. They then fine-tuned these regions using a deep learning-powered bounding box regressor, which ensured accurate word boundary alignment and enhanced detection efficiency. The study focused on detecting text in real-world scene images. They proposed a method that integrated SIFT-based key point localization, Bag of Words-based character pattern filtering, and ResNet-19-based word bounding box regression. Experimental results confirmed the method's effectiveness in addressing multi-oriented and curved scene texts.

In their paper, G. Liao et al. [14] significantly contributed to STD. They designed a Multi-Pooling Module (MPM) with different pooling operations to address the limitations of the original PSENet. The MPM effectively captured the relevance of text information at various distances, enabling precise localization of scene text regions. Y. Cai et al. [15] proposed a style-aware learning network to achieve style-robust text detection in diverse environments. M. Lu et al. [16] addressed the existing model's deficiencies in detecting long text regions by altering the shrinkage calculation, adding a feature enhancement module, and changing the loss function to Focal

loss. S. Yuchen et al. [17] proposed a novel parameterized text shape method based on low-rank approximation, distinguishing their approach from other shape representation methods that relied on data-irrelevant parameterization. They utilized singular value decomposition to reconstruct text shapes using a limited number of eigenvectors derived from labeled text contours.

In a study, M. Aluri et al. [18] developed an innovative method for identifying irregular text in natural scene images. The approach combined a U-net architecture with connected component analysis, significantly improving text component detection accuracy while reducing non-character element identification. Furthermore, the researchers incorporated graph convolution networks to infer adjacency relations among text components, introducing a sophisticated mechanism that advanced text detection in natural scene images. In their novel approach, H. Chen et al. [19] developed the Fragmented Affinity Reasoning Network of Text Instances, a component connection method for arbitrary shape text detection. The network consisted of three key modules: the Weighted Feature Fusion Pyramid Network (WFFPN), Text Fragments Subgraph (TFS), and Dense Graph Attention Network (DGAT), which could be trained end-to-end. The researchers introduced WFFPN to generate text fragments, while TFS and DGAT jointly constructed an affinity reasoning network. Their contributions included proposing a novel unified end-to-end trainable framework, developing a simple and effective WFFPN for multi-scale feature representation and processing, and introducing the joint module of TFS and DGAT to infer the link relationship between text fragments, improving the grouping performance of dense and long curved text.

In their work, Y. Zhu et al. [20] proposed a novel STD method called Text Mountain. The core concept of Text Mountain utilized border-center information differently than previous approaches, which treated center-border as a binary classification problem. Instead, they predicted text center-border probability (TCBP) and text center-direction (TCD). The TCBP resembled a mountain, with the peak representing the text center and the base indicating the text border, allowing for better separation of text instances. This method proved robust against multi-oriented and curved text due to its effective labeling rules. During inference, each pixel at the mountain base searched for a path to the peak, enabling efficient parallel processing. Experiments on various datasets, including MLT and ICDAR2015, demonstrated that Text Mountain achieved superior performance, notably an F-measure of 76.85% on MLT, surpassing previous methods significantly.

Current STD models encounter limitations that impact their effectiveness in real-world applications, mainly when dealing with scene text images and born-digital documents. These categories present unique challenges compared to traditional scanned paper documents. One significant difficulty is the presence of cluttered backgrounds. Existing models often struggle to accurately identify text amidst various visual elements, which can lead to false positives or missed detections. Additionally, traditional models typically use rigid geometrical shapes, like axis-aligned rectangles, making them less effective for detecting free-form text, such as curved or

rotated characters commonly found in natural environments. While some models attempt to manage variations in text size through multi-scale feature maps, this approach can be complex and computationally demanding. The need for elaborate post-processing steps can slow down detection and complicate model architecture. Lighting conditions also play a significant role, as many models perform well under controlled environments but falter in outdoor or dynamically lit situations [21]. Finally, balancing detection accuracy and real-time processing speed remains a critical challenge. Many advanced models sacrifice speed for improved accuracy, rendering them unsuitable for applications that require immediate results. Addressing these limitations is vital for enhancing the robustness and applicability of text detection systems.

III. PROPOSED METHODOLOGY

Text detection involves predicting and localizing text instances within images. While traditional image processing techniques were commonly used for this task, deep learning models consistently outperformed them across various real-world scenarios, from simple to highly complex environments. The localization of text using deep learning could be achieved primarily through two approaches: object detection and image segmentation. Object detection methods focused on identifying

and bounding text regions, providing a straightforward way to localize text. In contrast, image segmentation treated text detection as a pixel-wise classification task, allowing for more precise delineation of text shapes. Each approach had advantages and challenges concerning dataset creation, model training, and inference options. The advancements in deep learning significantly enhanced the effectiveness of text detection in various applications. Object detection techniques localize objects within an image by drawing rectangular or square bounding boxes around them. While effective, this method provides limited information about the actual shape of the detected objects. Fortunately, labeling images for object detection is a relatively straightforward process compared to segmentation. Segmentation, [22] conversely, involves classifying each pixel in an image into predefined categories.

Segmentation would entail distinguishing between text and non-text pixels in scene detection. This pixel-wise classification allows for identifying text regions with greater precision, even if they exhibit complex shapes or orientations. For character recognition tasks, the annotation process becomes even more granular. Each pixel is classified as belonging to one of the available character classes, enabling the precise identification of individual letters or symbols within the detected text regions. This process can be visualized in Fig. 2.

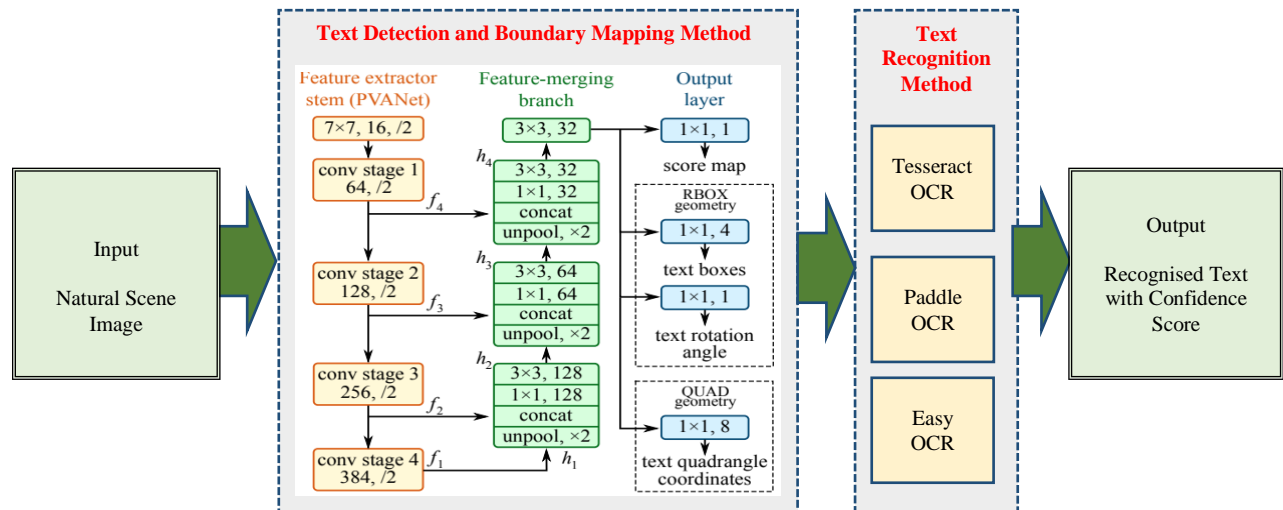


Fig. 2. The structure of the EAST text detection fully convolutional network.

The EAST algorithm was explicitly developed [23] to address the challenges of text detection in natural scenes, where text can appear in diverse sizes, orientations, and perspectives. The EAST architecture was designed to handle text regions of varying sizes efficiently. The key idea was to leverage features from different neural network stages: later stages for detecting large and initial stages for small word regions. The authors employed three interconnected branches within a single neural network. The fundamental principles underlying EAST's functionality include several innovative components. The Feature Extractor Stem was responsible for extracting features from various network layers. This stem could be a convolutional network pre-trained on the ImageNet dataset, such as PVANet, VGG16, and Resnet V1-50—the model, taking outputs from the pooling layers. This network is typically pre-trained on extensive datasets and subsequently

fine-tuned for the specific task of text detection, allowing it to learn the unique characteristics of text in various contexts effectively. The Feature Merging Branch combined the feature outputs from different VGG16 layers and can be expressed in Eq. (1) and Eq. (2).

$$g_i = \begin{cases} \text{unpool}(h_i) & \text{if } i \leq 3 \\ \text{conv}_{3 \times 3}(h_i) & \text{if } i = 4 \end{cases} \quad (1)$$

$$h_i = \begin{cases} f_i & \text{if } i = 1 \\ \text{conv}_{3 \times 3}(\text{conv}_{1 \times 1}[g_{i-1}; f_i]) & \text{otherwise} \end{cases} \quad (2)$$

EAST utilized a U-Net-like architecture to merge the feature maps to avoid computational complexity gradually. The process involved upsampling the $pool_{n-1}$ layer output to match the size of the $pool_n$ layer output, concatenating them, and applying convolutional layers to fuse the information. This

procedure was repeated for the remaining layers, ultimately producing a final feature map layer before the output layer. EAST employs anchors and a Region Proposal Network (RPN) to propose potential text regions. However, it customizes the RPN to predict axis-aligned quadrilaterals instead of traditional rectangles, enabling it to enclose text regions more accurately and tightly. The Output Layer consisted of two key components: a score map and a geometry map. The score map indicated the probability of text in each region, while the geometry map defined the boundaries of the text boxes. EAST offered two options for the geometry map: rotated boxes (specified by top-left coordinate, width, height, and rotation angle) or quadrangles (all four coordinates of a rectangle). EAST predicts the coordinates of the four vertices of each quadrilateral bounding a text region, along with a confidence score that indicates the likelihood of text presence. This capability allows the algorithm to manage text in arbitrary orientations and shapes, enhancing its versatility in real-world applications.

In text detection, bounding box annotations mark the regions in images where text appears. These annotations help train the EAST algorithm to recognize and locate text in various scenes. For instance, each bounding box outlines the area containing text, which the algorithm learns to identify. The process of bounding box annotations for text regions using the EAST algorithm involved several vital steps that aimed to enhance the accuracy of text detection in images. Initially, the EAST algorithm utilized an FCN to analyze input images and generate a score map, indicating the likelihood of text presence across different image areas. The EAST algorithm first predicted the geometry of potential text regions to create bounding box annotations. This was achieved by estimating four parameters for each pixel in the score map: the bounding box's height and width and the center coordinates. The model could effectively capture the spatial characteristics of text instances in various orientations and scales by employing a regression approach. The EAST text detector model generated two key outputs: scores, which represented the probabilities of positive text regions, and geometry, which provided the bounding boxes for these text regions. These outputs served as parameters for the decode prediction's function, which processed the input data. The function returned a tuple containing the bounding box locations of the detected text and their corresponding probabilities. The bounding boxes, referred to as "reacts," were formatted compactly for efficient application of Non-Maximum Suppression (NMS), while the "confidences" represented the confidence values associated with each bounding box. Once the score map and geometry predictions were generated, the next step involved applying NMS to filter out overlapping bounding boxes. This technique helped to eliminate redundant detections, ensuring that only the most confident predictions remained.

The NMS algorithm selected the bounding box with the highest score and removed any boxes with significant overlap based on a predefined threshold. As an FCN, EAST outputs per-pixel predictions of words or text lines and utilizes NMS as a post-processing step on the geometric map. This geometric map can be RBOX (with four channels for bounding box coordinates and one for text rotation) or QUAD (with eight

channels representing shifts from the four corner vertices). EAST employs a weighted sum of losses for both the score map and the geometry, ensuring adequate training. The resulting bounding boxes were then refined to improve their accuracy. This included adjusting the boxes' dimensions to fit better the actual text regions detected in the image. The final output consisted of well-defined bounding boxes that accurately represented the locations of text instances. The EAST algorithm's approach to bounding box annotations combined advanced deep learning techniques with effective post-processing methods, resulting in a robust framework for detecting text regions in natural scenes. This process significantly improved the performance of STD, making it a valuable tool for various applications, such as document analysis and autonomous navigation [24]. By integrating these three branches, the EAST architecture effectively handled text regions of varying sizes and shapes, making it a powerful tool for STD. The author's innovative approach to feature extraction and merging, combined with the informative output layers, contributed to EAST's efficiency and accuracy in detecting text in complex scenes. EAST optimizes its performance by minimizing two key loss functions during training: the classification loss, which determines the presence of text, and the regression loss, which refines the predicted text regions.

The classification loss can be expressed as in Eq. (3).

$$L_{cls} = -\frac{1}{N_{cls}} \sum_{i=1}^{N_{cls}} [g_i \log(p_i) + (1 + g_i) \log(1 + p_i)] \quad (3)$$

Where N_{cls} denotes the number of anchor regions used for classification. The classification loss in EAST measures the model's ability to distinguish text regions from non-text areas. It is calculated using cross-entropy loss, where p_i represents the predicted probability of the i -th region containing text and g_i is the ground truth label (1 for text, 0 for non-text). Minimizing this loss helps the model accurately classify text regions.

The regression loss can be expressed as in the Eq. (4).

$$L_{reg} = \frac{1}{N_{reg}} \sum_{i=1}^{N_{reg}} Smooth_{L1}(d_i - g_i) \quad (4)$$

Here N_{reg} represents the number of anchor regions used for regression and $Smooth_{L1}$ is the smoothing loss function. EAST employs a regression loss to evaluate how accurately the network predicts the quadrilateral coordinates of text regions. It utilizes $Smooth_{L1}$ loss, which compares the predicted geometry parameters. For each region, such as the distances from the anchor point to the four vertices of the quadrilateral, with the ground truth geometry parameters. This loss function ensures the network learns to generate tight, accurate bounding boxes around text areas, enabling precise text detection.

EAST-OCR Fusion Algorithm

Input: Natural Scene Image (I) from ICDAR 2013, ICDAR 2015, COCO-Text

Output: Recognized Text String (T), Confidence Score (C)

Step-I: Pre-processing (Pre-process (I))

i. $I = \text{resize}(I, (224, 224))$

- ii. *Grayscale Conversion:* $I_{gray} = \text{rgb2gray}(I)$
 - iii. *Noise Remove:* $I_{denoised} = \text{median_filter}(I_{gray})$
 - iv. *Normalization:*

$$I_{norm} = \frac{(I_{denoised}(score) - \min(I_{denoised}(score)))}{\max(I_{denoised}(score)) - \min(I_{denoised}(score))}$$
- Step-II: Text Detection ($Text_{Regions} = \text{DetectText}(I_{norm})$)
- i. *Text RegionDetection:* Apply the EAST algorithm to detect text regions in I_{norm}
 - ii. *Bounding Box Extraction:*
 - a. Calculate confidence (D), coordinates (C), and rotation angle (θ) using a 1D vector:
 - 1D vector = $\text{Conv1D}(\text{output})$
 - b. Use Non-Maximum Suppression (NMS) to refine bounding boxes:
 - final bounding box = $\text{NMS}(\text{start}_x, \text{start}_y, \text{end}_x, \text{end}_y)$
- Step-III: Text Segmentation ($\text{Segmentation}(I_{norm}, Text_{Regions})$)
- i. *Check for Detected Regions:*
 - a. If $Text_{Regions}$ is empty:
 - Segment I_{norm} into individual characters using connected component analysis.
 - b. Else
 - Segment each region in $Text_{Regions}$ into individual characters.
 - ii. Apply heuristics filtering to discard non-text regions based on size and aspect ratio.
- Step-IV: Feature Extraction ($\text{Features} = \text{FeatureExtractor}(\text{Character}_{Images})$)
- i. For each character image (c):
 - a. Extract features (f_c):
 - HOG features: $f_c = \text{hogfeature}(c)$
 - Binary Image: $f_c = \text{imbinarize}(c)$
 - b. Return a list of features (Features) for all characters
- Step-V: Text Recognition with OCR Methods ($\text{Text}_{Sequence}, \text{Confidence}_{Score} = \text{OCR}_{recog}(\text{Features})$)
- i. Apply an OCR network with an embedding layer, OCR layers, and a softmax output layer.
 - ii. For each feature vector (f_i) in Features :
 - a. Predict character probability distribution using $p(c|f_i) = \text{softmax}(\text{OCR}(f_i))$
 - b. Decode the predicted character sequence ($\text{Text}_{Sequence}$)
 - iii. Calculate the confidence score (C) where C_{ij} represents the probability of character j being at position i in the sequence.
- Step-VI: Post-processing ($\text{Text}_{Refined} = \text{Postprocess}(\text{Text}_{Sequence})$)
- i. Implement proofreading steps to enhance text quality, including spell-checking
- Step-VII: Output Display ($\text{Display}(\text{Text}_{Refined}, C)$)

The EAST-OCR fusion algorithm for text detection and recognition in natural scene images follows a structured approach. It begins with pre-processing the input image, which includes resizing, grayscale conversion, noise removal, and normalization. Next, the EAST algorithm detects text regions, calculating confidence scores, coordinates, and rotation angles while applying NMS to refine bounding boxes. Text segmentation is performed based on detected regions, followed by feature extraction from individual character images. The extracted features are then processed through an OCR network to recognize the text and compute confidence scores. Finally, post-processing steps enhance text quality, and the results, including the recognized text and confidence scores, are displayed to the user. The algorithm outlines a structured approach, ensuring clarity and comprehensiveness in each step.

TABLE I. DATASET STATISTICS

Parameter	Value
Dataset Names	ICDAR 2013, ICDAR 2015, COCO-Text
Total Images	65,598
Total Bounding Boxes	5,000
Average Bounding Boxes per Image	5
Total Text Instances	1,50,359
Text Instances Categories	machine-printed and handwritten text
Text Instances Language Categories	English script and non-English script
Training Set Size	70%
Validation Set Size	15%
Testing Set Size	15%

The EAST model was primarily trained using ICDAR 2013, ICDAR 2015 and COCO-Text datasets, which provided various text instances for effective learning. This dataset's statistics can be seen in Table I. Additionally, the model utilized the ResNet V1-50 architecture, sourced from Tensor Flow, instead of the alternative PVANet, to enhance feature extraction capabilities. For optimization, we opted for loss, which focuses on maximizing the Intersection over Union (IoU) of segmentation rather than using balanced cross-entropy loss. Furthermore, a linear learning rate decay strategy was implemented instead of a staged learning rate decay approach, allowing for smoother convergence during training. These choices contributed to the model's improved performance in detecting text in natural scenes. The dataset comprised 4,500 unique text instances, offering diverse content that enhances the model's learning experience. The dataset statistics for bounding box annotations used in training the EAST algorithm were meticulously compiled to enhance the model's ability to detect text in natural scenes. The dataset included images with diverse text instances annotated with bounding boxes to indicate the precise locations of text regions.

Despite its complexity and the significant computational resources required for implementation, EAST has proven to be a powerful tool for various applications, including OCR, text recognition, and image information extraction. Ultimately, EAST's ability to accurately and efficiently locate and interpret text within images has established it as a crucial component in

computer vision and OCR. Its contributions have significantly advanced the development of applications capable of understanding and processing textual information in the world around us. Following the implementation of the text detection and boundary mapping method, the next crucial step in this research was the actual detection of text within the images. Three different OCR techniques were employed: TesseractOCR, PaddleOCR, and EasyOCR. Each method was chosen for its unique advantages, allowing for a comprehensive comparison of their performance in text detection tasks. Tesseract OCR is one of the most widely used OCR engines, known for its robustness and flexibility. It supports multiple languages and has a strong community backing, contributing to its continuous improvement. Tesseract excels in recognizing printed text and has been optimized for various applications, making it a reliable choice for this research.

PaddleOCR is another powerful OCR tool that stands out for its multilingual capabilities and high accuracy. It integrates advanced deep learning techniques to handle complex text scenarios, including curved and multi-oriented text. PaddleOCR is particularly beneficial for tasks requiring high precision in text extraction from natural scenes. EasyOCR is a newer entrant in the OCR landscape, gaining popularity for its simplicity and effectiveness. It supports over 80 languages and is designed to be easy to use. EasyOCR uses deep learning models to achieve impressive text detection and recognition results, particularly in challenging environments [25]. By applying these three OCR techniques, the research aimed to evaluate their effectiveness in detecting text across various

scenarios. Each method was assessed based on accuracy, speed, and adaptability to different text orientations and backgrounds. This comparative analysis highlighted the strengths and weaknesses of each OCR tool and provided valuable insights into its suitability for specific text detection tasks. Ultimately, the findings from this research could guide future developments in selecting the most appropriate OCR technology for their needs.

IV. RESULT AND DISCUSSION

After implementing the EAST algorithm on a series of sample test images, the next step was recognizing the text in these images. The results of this process are illustrated in the accompanying Fig. 3: (a1-a3) display the sample testing images. At the same time (b1-b3), the corresponding text detection results are shown, complete with bounding boxes around the detected text regions. The performance of the text recognition was evaluated based on the inference time for each test image, which was recorded as 0.446 seconds for the first image, 0.439 seconds for the second, and 0.440 seconds for the third. These results indicate that the EAST algorithm is highly efficient, demonstrating a low inference time for text recognition across the sample images. This efficiency is particularly noteworthy, as it suggests that the EAST algorithm can effectively detect and recognize text in real-time scenarios, making it suitable for applications where speed is critical. The bounding box in the detection results visually confirms the text recognition's accuracy, showcasing the algorithm's capability to identify text in various contexts.

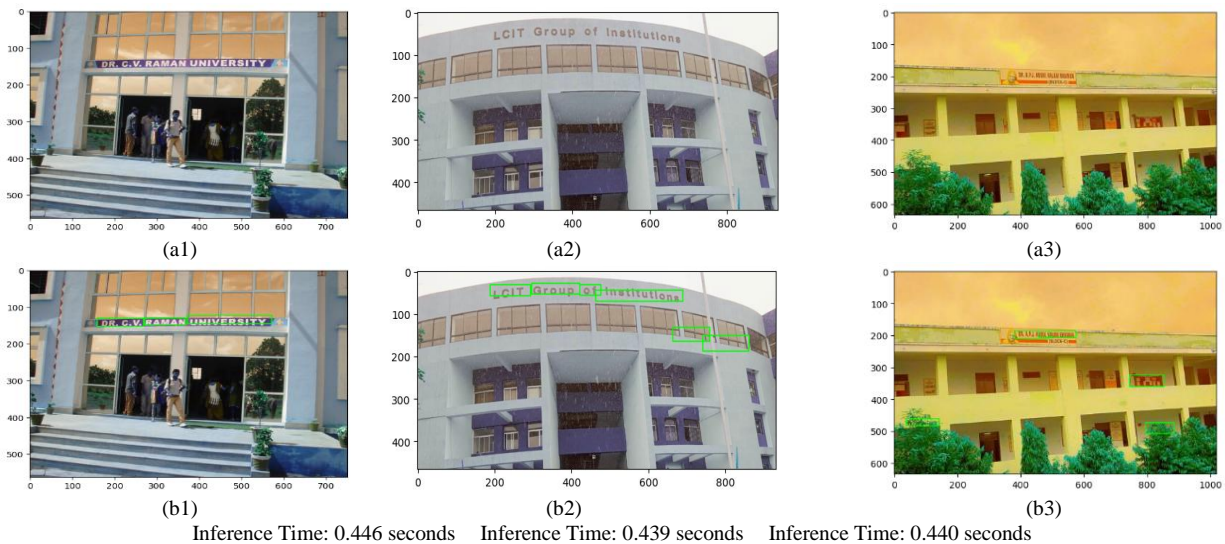


Fig. 3. (a1-a3): Sample testing images, (b1-b3) Text detection results with bounded box.

TABLE II. COMPARISON OF DIFFERENT TEXT RECOGNITION METHODS

Method/ Actual Text	DR C. V. RAMAN UNIVERSITY	LCIT Group of Institutions	DR. A.P.J. ABDUL KALAM BHAWAN (BLOCK-C)	Average Confidence Score
Easy OCR	DR CV RAMAN UNIVERSITY (Confidence: 0.86)	LCIT (Confidence: 0.98) Group (Confidence: 1.00) 6 f (Confidence: 0.58) Institulone (Confidence: 0.76)	DR.A,PJ, ABDUL KALAM BHAWAN (Confidence: 0.89) (BLOCK-C) (Confidence: 0.91)	0.85
Tesseract OCR	DR C.V. RAMAN UNIVERSITY (Confidence: 0.83)	LCIT, (Confidence: 86.00) Gr, (Confidence: 95.00)	DR.A,PJ, ABDUL KALAM BHAWAN (Confidence: 0.87) (BLOCK-C) (Confidence: 0.96)	0.89
Paddle OCR	DR.C.V.RAMAN UNIVERSITY (Confidence: 0.96)	LCIT (Confidence: 0.98) Group (Confidence: 1.00) of (Confidence: 0.78) Institution (Confidence: 0.89)	DR.A.P.J.ABOUL KALAM BHAWAN (Confidence: 0.92) BLOCK-C (Confidence: 0.98)	0.93

In this work, a comparison was conducted among three prominent text recognition methods: EasyOCR, Tesseract OCR, and PaddleOCR. Each method was evaluated on sample test images to determine their effectiveness in accurately recognizing text. As shown in Table II and Fig. 5, the results revealed average confidence scores of 0.85 for EasyOCR, 0.89 for Tesseract OCR, and an impressive 0.93 for PaddleOCR. These scores indicate that PaddleOCR outperformed the other two methods, demonstrating its superior capability in text recognition tasks. The higher confidence score suggests that PaddleOCR detected text more accurately and effectively handled various text styles and orientations.

Inference Time (In seconds)

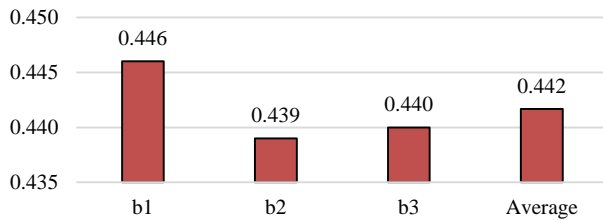


Fig. 4. Comparison between text detection inference times for test images.

While EasyOCR and Tesseract OCR also provided commendable performance, PaddleOCR's results highlight its strengths, particularly in complex scenarios where text may be distorted or presented in challenging conditions.

Average Confidence Score

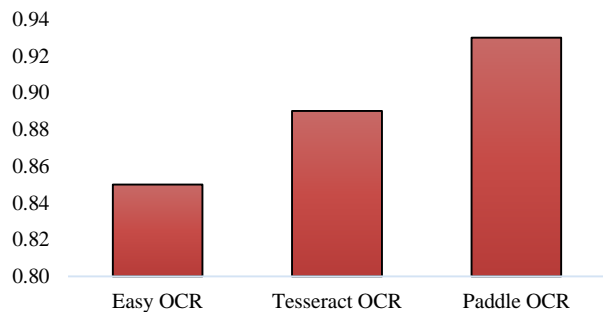


Fig. 5. Comparison of average confident score between different OCR methods for test images.

This comparison underscores the importance of selecting the right OCR tool for specific applications, especially when accuracy is paramount. Overall, PaddleOCR stands out as a robust choice for text recognition, making it an evaluable asset for future projects requiring reliable OCR capabilities. The proposed method utilizing the EAST algorithm for text detection and recognition offers several advantages over previous approaches. Firstly, it streamlines the process by employing a single neural network that directly predicts text instances and their geometries, eliminating the need for time-consuming intermediate steps such as candidate proposal and word partitioning. This end-to-end approach enhances speed, as shown in Fig. 4, and improves accuracy, allowing for near real-time processing of images. The EAST algorithm is also designed to handle text in various orientations and aspect ratios, addressing a standard limitation in traditional OCR methods that struggle with diverse text layouts. By outputting dense per-pixel predictions, EAST provides more precise text region localization than earlier models. Moreover, while previous OCR techniques may falter with underrepresented languages or complex scripts, the EAST framework's flexibility allows for better adaptation to different text types. Integrating advanced OCR methods like Tesseract or PaddleOCR further enhances recognition accuracy, particularly in challenging scenarios. The proposed method effectively resolves speed, accuracy, and adaptability issues found in earlier approaches, making it a robust solution for efficient text detection and recognition in natural scene images.

V. CONCLUSION AND FUTURE SCOPE

Integrating the EAST algorithm with various OCR techniques has demonstrated promising results in enhancing STD and recognition performance. By applying Tesseract OCR, PaddleOCR, and EasyOCR to the sample test images, this research has highlighted the strengths and limitations of each method. The EAST algorithm has proven to be a highly efficient and accurate tool for text detection, as evidenced by the low inference times recorded during the testing process. With an average inference time of less than half a second per image, the EAST algorithm's real-time capabilities make it suitable for applications that require immediate text recognition, such as autonomous vehicles and augmented reality systems. Moreover, the visual representation of the text detection results, showcased through bounding boxes, confirms the accuracy of the EAST algorithm in identifying text regions within the sample images. The comparative analysis of the

OCR techniques revealed distinct strengths and weaknesses. Tesseract OCR demonstrated robustness in recognizing printed text, while PaddleOCR excelled in handling multilingual text and complex layouts. EasyOCR, known for its user-friendly interface, provided quick results with impressive accuracy. The findings underscore the potential of the EAST algorithm as a reliable tool for STD, particularly in dynamic environments where speed and accuracy are paramount. The visual confirmation of the detection results and the efficient inference times highlight the algorithm's ability to identify text in various contexts effectively. Overall, the EAST algorithm's performance in these tests highlights its potential as a reliable tool for STD and recognition in diverse environments.

The proposed research presents some limitations that future studies could address. Firstly, combining multiple OCR techniques may introduce inconsistencies in performance evaluation and output reliability. Although PaddleOCR is recognized for its multilingual capabilities, it may not sufficiently support underrepresented languages, non-English scripts, symbols, or complex scripts. Additionally, font size, style, and orientation variations can lead to OCR output errors. Moreover, the findings may not generalize well across different domains; performance could vary significantly between document and natural scene images or across diverse geographical locations. Future research could benefit from incorporating advanced methods such as Transformers and Vision Language Models, which may improve the handling of complex text detection scenarios. Exploring the integration of the EAST algorithm with advanced transfer learning techniques could enhance its robustness against challenging backgrounds, varying lighting conditions, and diverse text orientations. Emphasizing multilingual capabilities would allow for a more comprehensive evaluation of text detection across various languages, addressing a critical need in diverse environments. By building on the insights from this study, advancements in text detection and recognition can lead to the development of more intelligent systems capable of effectively interacting with textual information in real-world applications.

REFERENCES

- [1] C. Luo, L. Jin, and Z. Sun, "MORAN: A Multi-Object Rectified Attention Network for Scene Text Recognition," *Pattern Recognition*, 2019.
- [2] J. Ghosh, A. Talukdar, and K. Sarma, "A lightweight natural scene text detection and recognition system," *Multimedia Tools and Applications*, 2023.
- [3] P. Naveen, and M. Hassaballah, "Scene text detection using structured information and an end-to-end trainable generative adversarial networks," *Pattern Analysis and Applications*, 2024.
- [4] R. Pegah, "Deep Learning Techniques for the Analysis of Soccer Matches," *Budapest University of Technology and Economics (Hungary)*, 2024.
- [5] M. Kantipudi, S. Kumar, and A. K. Jha, "Scene Text Recognition Based on Bidirectional LSTM and Deep Neural Network," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 2676780, pp. 1-11, 2021.
- [6] Y. Yuwei, L. Yuxin, Z. Zixu, and T. Minglei, "Arbitrary-Shaped Text Detection with B-Spline Curve Network," *Sensors*, 2023.
- [7] Z. Hu, X. Wu, and J. Yang, "TCATD: Text Contour Attention for Scene Text Detection," In *25th International Conference on Pattern Recognition (ICPR)*, 2021.
- [8] P. Cheng, and W. Wang, "A Multi-Oriented Scene Text Detector with Position-Sensitive Segmentation," In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval - ICMR '18*, 2018.
- [9] M. Ibrayim, Y. Li, and A. Hamdulla, "Scene Text Detection Based on Two-Branch Feature Extraction," *Sensors*, Basel, Switzerland, vol. 22, 16-6262, 2022.
- [10] X. Luan, J. Zhang, M. Xu, W. Silamu, Y. Li, "Lightweight Scene Text Recognition Based on Transformer," *Sensors (Basel)*, vol. 5; 23(9):4490, 2023.
- [11] K. Manjari, M. Verma, G. Singal, and S. Namasudra, "QEST: Quantized and Efficient Scene Text Detector using Deep Learning," *Association for Computing Machinery Asian and Low-Resource Language Information Processing*, pp. 1-18, 2022.
- [12] A. Dass, S. Srivastava, M. Gupta, M. Khari, J. P. Fuente, and E. Verdú, "Modelling and control of systems using intelligent water drop algorithm," *Expert Systems*, 2022.
- [13] R. Harizi, R. Walha, and F. Drira, "SIFT-ResNet Synergy for Accurate Scene Word Detection in Complex Scenarios," In *International Conference on Agents and Artificial Intelligence (ICAART)*, SCITEPRESS – Science and Technology Publications, Lda., vol 3, pp. 980-987, 2024.
- [14] G. Liao, Z. Zhu, Y. Bai, et al., "PSENet-based efficient scene text detection," *EURASIP Journal on Advances in Signal Processing*, vol. 97, 2021.
- [15] Y. Cai, F. Zhou, and R. Yin, "Exploring Style-Robust Scene Text Detection via Style-Aware Learning," *Electronics*, vol. 13(2):243, 2024.
- [16] M. Lu, Y. Mou, C-L. Chen, and Q. Tang, "An Efficient Text Detection Model for Street Signs," *Applied Sciences*, vol. 11(13):5962, 2021.
- [17] S. Yuchen, C. Zhineng, et al., "LRANet: Towards Accurate and Efficient Scene Text Detection with Low-Rank Approximation Network," *arXiv:2306.15142v5*, 2024.
- [18] M. Aluri and U.D. Tatavarthi, "Geometric deep learning for enhancing irregular scene text detection," *Revue d'Intelligence Artificielle*, Vol. 38, No. 1, pp. 115-125, 2024.
- [19] H. Chen, P. Chen, Y. Qiu, N. Ling, "FARNet: Fragmented affinity reasoning network of text instances for arbitrary shape text detection," *IET Image Process.* Vol. 17, pp. 1959–1977, 2023.
- [20] Y. Zhu and J. Du, "TextMountain: Accurate scene text detection via instance segmentation," *Pattern Recognition*, vol. 110 (2021) 107336, pp. 1-11, 2020.
- [21] B. A. Abubaker, J. Razmara, and J. Karimpour, "A Novel Approach for Target Attraction and Obstacle Avoidance of a Mobile Robot in Unknown Environments Using a Customized Spiking Neural Network," *Applied Sciences*, 2023.
- [22] L. Nandanwar, P. Shivakumara, R. Ramachandra, T. Lu, U. Pal, A. Antonacopoulos, and Y. Lu, "A New Deep Wavefront based Model for Text Localization in 3D Video," *IEEE Transactions on Circuits and Systems for Video Technology*, 2021.
- [23] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, "EAST: An Efficient and Accurate Scene Text Detector," In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 2642-2651, 2017.
- [24] M. K. Ebrahimi, H. Lee, J. Won, S. Kim, and S. S. Park, "Estimation of soil texture by fusion of near infrared spectroscopy and image data based on convolutional neural network," *Computers and Electronics in Agriculture*, 2023.
- [25] B. Myint, T. Onizuka, P. Tin, M. Aikawa, I. Kobayashi, and T. Zin, "Development of a real-time cattle lameness detection system using a single side-view camera," *Scientific Reports*, 2024.

AI-Powered Learning Pathways: Personalized Learning and Dynamic Assessments

Mohammad Abrar^{1*}, Walid Aboraya², Rawad Abdel Khaliq³,
Kabali P Subramanian⁴, Yousuf Al Husaini⁵, Mohammed Al Hussaini⁶
Faculty of Computer Studies, Arab Open University, Muscat 122, Oman^{1, 3, 5, 6}
Faculty of Education, Arab Open University, Muscat 122, Oman²
Cairo University, Giza, Egypt²
Faculty of Business Studies, Arab Open University, Muscat 122, Oman⁴

Abstract—Integrating artificial intelligence (AI) in education has introduced innovative approaches, particularly in personalized learning and dynamic assessment. Conventional teaching models often struggle to address learners' diverse needs and abilities, underscoring the necessity for AI-driven flexible learning frameworks. This study explores how AI-aided smart learning paths and dynamic assessments enhance learning efficiency by improving knowledge acquisition, optimizing task completion time, and increasing student engagement. A six-week quasi-experimental study was conducted with 200 students, divided into an experimental group using an AI-based learning system and a control group following traditional methods. Pre- and post-tests and engagement analyses were used to evaluate outcomes. The experimental group demonstrated a 25% improvement in performance, completed tasks 25% faster, and showed a 15% increase in engagement compared to the control group. These findings highlight the potential of AI to deliver personalized learning experiences and timely feedback, significantly enhancing student outcomes. Future research should involve larger participant groups across higher educational levels and examine the long-term impact of AI-supported education on students' knowledge retention and skill reinforcement.

Keywords—AI-powered learning; adaptive learning; dynamic assessments; education technology; personalized learning pathways; student engagement

I. INTRODUCTION

Artificial Intelligence (AI) is revolutionizing education in a way because it helps make learning more flexible, customized, and based on data. The conventional, institutionalized approach to learning that has been in practice for many years across many countries has been found unsuitable to handle the needs and interests of the students [1]. This gap has, therefore, created a new interest in using AI to develop an environment where a learning path is developed to suit personal learning rates, personal preferences, and learning abilities of learners. Furthermore, AI brings dynamic assessments into the picture; here, evaluating and even giving feedback are done in real-time, hence no stagnated learning process [2]. AI is being incorporated into various educational systems throughout the globe as an educational tool in areas such as intelligent tutoring systems, AI analytic and adaptive learning systems [3]. AI has been very useful in determining the special needs of individual students through the processing of large volumes of data

concerning the student's performance, attendance, and behavior, among others [4]. Solutions like auto-grading, virtual tutors, and learning analytics are assisting educators in knowing the students' achievements; therefore, they get to devise more effective methods of teaching [5]. That is why, by providing solutions for automating many administrative processes, such as grading work and tracking student attendance or performance, AI relieves educators from many concerns so they can dedicate more attention to the students, making the learning process as individual as possible.

It can also be claimed that such learning environments can be personalized and favorably optimized in real-time, depending on the needs of a student. For instance, if a student is doing poorly in each concept, then the system can change the level of difficulty, suggest materials and resources, or even ask questions that the student has to answer so that he gets it. Adaptive learning environments assist in reaching the maximum level of students' capabilities by adapting the learning process according to their specific needs and preferences, making the learning process productive and efficient [6]. It is, therefore, expected that the integration of AI can transform how education is provided and received, altering the learning methods for students in various settings. Individualization of education is the approach to delivering instructions to cater to each learner's needs, strengths, and preferred mode of learning. When AI is integrated into the learning system, it becomes easier to contemplate personalized learning because intelligent systems can explain performance data to formulate learning pathways that the student will appreciate and, at the same time, are on par with their abilities and learning curve. These systems permanently gather and analyze information about students' relationships with learning materials so they can advise on changes to the teaching methods preferred by each student [7]. Personalized learning is a concept that departs from the very rigid structure of a traditional learning system where the students are expected to work in groups or sets and are given grouping learning curves that tailor the learning needs of each person. Dynamic assessments supplement differentiated instruction by providing an assessment of student learning in a more frequent, real-time manner. Traditional forms of assessment can be static, timed, and administered at set points in time and, therefore, do not give a complete picture of the learning process of a learner [8]. In contrast, dynamic assessments, on the other hand, make use of AI, which assesses to be dynamic depending on the

*Corresponding Author.

The research leading to these results has received funding from the Arab Open University under the Internal Funding Program with ID AOU_OM/2023/FCS7.

performance of the student; the student is provided with feedback immediately and is given directions to further improvement or assistance where necessary [9]. This is very beneficial when determining a student progress since it presents teachers with guideline on aspects that requires special attention. Consequently, there is a benefit of dynamic assessments in that they can encourage a developmental view of assessment as a continuous process rather than a one-time event. This study is guided by the following research question: "How can AI-powered learning pathways, integrating personalized learning and dynamic assessments, enhance student engagement, motivation, and mastery across diverse educational contexts?" This question addresses a critical challenge in contemporary education by exploring the transformative potential of AI in tailoring learning experiences to individual needs. The study aims to contribute meaningful insights into the design and implementation of adaptive educational frameworks, offering a pathway to more effective and inclusive learning environments.

When it comes to components and interactions in making use of AI to support effective learning and related features, there are certain components and behaviors, as depicted in Fig. 1. Real-time feedback, individual student analysis, continuous assessment, content delivery, and learning adaptations are facilitated by the central mechanism of AI.

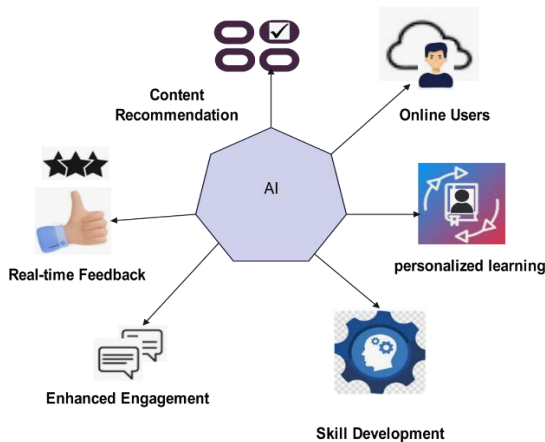


Fig. 1. AI-powered learning pathway components.

This paper aims to explore how AI can be effectively leveraged to create personalized learning pathways and dynamic assessment systems that respond to individual student needs. The research focuses on three primary objectives: first, to create an AI learning map that proposes models for formative and summative ‘dynamic assessments’ of learning pathways that will change depending on students’ performance; second, to offer a scoped real-world application of both formative and summative dynamic assessments within learning environments where timely feedback improves student learning outcomes; and, third, how AI is applied in practice in education, and to discuss the issues and potential solutions about scale-up of dynamic.

The key contribution of this paper is a closer inspection of how AI leverages the flipped classroom model to provide more contextualized and student-centric learning that is dynamic.

Besides, the paper offers an idea of how dynamic assessment models can be implemented practically and gives recommendations for educators and institutions willing to use AI-based learning systems. This study helps to fill the gaps in demand for learning environments that are learner-centered and are enhanced by the advent of technologies such as AI to develop an understanding for future studies under the proposed framework of AI in learning and assessments. The rest of the paper is organized as follows: The Literature Review in Section II explores existing work in personalized learning and dynamic evaluations, highlighting gaps addressed by this research. The Methodology in Section III details the data collection, AI model development, system implementation, and evaluation procedures. The Results and Discussion in Section IV presents the findings of the quasi-experimental study, comparing AI-based and traditional learning methods, followed by an analysis of their implications. Finally, the Conclusion in Section V summarizes the key contributions, discusses limitations, and suggests directions for future research.

II. RELATED WORK

Making use of AI in the education context has brought possibilities of reapportioning the traditional learning paradigms and processes. This section presents a discussion of the current state of the approach to learning personalization and the application of AI in general in learning environments. It highlights why the focus on dynamic educational paths built with the help of AI is relevant to the research.

A. Current Approaches to Personalized Learning

With the help of AI, the concept of learning has become Personalized learning, which is different from conventional teaching-learning processes [10]. One of the approaches is adaptive learning whereby the AI of the systems modifies the content of the lessons according to the results of the learners. This method widely adapts to the learner’s needs and delivers learning material at the learner’s acceptable speed, level of difficulty, and method. From interactions with the students like quizzes and learning activities, the next steps are recommended by an algorithm for the student [11]. Another approach includes the ITS, which acts as if tutoring one trainee [12]. These systems locate areas where a student may lack knowledge and provide their feedback together with self-practice. Students’ progress at ITS can be followed, and recommendations for interventions can be made, making the learning process individualized without the teacher’s interference. Another application of big data is in the early identification of students’ potential for improvement or worsening results in the light of existing data. This means that if a student develops some form of challenge, they are easily identified to receive early intervention [13]. However, there are still issues that stakeholders encounter while trying to apply personalized learning strategies on a large scale. This means that some considerations, like poor access to technology in low-performing schools and matters of privacy, limit its use in schools.

B. AI Applications in Education

In education, AI is not only used for learning but for all the processes that appear in the concept. For instance, automated grading systems have received consideration anew due to their

efficiency in evaluating student submissions, especially in formats common to standardized testing within a short time [14]. These systems employ the use of AI to grade answers so that the educators will not spend much time on this activity while at the same time, the students get useful feedback from the system. Even more current progress allows even AI to evaluate other tasks as writing an essay using natural language processors (NLP) [15]. It is also playing a role in developing a virtual learning environment. Automated intelligent chatbots are now being adopted for round-the-clock student support and to help learners navigate and understand the course content and respond to basic queries [16]. These bots also contribute positively to the educational process since students can get help from the bot without the need to involve a teacher. In the same manner, AI-based tools in analytics are useful in monitoring the progression of learning of the students as well as observing behaviors that may require further attention in learning [4]. Further, it is creating efficiencies in administration, as well as enrolling, scheduling, and resource management tasks. This, in turn, minimizes some of the administrative costs within institutions and enables institutions to devote adequate time to enhancing the quality of teaching and learning. It is only to be assumed that in the future, we will have even more advanced systems, individual tutoring-based virtual avatars, as well as a symbiotic relationship between AI and augmented reality technologies like virtual reality.

C. *Dynamic Assessment Models*

Dynamic assessment is a relatively new concept within the education sector that possesses a dynamic assessment model as against the more static assessment models such as examinations and quizzes [17]. Alphabets implementation makes it possible to achieve dynamic assessments since changes are made virtually, without affecting student performance in any way. Such models are supposed to indicate not only the knowledge of a student at a certain point in time but also their learning processes and developing knowledge incrementally. Dynamic assessments have been proven to be very effective since they give chances of quick feedback hence informing both learners and trainers on their strong areas and areas that require more focus. Unlike the static form of assessment that examines the extent of the student's knowledge at a certain point in time, an assessment of this type has a dynamic characteristic and can change when the student is being asked a question [18]. For example, if the student answers a question correctly the next may be more difficult. If a student exerts effort and accurately solves a problem, then the next one may be more difficult to solve, if a student is unable to solve a problem, then the system provides easy problems or additional materials to learn from. It can be done in real-time so that if the student gets stuck, it is easy to identify areas that the student needs to learn and adapt the content according to the student's requirements [19]. In addition, dynamic assessments may use formative components, which provide information regarding learning progress, rather than being used solely for the evaluation of the results of learning at the end of a course. This form of assessment helps the students come out of mistakes and also helps to tackle problems with more determination. AI predicts likely future learning difficulties, thus enabling the planner to put in measures long before big learning gaps surface [20]. These formative assessments offer a continual flow of information

about each student learner's development which makes it easier for a teacher to develop learning plans. An example of dynamic evaluation is the ITS, where testing is dynamic in that it adapts to the student's performance [21]. These assessments are not just score-based but also point to how well a student understands concepts of specific areas. This is especially helpful in understanding areas where there are learning hurdles and can be addressed early enough with the view of enhancing students' performance in the long run [22]. The study in [23] explored the use of advanced time-series models like RNN, LSTM, and GRU to predict student performance and dropout rates. It highlights the superiority of these models over traditional methods and emphasizes the importance of architecture and hyperparameter tuning for accurate predictions and effective interventions in platforms like MOOCs.

In study [24], the authors developed a machine learning-based system, with Random Forest identified as the most effective model for predicting student outcomes (graduate, dropout, or enrolled). By analyzing demographic, socioeconomic, and academic data, the system provides personalized learning strategies, demonstrating its potential to reduce dropout rates and improve academic success through data-driven interventions.

Dynamic assessments are particularly helpful in those learning situations where learners get the attention required to make change. Due to such an approach used in the assessment process, the students are not only evaluated but assisted in enhancing the right learning processes. However, the practical application of dynamic assessment models entails major systems' support, protection of data acquired and shared, and professional development to understand the implications of the assessments provided by these systems. A comparative analysis of existing studies reveals both advancements and limitations in the application of AI for personalized learning and dynamic assessments. While many studies have explored adaptive learning platforms that tailor content to individual preferences, these often lack integration with systems that provide continuous, real-time feedback based on a student's evolving performance. Conversely, some research has focused on dynamic assessment techniques but does not combine them with broader personalized learning frameworks. This study bridges these gaps by integrating AI-powered personalized learning pathways with real-time adaptive assessments into a unified system. This approach ensures not only individualized content delivery but also continuous evaluation and timely interventions, offering a more comprehensive and effective learning experience compared to existing methodologies.

III. PROPOSED APPROACH

In this section, an AI-driven intervention approach is proposed that would entail the differentiation of learning contracts for each learner based on their performance, choices, and past performance. This approach involves the use of AI in designing the flow, content, and modality of teaching, learning, and assessment so that any student's learning requirements are fully met. To achieve that, our model uses effective data collection, feedback, and adaptive learning approaches to provide students with the best learning experience. The general aim of this approach is to maximize the level of participation of

students and the efficiency of the learning process by designing paths that develop the actions and advancements of the learner.

A. AI-Powered Personalized Learning Pathways

AI-enabled learning pathways are expected to assist in delivering an education experience to learners that is tailored to their needs. These pathways aim at addressing different factors that may be hinging on the student such as the level of learning, how the student grasps this kind of information, the rates of learning, and the kind of learning that the student prefers. The basis of this strategy belongs to AI methodologies that interactively collect and process information on learners' engagement with learning materials. Concerning progress and areas of successful learning or learning challenges, the system adapts learning and the experience of each student.

1) *Data collection and analysis:* The system gathers information from multiple sources, such as students' engagement with multimedia solutions and tutorials, quiz scores, student engagement in activities, etc. So, the data collected regarding the students is analyzed with the help of Machine learning models to find patterns in their behavior and performance. The system then utilizes these to consider, in which segment of the curriculum the student needs more help and which part has been easily understood.

2) *Adaptive learning content:* According to the findings, the AI system offers consumable learning material that is within the student's grasp. For instance, if a student is performing well in a particular topic, the system may introduce the student to the hardest content in the topic in question; to the student who is poor in the topic in question, the system will provide simple content on the topic in question. Thus, using this adaptive approach, the students are kept alert by presenting them with tasks that are not overly complex, which enhances an optimal rate of learning.

3) *Real-time feedback and adjustments:* The personalized learning pathway is adjusted in real time depending on the student's activity while in the learning process. For example, suppose a student performs poorly in some specific subject area. In that case, the system may present them with more examples or a more detailed explanation of the concept being discussed. On the other hand, if the student can perform well, then the pathway can either engage material at a faster pace or even omit concepts that the student has already mastered.

4) *Dynamic assessment integration:* The system also incorporates dynamic assessments where it regularly gives feedback on students' performance. These assessments are not constant in their level of difficulty and are thus more specific in their approach to identifying a learner's learning requirements. Thus, evaluating the student's knowledge determines the areas that need to be filled and the path that should be followed.

5) *Customization based on learning preferences:* The system used incorporated AI to consider features of personal learning styles, for instance, whether a given student learns best with visual displays, articles, or exercises. In this way, by adjusting student's access to information and material, the system increases interest and saves useful information for

further use. For instance, action-oriented learners could be given more video lectures and illustrative diagrams, while others may get textual descriptions or other forms of simulation.

Fig. 2 shows the learning pathway model with references to AI. It captures information from the student's activities and, through the application of machine learning, performs analysis of the results. Following the assessments outlined, the pathway dynamically responds and delivers personalized content in addition to ongoing dynamic assessments. Archer's set-up of learning pathways and real-time feedback guarantees that the learning process is re-adjusted to increase efficiency.

This factor enhances users' interest because it involves learning that targets personal abilities and difficulties. Such an approach means that students do not get bored with content and, at the same time, do not face the overwhelming of complex information.

This approach can be of immense benefit when used in big classes, whereby it may be difficult for the instructors to attend to every single student in the class. AI systems with learning pathways provide every student with a method of learning that is unique to every student, the goal of which is also to reach the goal that has been set for learning and give the students incentives to learn more while doing it in a shorter amount of time.

This proposed approach offers a one-stop solution to improving a personalized approach to learning at large by using techniques such as adaptive learning, data analysis as well as continuous assessment.

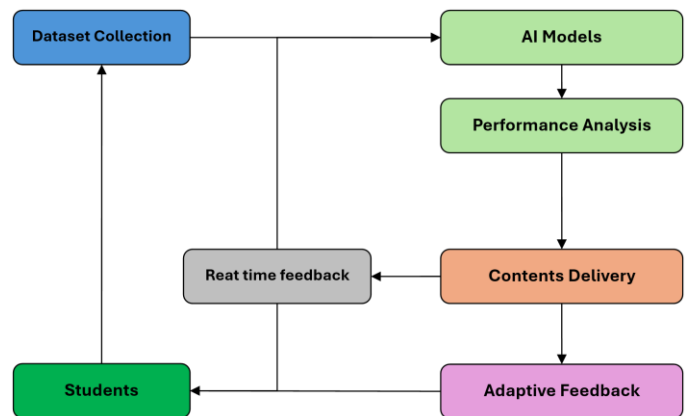


Fig. 2. AI-powered personalized learning pathway framework.

B. Dynamic Assessment Integration

In the proposed AI-based personalized learning pathway, the inclusion of dynamic assessment is envisaged to play a central role. Unlike typical assessment practices, which are pre-ordained and sequential, dynamic assessments are contingent and, occur in real-time and change depending on the student's performance. This approach makes it possible for the system to use AI algorithms to constantly assess the performance of a student and, therefore, improve the flexibility of the system in offering lessons to the students. Here, it is going to be described how dynamic assessment incorporates mathematics and how it fits into data-driven learning approaches.

Let's define the student's knowledge state as a vector $K(t)$ at any time t , where $K(t)$ is defined in Eq. (1).

$$K(t) = [k_1(t), k_2(t), \dots, k_n(t)] \quad (1)$$

Here, $k_i(t)$ represents the student's proficiency in the i^{th} topic or concept at time t , and n is the total number of topics in the learning pathway.

Dynamic assessments continuously update $K(t)$ based on the student's responses to questions, interaction with learning materials, and performance on exercises. The change in the knowledge state over time can be modeled as a differential equation, as shown in Eq. (2).

$$\frac{dK(t)}{dt} = \alpha A(t) - \beta L(t) \quad (2)$$

where $A(t)$ is the assessment score at time t , $L(t)$ represents the learning difficulty or cognitive load at a time t , α and β are weighting factors that balance the effect of assessments and cognitive load on knowledge acquisition.

The assessment score $A(t)$ is calculated based on the student's performance in a series of adaptive questions or tasks. Each question Q_i is associated with a difficulty level D_i and is chosen based on the current knowledge state $K(t)$. The score $A(t)$ is determined by Eq. (3).

$$A(t) = \sum_{i=1}^m w_i \cdot R_i(t) \quad (3)$$

where m is the number of questions in the assessment, w_i is the weight assigned to the i^{th} question based on its difficulty level D_i , $R_i(t)$ is the student's response to the i^{th} question, which is 1 for a correct answer and 0 for an incorrect answer.

The system dynamically adjusts the difficulty of subsequent questions based on the student's previous responses. If a student answers a question correctly, the system may increase the difficulty of the next question, while incorrect answers may result in easier questions being presented. Mathematically, the difficulty level of the next question D_{i+1} is updated as is Eq. (4).

$$D_{i+1} = D_i + \gamma(R_i(t) - 0.5) \quad (4)$$

where γ is a scaling factor that controls the sensitivity of the difficulty adjustment. A correct answer increases the difficulty of the next question, while an incorrect answer decreases it.

1) *Real-time feedback and adaptation:* As the system continuously monitors the student's performance through dynamic assessments, it updates the personalized learning pathway in real-time. The goal is to maintain the cognitive load within an optimal range to maximize learning efficiency. The cognitive load $L(t)$ is influenced by the difficulty level of the content and the student's current state of knowledge. It can be modeled as in Eq. (5).

$$L(t) = \sum_{i=1}^m \lambda_i \cdot D_i \cdot (1 - k_i(t)) \quad (5)$$

where λ_i is the weight associated with the importance of the i^{th} topic, D_i is the difficulty level of the i^{th} topic, $k_i(t)$ represents the student's proficiency in that topic.

The system aims to adjust the learning path by keeping $L(t)$ within a predefined threshold L_{opt} , which represents the optimal cognitive load for learning. If $L(t) > L_{opt}$, the system reduces the difficulty of subsequent topics or provides additional scaffolding. If $L(t) < L_{opt}$, the system increases the difficulty of keeping the student engaged and challenged.

2) *Optimization of learning pathway:* The integration of dynamic assessment into the learning pathway allows for continuous optimization. The system uses real-time data from assessments to update the knowledge state vector $K(t)$ and adjust the content accordingly. The objective is to minimize the difference between the desired knowledge state $K^*(t)$ and the actual knowledge state $K(t)$ at any given time, which can be formulated in Eq. (6) as a cost function J :

$$J(t) = \|K^*(t) - K(t)\|^2 \quad (6)$$

The learning pathway is optimized by minimizing $J(t)$, ensuring that the student's knowledge state converges toward the desired state over time. AI algorithms, such as reinforcement learning, can be applied to solve this optimization problem by selecting the most effective instructional strategies and assessment questions at each step.

C. Adaptive Algorithms and Feedback Loops

Algorithms are at the heart of AI-based personalized learning models because they have to incorporate flexibility. It means that these algorithms change the content, the rate, and the assessments according to the interactions and performances of the students in real-time. The idea is to deliver individual learning, which means the system should be adjusted to learner needs and in which the learner is challenged but not overwhelmed.

The mechanisms of adaptive algorithms focus on the integration of feedback loops to establish the effectiveness of a responsive learning environment. Student data include performance on the test, the interaction with peers as well as time spent on the task and such data are used to adapt the learning process for the student.

Adaptive algorithms use data collected at the time to determine what should happen shortly in the student's learning process. All these algorithms take into consideration various input variables, such as the performance of the students, the time they take to answer the questions, and even the engagement figures. The system monitors the accomplishments of students and how they were able to do it in the assessments and activities. The time a student takes to answer a question or complete a task can indicate their confidence or difficulty level. Data on how often a student interacts with learning materials helps the system adjust the difficulty and type of content delivered.

Based on these variables, adaptive algorithms continuously modify the content and assessments. The system's core objective is to maintain an optimal learning pace that challenges

students without overwhelming them, ensuring steady progress. Algorithm 1 is for how an adaptive learning system might function with integrated feedback loops:

Algorithm 1. Adaptive Learning with Feedback Loops

Input: Initial knowledge state K_0 , content difficulty D_0 , student response time τ , learning rate α , scaling factor γ , performance threshold ϵ .

Output: Updated learning parameters θ_t , optimized learning pathway.

1. **For** $t = 1$ to T (epochs) **do**
2. Present learning content L_t with difficulty D_t
3. Record student response R_t and response time τ_t
4. Update knowledge state: $K_t \leftarrow \beta_1 \cdot K_{t-1} + (1 - \beta_1)R_t$
5. Update learning objective: $L_t \leftarrow \gamma \cdot \tau_t \cdot (1 - K_t)$
6. Compute bias-corrected knowledge estimate: $\hat{K}_t \leftarrow \frac{K_t}{1 - \beta_1^t}$
7. Compute bias-corrected learning objective: $\hat{L}_t \leftarrow \frac{L_t}{1 - \beta_1^t}$
8. Update learning parameter: $\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \frac{\hat{K}_t}{\sqrt{\hat{L}_t + \epsilon}}$
9. Adjust content difficulty: $D_{t+1} \leftarrow D_t + \gamma \cdot (\hat{K}_t - 0.5)$
10. **End For**

Return θ_t (final optimized learning parameters)

Implementing the AI-powered learning pathways system involved a multi-layered approach to ensure its adaptability, functionality, and scalability. Python was selected as the primary programming language due to its extensive support for machine learning and data processing, with frameworks such as TensorFlow and PyTorch utilized for model development. The backend was built using Flask to enable seamless scalability, while the user interface was designed with React.js to provide an intuitive and engaging experience for educators and students. The raw data, including learning behaviors, preferences, and performance metrics, underwent extensive preprocessing using Pandas and NumPy to ensure consistency, handle missing values, and extract meaningful features. AI models were then trained to analyze this data, employing supervised learning techniques for predicting individual learning needs and reinforcement learning for optimizing dynamic assessments.

The conventional teaching sessions were structured using a standardized curriculum aligned with the study's objectives. Lesson plans were developed to cover the same content as the AI-based system, ensuring parity in learning objectives. Traditional instructional materials, including textbooks, printed handouts, and multimedia presentations, were utilized to deliver the content. Teaching techniques followed a lecture-based format supplemented with interactive classroom discussions and periodic assessments to monitor student progress. These details have been incorporated to enhance the transparency of the methodology and provide a clearer basis for interpreting the comparative results of the study.

The adaptive assessment system integrated natural language processing (NLP) for automated question generation and AI algorithms for real-time performance tracking, dynamically adjusting question difficulty and type based on the student's progress and mastery levels. The overall system architecture was designed with modularity in mind, comprising a data layer

for storage and retrieval, an AI engine for learning and assessment adaptation, and an application layer that hosted user-facing features like dashboards and progress reports. The entire system was deployed on a cloud platform, such as AWS or Google Cloud, to ensure accessibility and scalability, with continuous integration and deployment pipelines established using Jenkins and Docker for smooth updates. Pilot testing was conducted in real classroom settings to evaluate the system's performance, with feedback from users incorporated to refine its features and enhance usability.

D. Measuring Engagement Levels

To effectively evaluate the impact of the AI-powered learning pathways system, a robust framework for measuring student engagement levels is essential. Engagement is assessed through a combination of quantitative and qualitative metrics, ensuring a comprehensive understanding of how students interact with the platform and learning materials. Student interaction with the platform is monitored through log data, capturing behaviors such as the frequency of logins, time spent on individual activities, and the number of interactions with learning resources. These metrics provide insights into active participation and overall engagement with the system. The system tracks response times for quizzes and assessments, as well as the rate at which students complete assigned tasks. Quick response times and high completion rates indicate consistent engagement, while delays or unfinished tasks may signal a need for intervention. Engagement is also inferred from behavioral patterns, such as the use of optional resources, reattempts at challenging exercises, and participation in collaborative activities like discussion forums or peer reviews. These indicators reflect deeper involvement with the learning content. To complement behavioral data, students are regularly asked to provide self-reported feedback through in-platform surveys. These surveys measure perceived engagement, motivation, and satisfaction with the learning pathways and assessment system. AI algorithms analyze the collected data to identify trends and patterns in engagement. For example, machine learning models assess correlations between engagement metrics (e.g., time spent on tasks) and learning outcomes (e.g., assessment performance). This analysis enables the system to adapt to students' engagement levels by modifying learning content or assessment strategies to maintain interest and motivation.

IV. EXPERIMENTS AND RESULTS

The experiment was conducted to evaluate the effectiveness of AI-powered personalized learning pathways and dynamic assessments. A group of 100 students was divided into two groups: a control group using traditional learning methods and an experimental group using the AI-powered adaptive learning system. The subjects studied similar content in a mathematics course over 6 weeks. The performance was measured through pre-tests, post-tests, and continuous assessments. While the target system adjusted the level of the material according to the state of knowledge of the student, the control group used a set curriculum. Such data as the assessment and the time spent on the task, as well as the engagement, were obtained. The results obtained in the two groups were compared to determine the effects of the adaptive system on learning.

A. Results on Personalized Learning Efficiency

The experiment aimed to compare the efficiency of learning that is based on the use of new technologies, particularly, the AI-based personalized learning pathways. Efficiency was measured using three key metrics: (1) improvement in student performance (knowledge gain), (2) time spent on learning tasks, and (3) engagement levels. These metrics were compared between the experimental group (using the AI-powered personalized learning system) and the control group (using traditional learning methods).

1) *Improvement in student performance (Knowledge gain):* Students' knowledge gain was assessed by comparing their pre-test and post-test scores. The experimental group showed a significant improvement in their performance compared to the control group. On average, the experimental group improved from 55% to 80% in their post-test scores, whereas the control group only increased from 54% to 68%, as shown in Table I.

TABLE I. COMPARISON OF PRE-TEST AND POST-TEST SCORES

Group	Pre-test Average (%)	Post-test Average (%)	Performance Improvement (%)
Control Group	54%	68%	14%
Experimental Group	55%	80%	25%

The higher performance improvement in the experimental group suggests that the adaptive learning system helped students better understand and retain the material by tailoring the learning experience to their individual needs. Fig. 3 compares the pre-test and post-test performance improvement between the control and experimental groups.

2) *Time spent on learning tasks:* One of the major advantages of AI-powered personalized learning pathways is their ability to optimize the time students spend on tasks. By adjusting content difficulty in real-time, students in the experimental group spent 25% less time on tasks compared to the control group, as shown in Table II.

TABLE II. AVERAGE TIME SPENT ON LEARNING TASKS

Group	Average Time per Task (minutes)	Time Reduction (%)
Control Group	40	N/A
Experimental Group	30	25%

This reduction in time demonstrates that the adaptive learning system allows students to focus on areas where they need improvement, resulting in more efficient learning. Fig. 4 illustrates the comparison of the average time spent per task between the control and experimental groups.

3) *Engagement levels:* The AI-powered personalized learning system also resulted in higher engagement levels. The system provided content that was both challenging and suited to the student's learning pace, leading to higher interaction with the platform. The experimental group had, on average, 15% higher engagement than the control group, as shown in Table III.

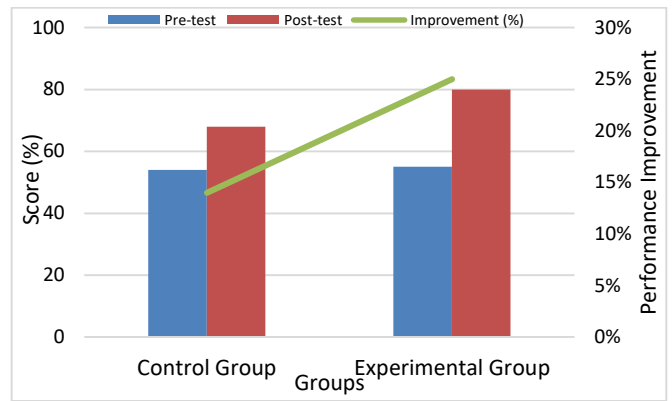


Fig. 3. Performance improvement across both groups.

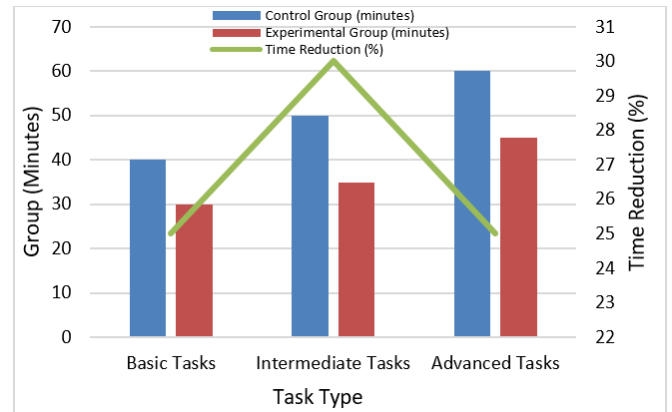


Fig. 4. Time efficiency comparison.

TABLE III. ENGAGEMENT METRICS COMPARISON

Group	Average Weekly Sessions	Average Session Duration (minutes)	Engagement Increase (%)
Control Group	3	45	N/A
Experimental Group	4	52	15%

Higher engagement in the experimental group indicates that students were more motivated and focused when using the adaptive system. Fig. 5 compares the weekly session frequency and session duration between the control and experimental groups.

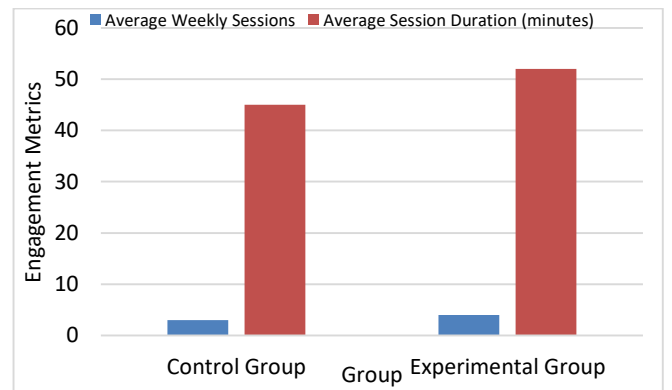


Fig. 5. Engagement levels comparison.

The results across all metrics — knowledge gain, time efficiency, and engagement levels — show that the AI-powered personalized learning pathways significantly enhanced learning efficiency compared to traditional methods. Students in the experimental group demonstrated higher performance improvement, spent less time completing tasks, and were more engaged with the learning content.

The increase in knowledge gain (Table I, Fig. 3) highlights the system’s ability to tailor learning materials to each student’s needs. The decrease in time spent on tasks (Table II, Fig. 4) shows that the adaptive system saves time in guiding students to pay more attention to the problematic material more efficiently. Therefore, the enhancement of the level of engagement (Table III, Fig. 5) supports the notion that the strategy of integrating individual interests kept the students engaged and interested in learning. The consolidation of the given evaluation of personalized learning indicates the possibility of applying AI systems to reform the educational process since the impact of the learning process might be enhanced for each student.

B. Comparison with Traditional Methods

To verify the application of AI intelligent learning pathways of learning, relevant literature that was majorly focused on traditional approaches to learning was compared. The comparison was made based on the number of new facts, the time it took to complete the activities, and the level of engagement of learners. As indicated in.

Table IV, all performances indicate that AI-powered methods have higher performances than traditional methods by large margins. The personalized learning group displayed a 25% performance increase as opposed to the 14% increase that the conventional learning group showed. Maintenance of routine tasks was done 25% faster for those students who employed an AI-powered system. In terms of the engagement rate, the experimental group proved to be 15% more engaged than the control group.

TABLE IV. COMPARISON OF AI-POWERED VS. TRADITIONAL LEARNING METHODS

Metric	AI-Powered Learning	Traditional Learning	Difference
Knowledge Gain	25% improvement	14% improvement	+11%
Time Efficiency	30 minutes/task	40 minutes/task	-25%
Engagement Increase	15%	N/A	+15%

V. CONCLUSION AND FUTURE WORK

The development of AI-powered smart learning paths marks a significant advancement in educational technology, offering a tailored approach to addressing the unique needs of individual learners. This study investigated the potential of improving learning outcomes and academic performance, particularly in distance education, through the use of AI-driven systems that adapt content dynamically based on feedback and assessments. The findings of this research provide compelling evidence in favor of personalized learning systems. Students

utilizing the AI-powered system achieved a post-test average that was 25% higher compared to a 14% improvement observed in those following traditional methods. This result emphasizes the superior efficacy of adaptive learning paths in enhancing academic achievement. Furthermore, the AI-supported system enabled students to complete tasks 25% faster than conventional learning approaches, demonstrating its capacity to streamline the learning process without compromising comprehension. Additionally, student engagement levels increased by 15%, facilitated by the system’s ability to maintain interest through personalized challenges, project-based learning, and dynamic feedback mechanisms.

It is observed that the AI-based learning system significantly improved time efficiency and performance compared to conventional methods, aligning with previous research findings that emphasize the potential of AI in optimizing learning processes [25]. This agreement with prior studies reinforces the reliability of AI-driven educational tools in similar contexts. These findings highlight a novel aspect: the ability of the AI-based system to dynamically adapt to student learning patterns, which has not been extensively addressed in prior literature. This discovery underscores the unique contribution of our research to the field of AI in education. Moreover, while previous research has focused primarily on long-term AI-based interventions, our short-term study demonstrates that measurable impacts can also be observed within a limited timeframe, providing complementary insights into the application of AI in education."

Despite these promising results, several important areas warrant further investigation. Scalability remains a critical consideration, as the implementation of AI-powered systems in larger and more diverse educational settings presents unique challenges. Future studies should explore how such systems can maintain effectiveness and adaptability in substantially broader and more heterogeneous learning environments. Additionally, while this research highlights short-term benefits, the long-term effects of AI-based personalized learning require closer examination. Establishing whether improvements in retention, comprehension, and performance persist over time is essential for validating the sustainability of these systems. While this study demonstrates the quantitative benefits of AI-based learning, future research must incorporate qualitative methods to understand the student experience and engagement with these systems fully. Another pressing issue involves ethical considerations, particularly in relation to data privacy, fairness, and transparency in AI algorithms. There is a pressing need for the development of robust ethical frameworks to guide the responsible deployment of AI technologies in education, ensuring equitable access and trustworthiness. This study underscores the transformative potential of AI in education, demonstrating its ability to deliver personalized, efficient, and engaging learning experiences. By addressing scalability challenges, investigating long-term effects, and developing ethical frameworks, future research can ensure that AI continues to revolutionize education in a way that is both impactful and responsible. The results contribute novel insights to the growing body of knowledge on AI in education, reinforcing its role as a catalyst for positive change while identifying critical areas for further exploration.

ACKNOWLEDGMENT

"The research leading to these results has received funding from the Arab Open University under the Internal Funding Program with ID AOU_OM/2023/FCS7."

REFERENCES

- [1] T. Bates, C. Cobo, O. Mariño, and S. Wheeler, "Can artificial intelligence transform higher education?," vol. 17, ed: Springer, 2020, pp. 1-12.
- [2] E. Dimitriadou and A. Lanitis, "A critical evaluation, challenges, and future perspectives of using artificial intelligence and emerging technologies in smart classrooms," *Smart Learning Environments*, vol. 10, no. 1, p. 12, 2023.
- [3] A. Prasanth, J. V. Densy, P. Surendran, and T. Bindhya, "Role of artificial intelligence and business decision making," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 6, 2023.
- [4] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education—where are the educators?" *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, pp. 1-27, 2019.
- [5] C. Xu and L. Wu, "The Application of Artificial Intelligence Technology in Ideological and Political Education," *International Journal of Advanced Computer Science & Applications*, vol. 15, no. 1, 2024.
- [6] I. Gligorea, M. Cioca, R. Oancea, A.-T. Gorski, H. Gorski, and P. Tudorache, "Adaptive learning using artificial intelligence in e-learning: a literature review," *Education Sciences*, vol. 13, no. 12, p. 1216, 2023.
- [7] K. Chrysafiadi and M. Virvou, "Student modeling approaches: A literature review for the last decade," *Expert Systems with Applications*, vol. 40, no. 11, pp. 4715-4729, 2013.
- [8] L. Dunn, C. Morgan, M. O'Reilly, and S. Parry, *The student assessment handbook: New directions in traditional and online assessment*. Routledge, 2003.
- [9] T. Kubiszyn and G. D. Borich, *Educational testing and measurement*. John Wiley & Sons, 2024.
- [10] O. Tapalova and N. Zhiyenbayeva, "Artificial intelligence in education: AIED for personalised learning pathways," *Electronic Journal of e-Learning*, vol. 20, no. 5, pp. 639-653, 2022.
- [11] J. F. Pane, E. D. Steiner, M. D. Baird, and L. S. Hamilton, "Continued Progress: Promising Evidence on Personalized Learning," *Rand Corporation*, 2015.
- [12] A. Alkhatlan and J. Kalita, "Intelligent tutoring systems: A comprehensive historical survey with recent developments," *arXiv preprint arXiv: 1812.09628*, 2018.
- [13] R. Shaun, J. De Baker, and P. Inventado, "Chapter 4: Educational Data Mining and Learning Analytics," ed: Springer, 2014.
- [14] J. C. Paiva, J. P. Leal, and Á. Figueira, "Automated assessment in computer science education: A state-of-the-art review," *ACM Transactions on Computing Education (TOCE)*, vol. 22, no. 3, pp. 1-40, 2022.
- [15] K. Taghipour and H. T. Ng, "A neural approach to automated essay scoring," in *Proceedings of the 2016 conference on empirical methods in natural language processing*, 2016, pp. 1882-1891.
- [16] D. Ramandanis and S. Xinogalos, "Investigating the Support Provided by Chatbots to Educational Institutions and Their Students: A Systematic Literature Review," *Multimodal Technologies and Interaction*, vol. 7, no. 11, p. 103, 2023.
- [17] E. G. Estrada-Araoz, B. T. Sayed, G. G. Niyazova, and D. Lami, "Comparing the effects of computerized formative assessment vs. computerized dynamic assessment on developing EFL learners' reading motivation, reading self-concept, autonomy, and self-regulation," *Language Testing in Asia*, vol. 13, no. 1, p. 39, 2023.
- [18] G. E. DeBoer et al., "Comparing three online testing modalities: Using static, active, and interactive online testing modalities to assess middle school student's understanding of fundamental ideas and use of inquiry skills related to ecosystems," *Journal of Research in Science Teaching*, vol. 51, no. 4, pp. 523-554, 2014.
- [19] R. J. Mislevy, J. T. Behrens, K. E. Dicerbo, and R. Levy, "Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining," *Journal of educational data mining*, vol. 4, no. 1, pp. 11-48, 2012.
- [20] K. Ahmad et al., "Data-driven artificial intelligence in education: A comprehensive review," *IEEE Transactions on Learning Technologies*, 2023.
- [21] Z. Jeremić, J. Jovanović, and D. Gašević, "Student modeling and assessment in intelligent tutoring of software patterns," *Expert Systems with Applications*, vol. 39, no. 1, pp. 210-222, 2012.
- [22] C. T. Ramey and S. L. Ramey, "Early learning and school readiness: Can early intervention make a difference?" *Merrill-Palmer Quarterly*, vol. 50, no. 4, pp. 471-491, 2004.
- [23] D. Herath, C. Dinuwan, C. Ihalagedara, T. Ambegoda, "Enhancing Educational Outcomes Through AI Powered Learning Strategy Recommendation System," *International Journal of Advanced Computer Science & Applications*. 2024 Oct 1;15(10).
- [24] S. Vanitha, "Towards finding the impact of deep learning in educational time series datasets—A systematic literature review," *International Journal of Advanced Computer Science and Applications*. 2023;14(3).
- [25] O. Onesi-Ozigagun, Y. Ololade, N. Eyo-Udo, and D. Ogundipe, "Revolutionizing education through AI: a comprehensive review of enhancing learning experiences," *International Journal of Applied Research in Social Sciences*, 2024, <https://doi.org/10.51594/ijarss.v6i4.1011>.

Feature Reduction and Anomaly Detection in IoT Using Machine Learning Algorithms

Adel Hamdan¹, Muhannad Tahboush², Mohammad Adawy³, Tariq Alwada'n⁴, Sameh Ghwanmeh⁵

Computer Science Dept., The World Islamic Sciences and Education University, Amman, Jordan¹

Information System and Network Dept., The World Islamic Sciences and Education University, Amman, Jordan^{2,3}

Network and Cybersecurity Dept., Teesside University, Middlesbrough, UK⁴

Computer Science Dept., American University in the Emirates, Dubai, UAE⁵

Abstract—Anomaly detection in IoT is a hot topic in cybersecurity. Also, there is no doubt that the increased volume of IoT trading technology increases the challenges it faces. This paper explores several machine-learning algorithms for IoT anomaly detection. The algorithms used are Naïve Bayesian (NB), Support Vector Machine (SVM), Decision Tree (DT), XGBoost, Random Forest (RF), and K-nearest Neighbor (K-NN). Besides that, this research uses three techniques for feature reduction (FR). The dataset used in this study is RT-IoT2022, which is considered a new dataset. Feature reduction methods used in this study are Principal Component Analysis (PCA), Particle Swarm Optimization (PSO), and Gray Wolf Optimizer (GWO). Several assessment metrics are applied, such as Precision (P), Recall(R), F-measures, and accuracy. The results demonstrate that most machine learning algorithms perform well in IoT anomaly detection. The best results are shown in SVM with approximately 99.99% accuracy.

Keywords—Machine learning; Internet of Things (IoT); anomaly detection; feature reduction; Naïve Bayesian (NB); Support Vector Machine (SVM); Decision Tree (DT); XGBoost; Random Forest (RF); K-Nearest Neighbor (K-NN)

I. INTRODUCTION

Detecting anomalies on the Internet of Things (IoT) is a major security issue that has been investigated and studied for centuries. The Internet of Things (IoT) involves several devices capable of processing, collecting, storing data, and communicating. The adoption of the IoT brought many innovations to industries, homes, and businesses, and undoubtedly, it has improved the quality of life.

Recently, the Internet of Things (IoT) has experienced quick growth in many specific applications. Also, IoT has become a driving force for the current technology revolution. IoT captures valuable data daily, allowing individuals or users to make critical decisions. There are many applications for IoT, such as healthcare, transportation, agriculture, and others. Also, there is no doubt that IoT devices have some limitations, such as CPU, memory, and low-energy storage. IoT devices comprise several interconnected sensors, actuators, and other devices [1],[2]. A lot of research expected tremendous growth in IoT. For example, cisco predicted an average of 75.3 billion linked devices by 2025 [3], [4].

IoT devices are extremely vulnerable to cyber-security threats targeting integrity and availability, and it is necessary to prevent cyber-security accidents. Thus, a Network Intrusion

Detection System (NIDS) is needed. NIDS can detect any anomaly to protect the IoT network and the device. NIDS has the ability to monitor all traffic across the IoT network and acts as a first defense line. Also, NIDS can identify networks against intruders and suspicious activity. In addition, NIDS can examine and investigate the devices on the network [5], [6], [7].

Anomaly recognition can be divided into three categories based on the function of the training data stated as follows [2], [3], [4].:

Supervised Anomaly Detection: The normal and abnormal training datasets contain labeled cases. Thus, this methodology is about creating a predictive model for the abnormal and normal classes and then comparing both together.

Semi-supervised anomaly detection: The learning here involves only common cases of the class. Thus, anything that cannot be classified as usual is marked as abnormal.

Unsupervised anomaly detection: The training datasets will not be necessary for the methods. Thus, these methods indicate that regular cases are much more common than anomalies in the test data sets. Even if the hypothesis fails, this leads to a high false alarm rate for this practice.

This research proposes a new approach for IoT anomaly detection combined with artificial intelligence (AI) using detection mechanisms. The proposed approach combines three techniques for feature reduction (FR). Principal Component Analysis (PCA), Particle Swarm Optimization (PSO), and Gray Wolf Optimizer (GWO) were implemented for IoT cybersecurity.

Several research papers and surveys related to IoT have been proposed and published. Some of this research discusses security frameworks, privacy issues, security challenges, models, and tools [8], [9], [10], [11]. When Artificial Intelligence (AI) and the IoT combine, anomaly detection becomes more effective and reliable. AI-based anomaly detection can detect a wide range of threats. This paper will focus on machine learning (ML) algorithms and techniques for IoT security; the contribution of this paper can be reviewed in the following points:

- Using several machine learning algorithms for anomaly detection in IoT.

- Using The RT-IoT2022 proprietary dataset taken from a real-time IoT infrastructure.
- Using several up-to-date techniques for feature reduction, such as PSO, GWO, and PCA.

The rest of this paper is organized as follows: Section II will discuss previous studies related to this research. Section III will introduce machine learning algorithms for anomaly detection in an IoT environment. Section IV will discuss feature reduction and the dataset used in this paper. Section V will demonstrate experiments and results. Finally, the paper is concluded in Section VI.

II. RELATED WORK

In this section, the authors will concentrate on some of the most prevailing solutions and demonstrate several research talks about IoT anomaly discovering methods and techniques.

Ayan Chatterjee [12] demonstrates a complete survey of IoT anomaly detection methods and applications. This survey examines 64 articles among publications between 2019 and 2021. The authors explain that they witnessed a shortage of IoT anomaly detection methodologies. Also, the authors present challenges and offer a new perspective where more research is needed. Besides that, the authors show that the publication of IoT detection is still in its early stages. Finally, they present no single best generic algorithm, but several methods are specific to a particular application.

Rafique Saida [13] presents a variety of literature on anomaly detection in IoT using both ML and DL. The authors discuss various challenges in anomaly detection in IoT infrastructure. Also, this research presents an increasing number of attacks. Finally, this work summarizes the most available literature and concludes that further development of the current detection technique is needed.

Maryam Khan [14] presents a machine learning anomaly detection model for cybersecurity using the Canadian Institute for Cybersecurity (CIC) dataset. The dataset presented in this work consists of 33 types of IoT attacks divided into seven categories. Techniques used in this work are Random Forest (RF), Adaptive Boosting (AB), Logistic Regression (LR), and Neural Network (NN). RF performs 99.55% accuracy.

Edwin Omo [15] presents several machine-learning algorithms for anomaly detection. The algorithms used in this work are isolation forest, One-Class SVM, Autoencoders, and Random Forest (RF). The study also examines the performance evaluation, efficiency process, and model selection methods. Besides that, the research sheds light on the main IoT aspects.

Adel Abusitta [16] presents a deep learning-powered anomaly recognition for IoT. The proposed model is designed based on a denoising autoencoder. Also, the denoising autoencoder allows the system to obtain features. Finally, experiments were conducted using the DS2OS traffic dataset.

Bhawana Sharma [17] provides an overview of anomaly detection using both machine learning and deep learning methods. This research addresses the key issues and

challenges related to deep anomaly detection techniques in IoT.

Sahu [18] presents a supervised learning model to predict anomalies. This research uses several machine learning algorithms to predict anomalies on the 350K dataset. Two different approaches are used in this research. Also, classification algorithms were applied to the whole dataset in two different ways. The algorithms used were Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), and Artificial Neural Network (ANN). Finally, accuracy achieved an average of 99.4%.

Muhammad Inuwa [19] presents the comprehensive difficulties and challenges of cybersecurity in the context of IoT. This research uses machine learning (ML) methods to detect cyber anomalies within IoT systems. The algorithms used were Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Tree (DT), Logistic Regression (LR), and K-Nearest Neighbors (k-NN). Results demonstrate that ANN performs better than other models.

Abebe Diro [20] aim to provide a deep review of available works in anomaly discovery based on machine learning methods for IoT protection. This work indicates that blockchain-based anomaly detection can be effective. The future work of this research is to provide the implementation of a blockchain-based anomaly detection system.

A. Pathak [21] addresses the tampering of IoT security sensors in an office environment. Data is collected from real-life settings, and machine learning is applied to detect sensor tampering. The classification accuracy of the proposed model is 91.62%, with the lowest false positive rate.

Grace Hannah [22] explores several ML algorithms for anomaly discovery. This research explores supervised, unsupervised, and semi-supervised techniques. Also, the authors discuss the challenges and difficulties in implementing these algorithms in an IoT environment. Preprocessing techniques are examined. Besides that, this research demonstrates a case study on anomaly discovery in an IoT-based temperature monitoring system using a Gaussian Mixture Model (GMM). Precision, recall, and F1 score are used for evaluation.

III. MACHINE LEARNING ALGORITHMS FOR IOT ANOMALY DETECTION

Machine Learning (ML) algorithms can be used for different objectives and objectives and can impact every part of our lives. ML algorithms can be employed for pattern recognition, speech recognition, fraud detection, spam detection, phishing, and others. Also, machine learning procedures are used for prediction and classification, such as Decision Tree (DT), Random Forest (RF), Support Vector Machines (SVM), K-Nearest Neighbor (k-NN), Naïve Bayes Theorem (NB), K-Mean Clustering, Artificial Neural Network (ANN), and others. [23] [24], [25], [26], [27], [28], [29], [30].

Machine learning can be used for anomaly detection in IoT environments. The noun anomaly comes from the Greek word anomolia, meaning “irregular” which means that something is unusual if compared to similar things around it [31]. This

paper will introduce several machine learning algorithms in IoT anomaly detection. An anomaly in IoT is a pattern or series of samples in the IoT network that is different from a normal pattern. Also, anomaly detection can be defined as suspicious activity that falls outside normal patterns or behavior. Generally, anomalies can be divided into three categories: global outliers, contextual, and collective outliers [32], [33], [34], [35].

IV. FEATURE REDUCTION AND SELECTION

Feature reduction or dimensionality reduction is the process of reducing the number of features in a dataset. Minimizing the number of features in a dataset is very important since the number of features could be huge. Also, Reducing the number of features could be useful and retaining the most helpful information. Besides that, reducing features means reducing processing time in CPU, memory usage, and other resources [26], [27], [28]. In other words, feature reductions mean assigning a weight to each feature to decide how important they are. On the other hand, Feature selection means selecting the most powerful features in the training phase. In summary, if feature reduction is done properly, this means that selecting a partial subset of features could be enough to represent all features. In this paper, the authors will use the Principal Component Analysis (PCA), Grey Wolf Optimizer (GWO), and Particle Swarm Optimizer (PSO) [28], [29], [30], [33].

A. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a feature extraction method and is often used to reduce a higher-dimensional feature space to a lower-dimensional feature space. PCA is a statistical method that is employed to convert a set of possibly correlated variables into linearly unrelated variables known as principal components. The main objective of PCA is to capture the maximum variance available in the dataset with the fewest number of principal components. The transformation is defined mathematically as [25]:

$$\Sigma = \frac{1}{m-1} \sum_{i=1}^m (xi - \mu)(xi - \mu)^T \quad (1)$$

Where:

Σ : Covariance

xi : Data point

μ : Mean Vector

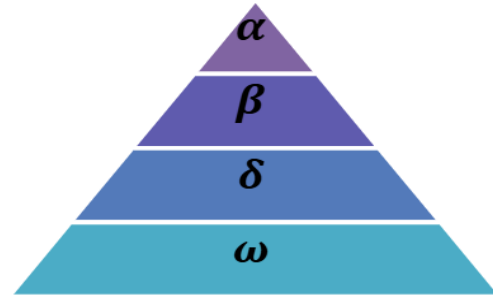
m : Number of data points.

B. Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) is a powerful meta-heuristics optimization algorithm. This algorithm is inspired by natural swarm activities, such as that of fish and birds. PSO can be used to find the optimal values for specific parameters of a given system. In PSO, particles are moved according to a simple formula. Besides that, swarms move through the search space in order to find the optimal value. Every time a better position is found, movement is done. This process is repeated until finding the optimal solution [36], [37], [38], [39], [40].

C. Gray Wolf Optimizer (GWO)

The Gray Wolf Optimizer (GWO) algorithm is a population-based meta-heuristics algorithm that simulates the leadership hierarchy and hunting mechanism of grey wolves in nature Fig. 1 [41] [42].



Hierarchy of grey wolf (dominance decreases from top down)

Fig. 1. Wolves' hierarchy.

Alpha wolves (α) wolf is the dominant, and his orders must be followed. Beta wolves (β) are subordinate wolves, which support alpha in decision-making. Delta wolves (δ) have to submit to alpha and beta. Omega wolves (ω) are the least important individuals in the pack [41], [42], [43], [44].

D. Dataset

The RT-IoT-2022 dataset, this dataset is proprietary and derived from a real-time IoT infrastructure. The RT-IoT-2022 provides comprehensive resources and a diverse range of IoT network machines. This dataset contains both normal and adversarial network behaviors. The RT-IoT-2022 contains 123117 instances and 83 features. Table I summarizes the RT-IoT-2022 dataset [45].

TABLE I. RT-IOT-2022 DATASET

No	Service	No of instances	Patterns
1	MQTT	4146	Normal Patterns
2	Thing_speak	8108	Normal Patterns
3	Wipro_bulb	253	Normal Patterns
	Total	12507	
4	ARP_poisoning	7750	Attacks patterns
5	DDOS_Slowloris	534	Attacks patterns
6	DOS_SYN_Hping	94659	Attacks patterns
7	Metasploit_Brute_Force_SSH	37	Attacks patterns
8	NMAP_FIN_SCAN	28	Attacks patterns
9	NMAP_OS_DETECTION	2000	Attacks patterns
10	NMAP_TCP_scan	1002	Attacks patterns
11	NMAP_UDP_SCAN	2590	Attacks patterns
12	NMAP_XMAS_T+REE_SCAN	2010	Attacks patterns
	Total	110610	

V. EXPERIMENTS AND RESULTS

This section will display the authors' experiments and results. Also, it will display evaluation matrices and important features. This study also uses the Anaconda platform (Python) and MATLAB 2020a. Finally, experiments were done using a Dell Machine, 11th Gen -1165G7 @ 2.80GHz, RAM 32 GB, Windows 11.

A. Experimental Metrics

In machine learning, there are several criteria for evaluation, such as accuracy, precision, recall, F-measure, True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). This is demonstrated in Table II and Eq. (2) to (8).

TABLE II. MATRIX OF CONFUSION

		Prediction.	
		Normal.	Phishing
Act.	Normal.	x (TP)	y (FN)
	Phishing	z (FP)	w (TN)

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

$$\text{TPR} = x/(x+y) \quad (3)$$

$$\text{FPR} = z/(z+w) \quad (4)$$

$$\text{FNR} = y/(x+y) \quad (5)$$

$$P = TP / (TP + FP) \quad (6)$$

$$R = TP / (TP + FN) \quad (7)$$

$$\text{F-Measure} = 2 * P * R / (P + R) \quad (8)$$

B. Experimental Results

In this section. The authors will demonstrate the results of feature reduction and selection using PCA, PSO, and GWO. PCA is evaluated using 10, 20, 30, 40, 50, 60 and 70 features.

Feature Reduction (FR) is done by using PCA, PSO and GWO. The PSO and GWO algorithms are executed

independently for (10) iterations; then, the number and the name of the features are written. Then, the most important features of each algorithm are determined and picked for the classification stage. The testing part of the dataset represents only 20% of the dataset, meaning only 24624 instances.

Fig. 2 represents the results using Fine Tree without any feature reduction (PCA Disabled) using MATLAB 2020a. The figure demonstrated a good result, but the high number of features required extensive CPU and RAM resources.

The results of the experiments using feature reduction are demonstrated in Tables III-VIII. Most of the algorithm's performance is highly accepted. Also, FR techniques are very helpful since reducing the number of features from 83 to any number will reduce processing time and memory storage.

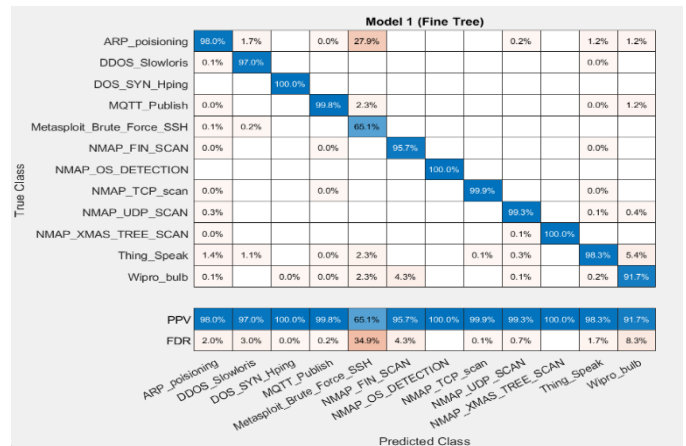


Fig. 2. Fine tree results (PCA disabled).

The above table shows that feature reduction using NB, (PCA-40) provides the best accuracy and optimal values of TP, TN, and FP compared with other types of feature reduction.

The above table shows that feature reduction using SVM, (PCA-50, PCA-60, and PCA-70) provides the best accuracy and optimal values of TP, TN, and FP compared with other types of feature reduction.

TABLE III. NAÏVE BAYESIAN EXPERIMENTS

FR	TP	TN	FP	FN	Pr.	Re.	F-Me.	Acc.
PCA-10	21324	1052	1501	747	93.42%	96.62%	94.99%	90.87%
PCA-20	21281	995	1558	790	93.18%	96.42%	94.77%	90.46%
PCA-30	21211	963	1590	860	93.03%	96.10%	94.54%	90.05%
PCA-40	21352	1055	1498	719	93.44%	96.74%	95.06%	91.00%
PCA-50	21345	886	1667	726	92.76%	96.71%	94.69%	90.28%
PCA-60	21369	834	1719	702	92.55%	96.82%	94.64%	90.17%
PCA-70	21385	793	1760	686	92.40%	96.89%	94.59%	90.07%
GWO-55	21395	800	1750	679	92.44%	96.92%	94.63%	90.14%
PSO-58	21400	815	1740	669	92.48%	96.97%	94.67%	90.22%

TABLE IV. SUPPORT VECTOR MACHINE EXPERIMENTS

FR	TP	TN	FP	FN	Pr.	Re.	F-Me.	Acc.
PCA-10	21874	2485	68	197	99.69%	99.11%	99.40%	98.92%
PCA-20	22039	2524	29	32	99.87%	99.86%	99.86%	99.75%
PCA-30	22067	2552	2	3	99.99%	99.99%	99.99%	99.98%
PCA-40	22067	2552	2	3	99.99%	99.99%	99.99%	99.98%
PCA-50	22070	2551	2	1	99.99%	100.00%	99.99%	99.99%
PCA-60	22070	2552	1	1	100.00%	100.00%	100.00%	99.99%
PCA-70	22070	2552	1	1	100.00%	100.00%	100.00%	99.99%
GWO-55	22040	2520	35	29	99.84%	99.87%	99.86%	99.74%
PSO-58	22041	2519	36	28	99.84%	99.87%	99.86%	99.74%

TABLE V. DECISION TREE EXPERIMENTS

FR	TP	TN	FP	FN	Pr.	Re.	F-Me.	Acc.
PCA-10	22053	2538	15	18	99.93%	99.92%	99.93%	99.87%
PCA-20	22052	2542	11	19	99.95%	99.91%	99.93%	99.88%
PCA-30	22053	2537	16	18	99.93%	99.92%	99.92%	99.86%
PCA-40	22056	2538	15	15	99.93%	99.93%	99.93%	99.88%
PCA-50	22055	2532	21	16	99.90%	99.93%	99.92%	99.85%
PCA-60	22056	2537	16	15	99.93%	99.93%	99.93%	99.87%
PCA-70	22052	2538	15	19	99.93%	99.91%	99.92%	99.86%
GWO-55	22050	2536	17	21	99.92%	99.90%	99.91%	99.85%
PSO-58	22048	2534	19	23	99.91%	99.90%	99.90%	99.83%

The above table shows that feature reduction using DT, (PCA-20, PCA-40) provides the best accuracy and optimal values of TP, TN, and FP compared with other types of feature reduction.

The above table shows that feature reduction using XGBoost, (GWO-55, PSO-58) provides the best accuracy and

optimal values of TP, TN, and FP compared with other types of feature reduction.

The above table shows that feature reduction using RF, (PCA-30) provides the best accuracy and optimal values of TP, TN, and FP compared with other types of feature reduction.

TABLE VI. XGBOOST EXPERIMENTS

FR	TP	TN	FP	FN	Pr.	Re.	F-Me.	Acc.
PCA-10	22064	2538	15	7	99.93%	99.97%	99.95%	99.91%
PCA-20	22070	2543	10	1	99.95%	100.00%	99.98%	99.96%
PCA-30	22069	2544	9	2	99.96%	99.99%	99.98%	99.96%
PCA-40	22069	2547	6	2	99.97%	99.99%	99.98%	99.97%
PCA-50	22069	2545	8	2	99.96%	99.99%	99.98%	99.96%
PCA-60	22069	2546	7	2	99.97%	99.99%	99.98%	99.96%
PCA-70	22069	2547	6	2	99.97%	99.99%	99.98%	99.97%
GWO-55	22070	2548	4	2	99.98%	99.99%	99.99%	99.98%
PSO-58	22070	2549	3	2	99.99%	99.99%	99.99%	99.98%

TABLE VII. RANDOM FOREST EXPERIMENTS

FR	TP	TN	FP	FN	Pr.	Re.	F-Me.	Acc.
PCA-10	22064	2533	20	7	99.91%	99.97%	99.94%	99.89%
PCA-20	22064	2538	15	7	99.93%	99.97%	99.95%	99.91%
PCA-30	22066	2542	11	5	99.95%	99.98%	99.96%	99.94%
PCA-40	22064	2537	16	7	99.93%	99.97%	99.95%	99.91%
PCA-50	22063	2538	15	8	99.93%	99.96%	99.95%	99.91%
PCA-60	22063	2534	19	8	99.91%	99.96%	99.94%	99.89%
PCA-70	22061	2537	16	10	99.93%	99.95%	99.94%	99.89%
GWO-55	22060	2541	14	9	99.94%	99.96%	99.95%	99.91%
PSO-58	22061	2540	12	11	99.95%	99.95%	99.95%	99.91%

TABLE VIII. K-NEAREST NEIGHBOR EXPERIMENTS

FR	TP	TN	FP	FN	Pr.	Re.	F-Me.	Acc.
PCA-10	22060	2535	18	11	99.92%	99.95%	99.93%	99.88%
PCA-20	22067	2533	20	4	99.91%	99.98%	99.95%	99.90%
PCA-30	22068	2544	9	3	99.96%	99.99%	99.97%	99.95%
PCA-40	22069	2544	9	2	99.96%	99.99%	99.98%	99.96%
PCA-50	22066	2544	9	5	99.96%	99.98%	99.97%	99.94%
PCA-60	22066	2544	9	5	99.96%	99.98%	99.97%	99.94%
PCA-70	22068	2542	11	3	99.95%	99.99%	99.97%	99.94%
GWO-55	22070	2540	9	5	99.96%	99.98%	99.97%	99.94%
PSO-58	22072	2538	11	3	99.95%	99.99%	99.97%	99.94%

The above table shows that feature reduction using KNN, (PCA-40) provides the best accuracy and optimal values of TP, TN, and FP compared with other types of feature reduction.

As demonstrated in the above tables. The performance of machine learning algorithms with feature reduction techniques is highly acceptable. Having too many processing features makes the ML model complex. There is no doubt that reducing the number of features has a lot of advantages, such as reducing time, improving computational efficiency, and preventing overfitting.

Fig. 3 and Fig. 4 show the accuracy of the machine-learning algorithms used in this paper. The figures demonstrated that the accuracy results are highly acceptable, especially with the number of features selected.

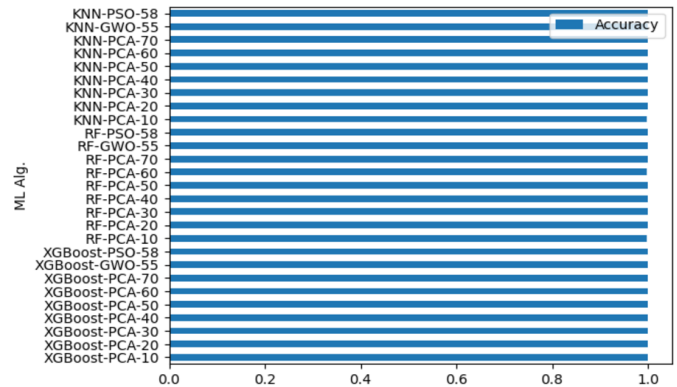


Fig. 4. KNN, RF, and XGBoost algorithms.

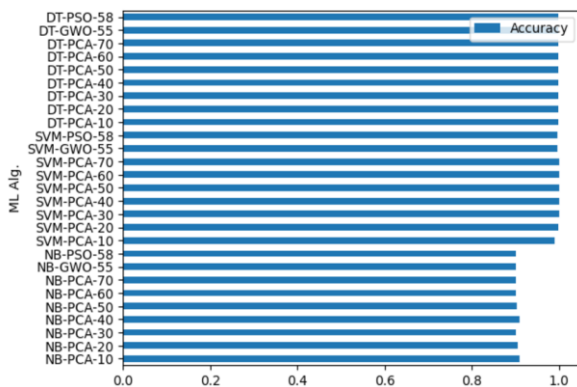


Fig. 3. DT, SVM, and NB algorithms.

VI. CONCLUSION AND FUTURE WORKS

The Internet of Things (IoT) or “Smart Objects” refers to physical devices embedded with sensors, software, and network connectivity. IoT devices can be used in smart homes, smart cities, and complex industries. IoT enables smart devices to communicate with each other and with the Internet. In the last decades, IoT devices have faced several threats and difficulties. This paper demonstrates several machine learning algorithms used in anomaly detection in IoT environments. This paper also uses PCA, GWO, and PSO as feature-reduction techniques. Several criteria are used for evaluation, such as precision, recall, F-measure, and accuracy. Most of the algorithms show excellent performance except the Naïve Bayesian. The support vector machines (SVM) show the best results with 99.99 accuracy with PCA-60 and PCA-70.

ACKNOWLEDGMENT

The researchers would like to provide a special thanks to the editor and reviewers for their time in reviewing the manuscript and overall suggestions to improve the manuscript. In addition, they are very grateful to WISE University.

REFERENCES

- [1] E. Gyamfi and A. Jurcut, "Intrusion Detection in Internet of Things Systems: A Review on Design Approaches Leveraging Multi-Access Edge Computing, Machine Learning, and Datasets," *Sensors*, vol. 22, no. 10, 2022, doi: 10.3390/s22103744.
- [2] M. Stoyanova, Y. Nikoloudakis, S. Panagiotakis, E. Pallis, and E. K. Markakis, "A Survey on the Internet of Things (IoT) Forensics: Challenges, Approaches, and Open Issues," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1191–1221, 2020, doi: 10.1109/COMST.2019.2962586.
- [3] A. Yastrebova, R. Kirichek, Y. Koucheryavy, A. Borodin, and A. Koucheryavy, "Future Networks 2030: Architecture & Requirements," *2018 10th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, pp. 1–8, 2018, [Online]. Available: <https://api.semanticscholar.org/CorpusID:59601484>
- [4] A. Jurcut, T. Niculcea, P. Ranaweera, and N.-A. Le-Khac, "Security Considerations for Internet of Things: A Survey," *CoRR*, vol. abs/2006.10591, 2020, [Online]. Available: <https://arxiv.org/abs/2006.10591>
- [5] I. Butun, S. D. Morgera, and R. Sankar, "A Survey of Intrusion Detection Systems in Wireless Sensor Networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 266–282, 2014, doi: 10.1109/SURV.2013.050113.00191.
- [6] M. A. Alsoufi *et al.*, "Anomaly-Based Intrusion Detection Systems in IoT Using Deep Learning: A Systematic Literature Review," *Applied Sciences*, vol. 11, no. 18, 2021, doi: 10.3390/app11188383.
- [7] L. Njilla, L. Pearlstein, X.-W. Wu, A. Lutz, and S. Ezekiel, "Internet of Things Anomaly Detection using Machine Learning," in *2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, 2019, pp. 1–6. doi: 10.1109/AIPR47015.2019.9174569.
- [8] M. Ammar, G. Russello, and B. Crispo, "Internet of Things: A survey on the security of IoT frameworks," *Journal of Information Security and Applications*, vol. 38, pp. 8–27, 2018, doi: <https://doi.org/10.1016/j.jisa.2017.11.002>.
- [9] Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, "A Survey on Security and Privacy Issues in Internet-of-Things," *IEEE Internet Things J.*, vol. 4, no. 5, pp. 1250–1258, 2017, doi: 10.1109/JIOT.2017.2694844.
- [10] M. Sain, Y. J. Kang, and H. J. Lee, "Survey on security in Internet of Things: State of the art and challenges," in *2017 19th International Conference on Advanced Communication Technology (ICACT)*, 2017, pp. 699–704. doi: 10.23919/ICACT.2017.7890183.
- [11] R. Benabdessalem, M. Hamdi, and T. Kim, "A Survey on Security Models, Techniques, and Tools for the Internet of Things," *2014 7th International Conference on Advanced Software Engineering and Its Applications*, pp. 44–48, 2014, [Online]. Available: <https://api.semanticscholar.org/CorpusID:18825070>
- [12] A. Chatterjee and B. S. Ahmed, "IoT anomaly detection methods and applications: A survey," *Internet of Things*, vol. 19, p. 100568, 2022, doi: <https://doi.org/10.1016/j.iot.2022.100568>.
- [13] S. H. Rafique, A. Abdallah, N. S. Musa, and T. Murugan, "Machine Learning and Deep Learning Techniques for Internet of Things Network Anomaly Detection—Current Research Trends," *Sensors*, vol. 24, no. 6, 2024, doi: 10.3390/s24061968.
- [14] M. M. Khan and M. Alkhatami, "Anomaly detection in IoT-based healthcare: machine learning for enhanced security," *Sci Rep*, vol. 14, no. 1, p. 5872, 2024, doi: 10.1038/s41598-024-56126-x.
- [15] E. Omol, L. Mburu, and D. Onyango, "Anomaly Detection In IoT Sensor Data Using Machine Learning Techniques For Predictive Maintenance In Smart Grids," *International Journal of Science, Technology & Management*, vol. 5, no. 1, pp. 201–210, Jan. 2024, doi: 10.46729/ijstm.v5i1.1028.
- [16] A. Abusitta, G. H. de Carvalho, O. Abdel Wahab, T. Halabi, B. C. M. Fung, and S. Al Mamoori, "Deep learning-enabled anomaly detection for IoT systems," *SSRN Electron. J.*, 2022.
- [17] B. Sharma, L. Sharma, and C. Lal, "Anomaly Detection Techniques using Deep Learning in IoT: A Survey," in *2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE)*, 2019, pp. 146–149. doi: 10.1109/ICCIKE47802.2019.9004362.
- [18] N. K. Sahu and I. Mukherjee, "Machine Learning based anomaly detection for IoT Network: (Anomaly detection in IoT Network)," in *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, 2020, pp. 787–794. doi: 10.1109/ICOEI48184.2020.9142921.
- [19] M. M. Inuwa and R. Das, "A comparative analysis of various machine learning methods for anomaly detection in cyber attacks on IoT networks," *Internet Things*, vol. 26, p. 101162, 2024, [Online]. Available: <https://api.semanticscholar.org/CorpusID:268402373>
- [20] A. Diro, N. Chilamkurti, V.-D. Nguyen, and W. Heyne, "A Comprehensive Study of Anomaly Detection Schemes in IoT Networks Using Machine Learning Algorithms," *Sensors*, vol. 21, no. 24, 2021, doi: 10.3390/s21248320.
- [21] A. K. Pathak, S. Saguna, K. Mitra, and C. Åhlund, "Anomaly Detection using Machine Learning to Discover Sensor Tampering in IoT Systems," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6. doi: 10.1109/ICC42927.2021.9500825.
- [22] Dr. J. G. Hannah, Dr. A. S. D. Murthy, Dr. G. Kalnoor, M. Vetrivelan, and Dr. M. S. Nidhya, "Machine Learning Algorithms for Anomaly Detection in IoT Networks," *Migration Letters*, vol. 20, no. S13, pp. 560–565, Dec. 2023, [Online]. Available: <https://migrationletters.com/index.php/ml/article/view/6728>
- [23] R. Mahajan and I. Siddavatam, "Phishing website detection using machine learning algorithms," *Int. J. Comput. Appl.*, vol. 181, no. 23, pp. 45–47, Oct. 2018.
- [24] D. T. Mosa, M. Y. Shams, A. A. Abohany, E.-S. M. El-kenawy, and M. Thabet, "Machine Learning Techniques for Detecting Phishing URL Attacks," *Computers, Materials and Continua*, vol. 75, no. 1, pp. 1271–1290, 2023, doi: <https://doi.org/10.32604/cmc.2023.036422>.
- [25] M. Altin and A. Cakir, "Exploring the influence of dimensionality reduction on anomaly detection performance in multivariate time series," 2024.
- [26] F. Abbasi, M. Naderan, and S. E. Alavi, "Anomaly detection in Internet of Things using feature selection and classification based on Logistic Regression and Artificial Neural Network on N-BaIoT dataset," in *2021 5th International Conference on Internet of Things and Applications (IoT)*, 2021, pp. 1–7. doi: 10.1109/IoT52625.2021.9469605.
- [27] A. G. Ayad, N. A. Sakr, and N. A. Hikal, "A hybrid approach for efficient feature selection in anomaly intrusion detection for IoT networks," *J Supercomput*, vol. 80, no. 19, pp. 26942–26984, 2024, doi: 10.1007/s11227-024-06409-x.
- [28] A. H. Mohammad, "Intrusion Detection Using a New Hybrid Feature Selection Model," *Intelligent Automation & Soft Computing*, vol. 30, no. 1, pp. 65–80, 2021, doi: 10.32604/iasc.2021.016140.
- [29] A. Mandadi, S. Boppana, V. Ravella, and R. Kavitha, "Phishing Website Detection Using Machine Learning," in *2022 IEEE 7th International conference for Convergence in Technology (I2CT)*, 2022, pp. 1–4. doi: 10.1109/I2CT54291.2022.9824801.
- [30] A. H. Mohammad, T. Alwada'n, O. Almomani, S. Smadi, and N. ElOmari, "Bio-inspired Hybrid Feature Selection Model for Intrusion Detection," *Computers, Materials and Continua*, vol. 73, no. 1, pp. 133–150, 2022, doi: <https://doi.org/10.32604/cmc.2022.027475>.
- [31] "<https://www.vocabulary.com/dictionary/anomalyWebsite>".
- [32] A. H. Mohammad, S. Smadi, and T. Alwada'n, "Email Filtering Using Hybrid Feature Selection Model," *CMES - Computer Modeling in Engineering and Sciences*, vol. 132, no. 2, pp. 435–450, 2022, doi: <https://doi.org/10.32604/cmcs.2022.020088>.
- [33] S. Alrefaai, G. Özdemir, and A. Mohamed, "Detecting Phishing Websites Using Machine Learning," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2022, pp. 1–6. doi: 10.1109/HORA55278.2022.9799917.

- [34] M. Tahboush, A. Hamdan, F. Alzobi, M. Husni, and M. Adawy, "NTDA: The mitigation of denial of service (DoS) cyberattack based on network traffic detection approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 3, 2024.
- [35] A. Hamdan, M. Tahboush, M. Adawy, T. Alwada'n, S. Ghwanmeh, and M. Husni, "Phishing detection using grey wolf and particle swarm optimizer," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 14, no. 5, p. 5961, Oct. 2024.
- [36] J. A. W. A. S. Saeed M. Alshahrani Nayyar Ahmed Khan, "URL Phishing Detection Using Particle Swarm Optimization and Data Mining," *Computers, Materials & Continua*, vol. 73, no. 3, pp. 5625–5640, 2022, doi: 10.32604/cmc.2022.030982.
- [37] W. Ali and S. Malebary, "Particle Swarm Optimization-Based Feature Weighting for Improving Intelligent Phishing Website Detection," *IEEE Access*, vol. 8, pp. 116766–116780, 2020, doi: 10.1109/ACCESS.2020.3003569.
- [38] F. Marini and B. Walczak, "Particle swarm optimization (PSO). A tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 149, pp. 153–165, 2015, doi: <https://doi.org/10.1016/j.chemolab.2015.08.020>.
- [39] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95 - International Conference on Neural Networks*, 1995, pp. 1942–1948 vol.4. doi: 10.1109/ICNN.1995.488968.
- [40] K. Ishaque, Z. Salam, M. Amjad, and S. Mekhilef, "An Improved Particle Swarm Optimization (PSO)-Based MPPT for PV With Reduced Steady-State Oscillation," *IEEE Trans Power Electron*, vol. 27, no. 8, pp. 3627–3638, 2012, doi: 10.1109/TPEL.2012.2185713.
- [41] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey Wolf Optimizer," *Advances in Engineering Software*, vol. 69, pp. 46–61, 2014, doi: <https://doi.org/10.1016/j.advengsoft.2013.12.007>.
- [42] J.-S. Wang and S.-X. Li, "An Improved Grey Wolf Optimizer Based on Differential Evolution and Elimination Mechanism," *Sci Rep*, vol. 9, no. 1, p. 7181, 2019, doi: 10.1038/s41598-019-43546-3.
- [43] E. M. R. Devi and R. C. Suganthe, "Feature selection in intrusion detection grey wolf optimizer," *Asian J. Res. Soc. Sci. Humanit.*, vol. 7, no. 3, p. 671, 2017.
- [44] Q. M. Alzubi, M. Anbar, Z. N. M. Alqattan, M. A. Al-Betar, and R. Abdullah, "Intrusion detection system based on a modified binary grey wolf optimisation," *Neural Comput Appl*, vol. 32, pp. 6125–6137, 2019, [Online]. Available: <https://api.semanticscholar.org/CorpusID:128021795>
- [45] B. & N. R. (2023). R.-I. [Dataset]. U. M. L. Repository. S., "https://archive.ics.uci.edu/dataset/942/rt-iot2022".

Network Security Based on GCN and Multi-Layer Perception

Wei Yu*, Huitong Liu, Yu Song, Jiaming Wang

Guangzhou Bureau, EHV Power Transmission Company of China, Southern Power Grid, Guangzhou, 510663, China

Abstract—With the continuous progress of network technology, network security has become a critical issue at present. There are already many network security intrusion detection models, but these detection models still have problems such as low detection accuracy and long interception time of intrusion information. To address these drawbacks, this study utilizes graph convolutional network to optimize multi-layer perceptron. An optimization algorithm based on multi-layer perceptron is innovatively proposed to construct an intrusion detection model. Comparative experiments are conducted on the improved algorithm. The accuracy of the algorithm was 0.98, the F1 value was 0.97, and the detection time was 1.1s. The overall performance was much better than comparison algorithms. Subsequently, the intrusion detection model was applied to network security detection. The detection time was 0.1s, the accuracy was 0.98, and the overall performance outperformed other comparison algorithms. The results demonstrate that the intrusion detection method on the basis of optimized multi-layer perceptron can enhance the detection ability of illegal intrusion information. This study optimizes the performance of detecting illegal network intrusion information, providing a theoretical basis for further development of network security. However, the types of intrusion information in this study are limited and there is still uncertainty. In the future, data augmentation techniques can be used to oversample minority class samples, synthesize new minority class samples, expand sample size, increase detection information, and improve the overall detection performance of the model.

Keywords—Network security; graph convolutional network; multi-layer perceptron; intrusion detection model

I. INTRODUCTION

In the current era of rapid digital development, network security is becoming increasingly prominent, which is an important challenge that countries, enterprises, and individuals must face [1]. Affected by the popularity of information technology and the Internet, network attacks are constantly evolving, and the traditional security measures have been difficult to deal with. Therefore, exploring new methods for network security protection is of great significance [2]. Many scholars have conducted research on network intrusion detection models. For example, Fu et al. proposed an intrusion detection model based on attention mechanism to enhance the performance of traditional network firewalls and data encryption methods. Through experimental verification, the model achieved a detection accuracy of 90.73% [3]. In addition, Hnamte et al. designed a network intrusion detection model based on deep neural networks for network attacks. Then, the model was applied to detect in practical situations. The results showed that the model could detect most of the intrusion information in the network [4]. In recent years, Graph

Convolutional Network (GCN) has shown strong feature extraction and relationship learning capabilities in multiple fields. Especially when dealing with non-Euclidean structured data, it has significant advantages [5]. Therefore, multiple scholars have applied it to network security protection. Diao et al. developed a spatiotemporal multi-scale GCN security model to improve the security of network data in vehicle prediction. After using this network security protection model, the security of network data in vehicle prediction was significantly improved [6]. To optimize the intrusion detection performance of labeled IoT networks, Deng et al. developed a GCN on the basis of flow topology. Comparative experiments on this model demonstrated that the intrusion detection accuracy of the GCN based on flow topology for labeled IoT networks was 92.31%, significantly better than other traditional methods [7]. Afterwards, Al-Ibraheemi et al. built an intrusion detection method on the basis of GCN and deep reinforcement learning algorithms to response the insufficient performance of intrusion detection models in software defined networks. The accuracy of this intrusion detection model was enhanced by 15.32% than the traditional intrusion detection model [8].

Meanwhile, Multi-layer Perceptron (MLP), as a classic deep learning model, also performs well in dealing with linearly inseparable problems [9]. Therefore, many scholars have also applied it to network security protection. Specifically, Shewale et al. designed an intrusion detection approach on the basis of MLP and Long Short-Term Memory Network (LSTM) to improve the network security. Comparative experiments showed that the intrusion detection accuracy was 91.83%, significantly better than traditional models [10]. In addition, to address the difficulty of detecting distributed denial of service attacks, Najjar et al. designed a hybrid algorithm based on MLP and random forest. Comparative experiments were conducted on a distributed denial of service attack dataset. It was found that the detection accuracy was 93.85% [11].

The above research indicates that in the field of network security, although some research have attempted to apply advanced technologies such as GCN and MLP, there are still some drawbacks. At present, the research mainly focuses on using machine learning frameworks for network intrusion detection and abnormal behavior recognition, but these methods ignore complex relationships between network data, resulting in limited detection accuracy and efficiency. In addition, existing research lacks sufficient flexibility and adaptability in dealing with constantly changing network threats. Therefore, this study designs a network security protection method on the basis of GCN and MLP. This method aims to combine the powerful relationship learning ability of

GCN with the nonlinear processing ability of MLP to process complex network data. At the same time, the GCN algorithm is used to optimize the initial parameters in MLP, improve its flexibility and generalization ability, reduce the impact of complex data on detection results in previous intrusion detection models, and more effectively identify and defend against network attacks. The innovation of the research lies in the organic combination of GCN and MLP, forming a new network security protection framework. This framework can not only handle complex network relational data, but also adaptively learn and respond to constantly changing network threats. It is expected to provide a new and more effective technological means for network security, contributing to building a more secure and reliable network environment. The contribution of this study is to timely detect abnormal information in the network through the GCN-MLP intrusion detection model, timely identify potential security threats, and reduce the damage caused by network attacks. This model promptly prevents malicious attackers from invading, protects the secure operation of networks or systems, and ensures the integrity and confidentiality of data information in the network. This model ensures that users or processes use system resources according to prescribed permissions, preventing resources from being illegally occupied.

The article is divided into five sections for discussion. Section II mainly covers network security related content and research on MLP and GCN algorithms. Section III construct an network intrusion detection model based on GCN and MLP algorithms. Section IV analyzes the effectiveness of the proposed intrusion detection model. Section V summarizes the entire text.

II. METHODS AND MATERIALS

A. Multi-Layer Perceptron Optimization Integrating Graph Convolutional Network

At present, people are paying more attention to network security issues, and there are also more network information intrusion detection models. However, these models still have problems such as false positives and missed detection [12]. MLP is a deep learning algorithm based on feedforward neural networks, which is composed of multiple neural structures. This

algorithm has strong representation and generalization capabilities, which can process various complex data, reducing the false detection rate of dangerous intrusion detection [13]. Fig. 1 displays the basic structure of the MLP.

From Fig. 1, the perceptron contains input and output layers. The perceptron allocates weights and assigns values to the input vector, then sums up the calculated data, and iteratively updates the weights until the error is reduced to the allowable range. The obtained values are then outputted [14]. MLP introduces a Hidden Layer (HL) based on single-layer neural network, making the neural network have multiple layers. MLP can adjust the number and dimensions of hidden layers, input layers, and output layers as necessary. Each node in the HL is a perceptron, and each perceptron contains some parameters. These nodes in the HL are all fully connected, that is, the previous node output is connected together as the next layer node input. The output result of the HL is shown in Eq. (1).

$$H = XW_h + b_h \quad (1)$$

In Eq. (1), H represents the output result of the HL. X signifies the given sample. W_h represents the weight of the HL. b_h signifies the deviation coefficient of the HL. If it is a single HL, the output of HL is shown in Eq. (2).

$$O = HW_0 + b_0 \quad (2)$$

In Eq. (2), W_0 signifies the weight of the output layer. b_0 represents the deviation coefficient of the output layer. Eq. (1) and (2) are combined to obtain the input of the output layer, as displayed in Eq. (3).

$$O = (XW_h + b_h)W_0 + b_0 = XW_hW_0 + b_hW_0 + b_0 \quad (3)$$

In equation (3), the weight coefficient of the output layer is changed to W_hW_0 . The deviation coefficient is changed to $b_hW_0 + b_0$. The ReLu activation function is introduced to perform nonlinear function transformation on hidden variables, making them the input of the next fully connected layer. The ReLu activation function is displayed in Eq. (4).

$$ReLu(x) = \max(x, 0) \quad (4)$$

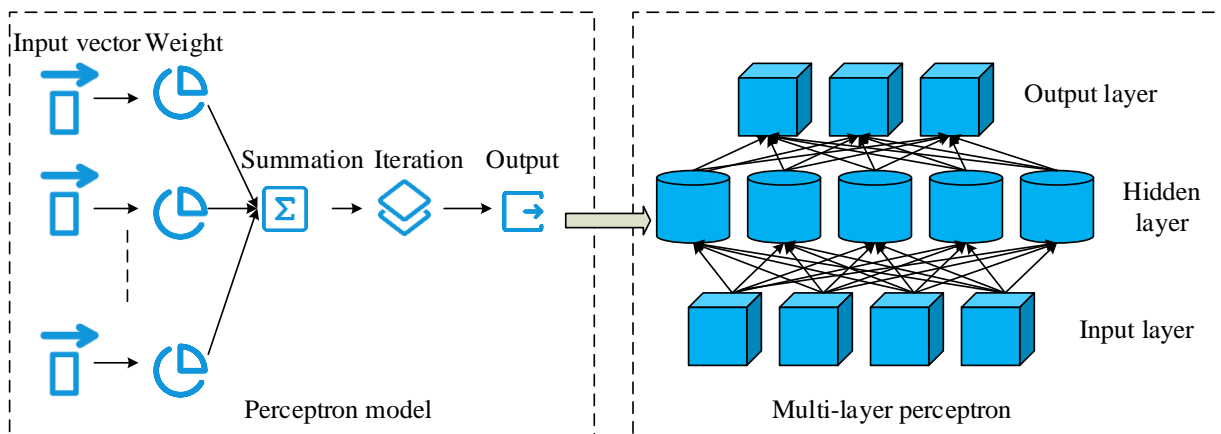


Fig. 1. Basic structure of multi-layer perceptron.

In Eq. (4), x signifies the input sample. The output expression of the MLP combined with the activation function is displayed in Eq. (5).

$$\begin{cases} H = R(XW_h + b_h) \\ O = HW_0 + b_0 \end{cases} \quad (5)$$

In Eq. (5), R represents the activation function ReLu. Afterwards, the error information is computed in Eq. (6).

$$E_i = \sum_{i=1}^n \frac{1}{2} (\tilde{y}_i - y_i)^2 \quad (6)$$

In Eq. (6), E_i represents the prediction error of the i -th output unit. \tilde{y}_i signifies the predicted value of the i -th output unit. y_i is the i -th output unit. n signifies the number of neurons in the output layer. The influence of weights on the overall error is shown in Eq. (7).

$$\frac{\partial E}{\partial w_j} = \frac{\partial E}{\partial y_l} \cdot \frac{\partial y_l}{\partial s_{y_l}} \cdot \frac{\partial s_{y_l}}{\partial w_j} \quad (7)$$

In Eq. (7), s_{y_l} signifies the weighted sum of input y_i . w_j signifies the weight of the j -th HL. The weight value is updated, as shown in Eq. (8).

$$w_j^+ = w_j - \eta \frac{\partial E}{\partial w_j} \quad (8)$$

In Eq. (8), η represents the learning rate. The above is the calculation method of MLP, which updates weights to iterate continuously. Finally, the error is reduced to the minimum allowable range. However, the training time is long, the number of calculated parameters is too large, and over-fitting is prone to occur, which can affect the detection accuracy and efficiency. The GCN algorithm has data normalization, small parameter size, and strong extraction ability [15]. Therefore, the GCN is used to optimize the MLP to improve its accuracy and efficiency. The GCN is displayed in Fig. 2.

As shown in Fig. 2, the GCN algorithm contains an input

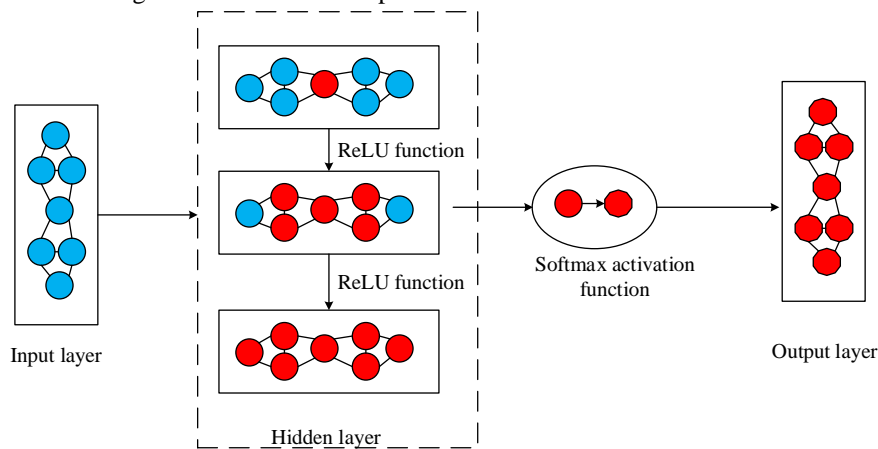


Fig. 2. GCN algorithm and basic structure diagram.

layer, multiple hidden convolutional layers, an activation layer, and an output layer [16]. In the input layer, data is clustered, and its feature information can be obtained from the neighboring nodes of that node during clustering. Then, the clustered data is passed into the HL, which is the core layer of the algorithm. In the HL, data graph convolution operations are performed. The features of each node in the clustered data are transformed through convolutional propagation to extract and retain their own feature information, removing irrelevant information. Finally, the data is normalized using the Softmax activation function. The propagation rule for each convolutional layer is shown in Eq. (9) [17].

$$M^{(l+1)} = \sigma(D^{-\frac{1}{2}} A D^{-\frac{1}{2}} B^{(l)} C^{(l)}) \quad (9)$$

In Eq. (9), A signifies the sum of the adjacency matrix and the closed-loop self-connection in the undirected graph. D signifies the degree matrix of A . $B^{(l)}$ is the activation unit matrix of l -th layer. $C^{(l)}$ signifies the parameter matrix of l -th layer. The nodes in the l -th layer complete the feature transformation operation, and the expression for this process is displayed in Eq. (10).

$$X^{(l+1)} = \sigma(NX^{(l)}K^{(l)} + m^{(l)}) \quad (10)$$

In Eq. (10), $X^{(l)}$ signifies the node feature of the l -th layer in the GCN. $K^{(l)}$ signifies the weight defined in layer l . σ is a nonlinear transformation. The adjacency matrix is normalized through a degree matrix, and the final expression is shown in Eq. (11).

$$x_i^{(l+1)} = \sigma\left(\sum_{j=N_i} D^{-\frac{1}{2}} A D^{-\frac{1}{2}} X^{(l)} C^{(l)} + b^{(l)}\right) \quad (11)$$

In Eq. (11), D signifies the degree matrix of A . All numbers on the diagonal of the adjacency matrix are changed to 1 through Eq. (11). The forward propagation is shown in Eq. (12).

$$Z = \text{soft max}(A \text{Re Lu}(AXC^{(0)})W^{(l)}) \quad (12)$$

Finally, the loss function of all points is calculated, as shown in Eq. (13).

$$L = - \sum_{l=y_L} \sum_{f=1} Y_l f \ln l_f \quad (13)$$

In Eq. (13), f represents the soft activation function. The extraction and preprocessing of data feature information are completed through the above process. Then, the information is transmitted into MLP for data analysis. The basic flowchart of MLP optimized by GCN is shown in Fig. 3.

From Fig. 3, the optimized MLP has a one-step data preprocessing process compared with the previous one. Firstly, the data is input into the GCN module for clustering analysis, and then the convolution operation is carried out to extract the data features. Afterwards, the data is normalized through the activation function to make the data the same form. Then, the data is taken as the input value of the MLP module. In the MLP module, the weight of the received data is allocated, and then it is calculated. Through continuous iteration, the data error is reduced to the allowable range. Then, the data is output. GCN is applied to preprocess the data, unify the data type and reduce the data volume, so as to enhance the operation speed and accuracy of MLP module.

B. Construction of Intrusion Detection Model Based on GCN-MLP

This study uses an intrusion detection model on the basis of GCN-MLP algorithm to detect information intrusion behavior in network security. It is hoped that this model can solve the low detection efficiency, false positives, and missed detection in current network intrusion detection models. The network security detection model is displayed in Fig. 4.

In Fig. 4, the network security detection model contains a detection layer, a transmission layer, a monitoring layer, and an application layer. In the detection layer of the model, network intrusion information is captured and transmitted to the algorithm detection model through sensors. The intrusion information is judged in the algorithm detection model. The transmission layer transmits the judged network intrusion information to the network through the server. In the monitoring layer, the intrusion information is monitored based on the judged intrusion information. Then, the user is searched through the database server and browser, and the intrusion information is transmitted to the user. This study uses an intrusion detection model based on GCN-MLP to investigate the module in the network security detection model. The basic structure diagram of the GCN-MLP intrusion detection model is shown in Fig. 5.

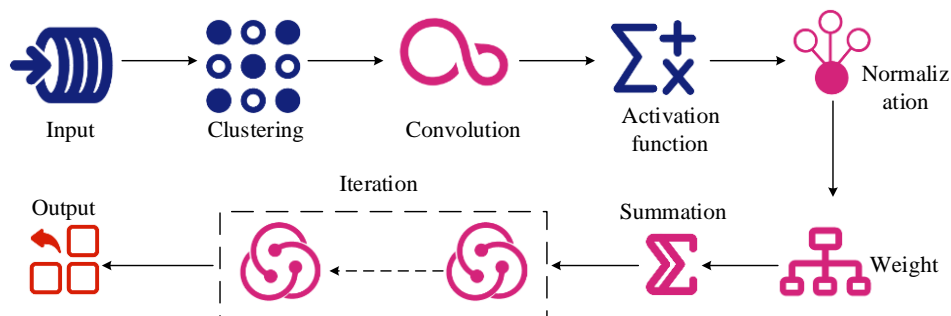


Fig. 3. Basic flowchart of GCN-MLP algorithm.

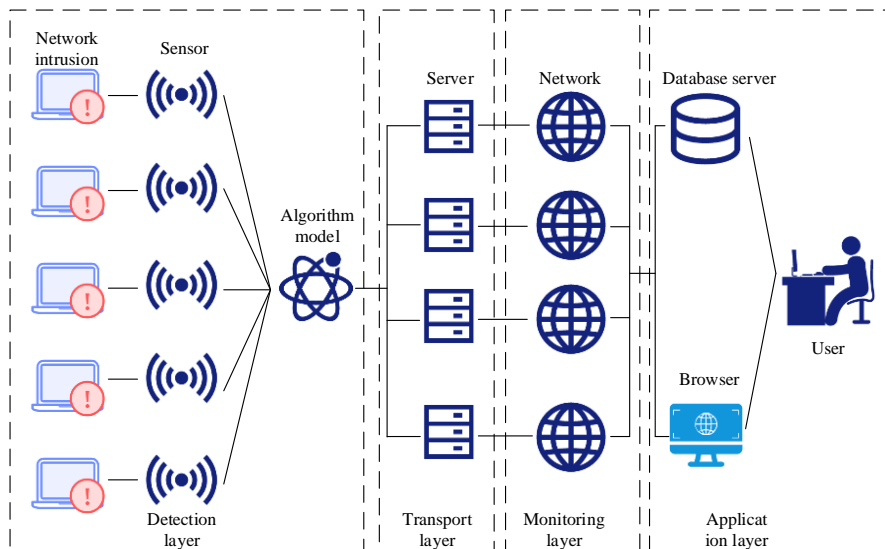


Fig. 4. Basic structure diagram of network security detection model.

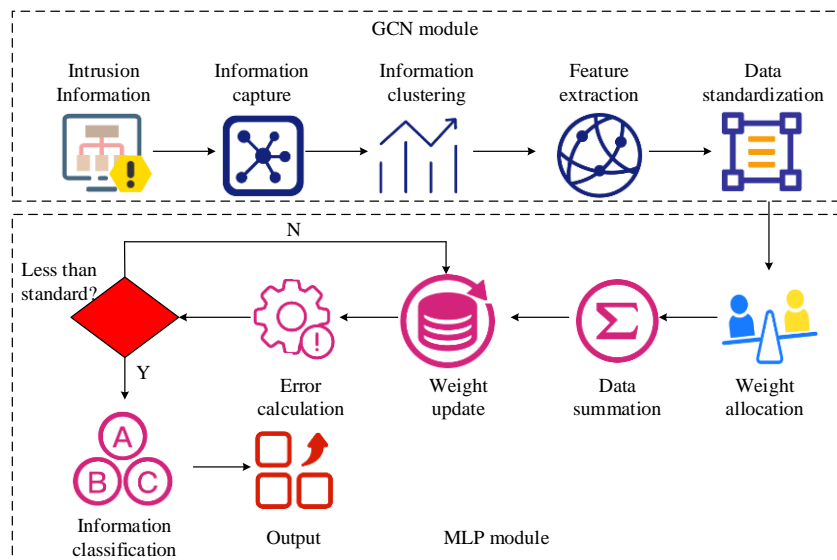


Fig. 5. GCN-MLP intrusion detection model.

As shown in Fig. 5, the model is divided into GCN module and MLP module. The GCN module captures the network intrusion information, clusters and integrates the captured information into data, extracts the features of the integrated data, and then standardizes the extracted feature information data to unify the data type. Then, the preprocessed data is sent as input information to the MLP. In this module, the incoming data is assigned weights, the weighted data is summed, the weights are updated, and the error value of the data is calculated. The error is compared with the minimum allowable error. If it is less than the allowable error, the intrusion information is classified based on the output data size to confirm the type of intrusion information. If the calculated error exceeds the allowable error, the weight is updated and the error is recalculated until the error is less than the allowable error value. The output calculation method of this model is shown in Eq. (14).

$$\hat{y} = h^T \begin{bmatrix} \Phi_{GCN} \\ \Phi_{MLP} \end{bmatrix} \quad (14)$$

In Eq. (14), h^T represents the weight matrix. The square loss is used as the loss function of the output model, as displayed in Eq. (15).

$$L = \sum_{(u,i) \in S} (\hat{y}_{ui} - y_{ui})^2 + \lambda \|\Theta\|^2 \quad (15)$$

In Eq. (15), (u,i) represents any number in the GCN dataset and MLP dataset, respectively. S represents the training dataset. \hat{y}_{ui} represents the predicted score. y_{ui} is the true score. Θ represents the weight parameter. λ represents the regularization parameter. To demonstrate the model effectiveness, the root mean square error is used as the evaluation index, as displayed in Eq. (16).

$$R = \sqrt{\frac{\sum_{(u,i) \in S} (\hat{y}_{ui} - y_{ui})^2}{N}} \quad (16)$$

In Eq. (16), N signifies the total data contained. The network intrusion detection model can timely detect various security risks in the network and effectively prevent network intrusion, thereby protecting network security.

III. RESULTS

A. Performance Analysis of GCN-MLP

To prove the superiority of GCN-MLP, the GCN-MLP is compared with Fusion Algorithm combined Convolutional Neural Network algorithm with Convolutional Attention Module (CNN-CBAM), Fusion Algorithm based on Time Convolutional Network and Bidirectional LSTM (TCN-BiLSTM), as well as Fusion Algorithm combined Principal Component Analysis algorithm with K-means clustering (PCA-K-means). Table I displays the configuration.

TABLE I. EXPERIMENTAL CONFIGURATION TABLE

Environment	Index	Type
Hardware environment	OS system	Winds 10
	Hardpan	500G
	CPU	I7 3.4Hz
	Internal memory	4GB
Software environment	Pyrhon	Pyrhon 3.x
	Matlab	Matlab7.0

According to Table I, the environmental configuration conditions during the experiment are obtained. During the experiment, the node features, HL features, and output layer features of the GCN are 50, and the number of layers in GCN is 5. The penalty coefficient is 0.001, and the learning rate is 0.005 in the MLP. The learning rate is 0.1, the capacity is 100, the weight attenuation is 0.005, and the training frequency is 50 in the CNN. The convolution kernel in CBAM is 9, the

convolution kernel size is 3*3, the weight threshold is 0.5, and the maximum pooling layer is set to 3*3. The number of neurons in BiLSTM is 100, and the batch size is 10. The n_components in the PCA algorithm is set to none, the copy value is True, and the white value is False. The K-value in the K-means is 50, and the maximum iteration is 500. Comparative experiments are carried out on the KDD CUP 99 dataset based on the parameter settings mentioned above. The superiority of the proposed algorithm was verified by comparing the accuracy, loss function value, F1 value, detection time, and ROC curve of four algorithms. The comparison between the predicted and the actual results, as well as the accuracy results, are shown in Fig. 6.

According to Fig. 6 (a), the GCN-MLP algorithm had the closest predicted result and the smallest difference. The difference of CNN-CBAM algorithm and TCN-BiLSTM algorithm was greater than that of GCN-MLP algorithm. The PCA-K-means algorithm had the greatest difference. In Fig. 6 (b), the accuracy of the four algorithms increased when the iteration was between 0 and 20. However, when the iteration exceeded 20, the accuracy stabilized. The accuracy of the GCN-MLP algorithm stabilized at 0.98 after more than 20 iterations.

The accuracy of the CNN-CBAM algorithm, TCN-BiLSTM algorithm, and PCA-K-means algorithm were 0.81, 0.69, and 0.61, respectively. Afterwards, comparative experiments are conducted on the F1 values and loss function values, as displayed in Fig. 7.

From Fig. 7, the F1 values and loss function values varied with the increase of iterations. From the Figure, the F1 value of the GCN-MLP algorithm reached its maximum value at 5 iterations, with a maximum F1 value of 0.97. However, the CNN-CBAM algorithm, TCN-BiLSTM algorithm, and PAC-K-means algorithm only reached their maximum F1 value at 10 iterations. The maximum F1 of these three algorithms was 0.92, 0.87, and 0.78. The loss function values decreased with the increase of iterations. In Fig. 7, the loss function value of the GCN-MLP algorithm stabilized at 0.03, which was much lower than the CNN-CBAM at 0.09, TCN-BiLSTM at 0.12, and PCA-K-means at 0.15. In Fig. 7, when the number of iterations was greater than 20, the loss function fluctuation range of the TCN-BiLSTM algorithm and PAC-K-means algorithm was larger, with the PCA-K-means algorithm having the largest fluctuation range and the smallest stability. Further analysis is conducted on the detection time and ROC curves, as displayed in Fig. 8.

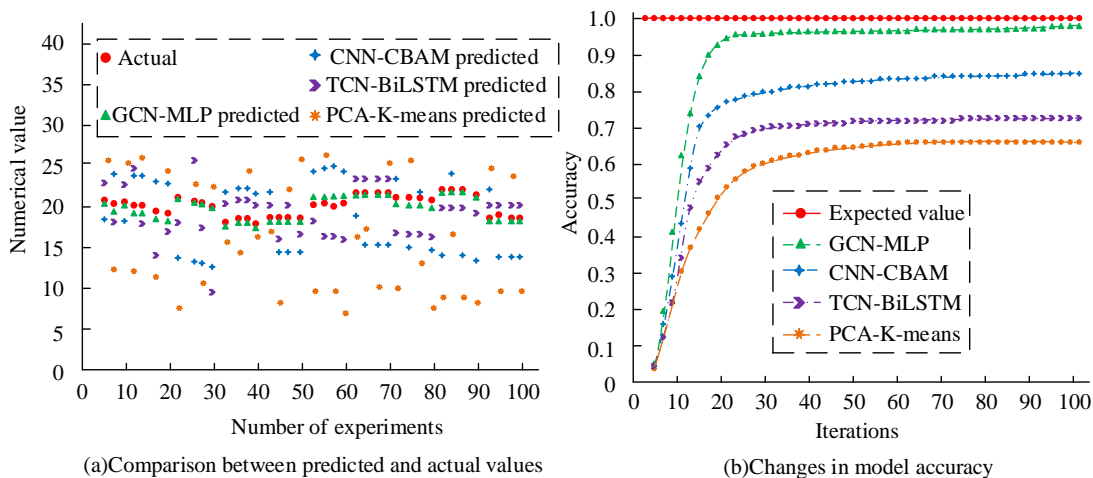


Fig. 6. Algorithm prediction results and accuracy.

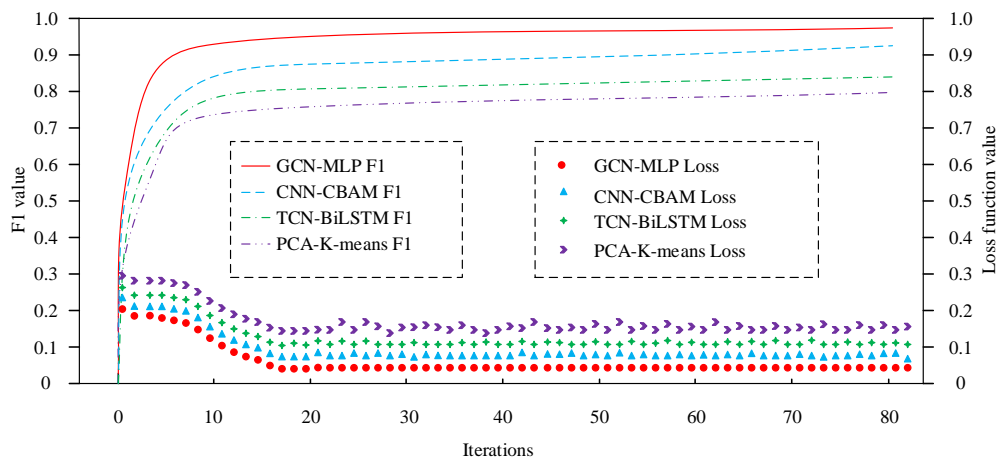


Fig. 7. Comparison of F1 value and loss function value.

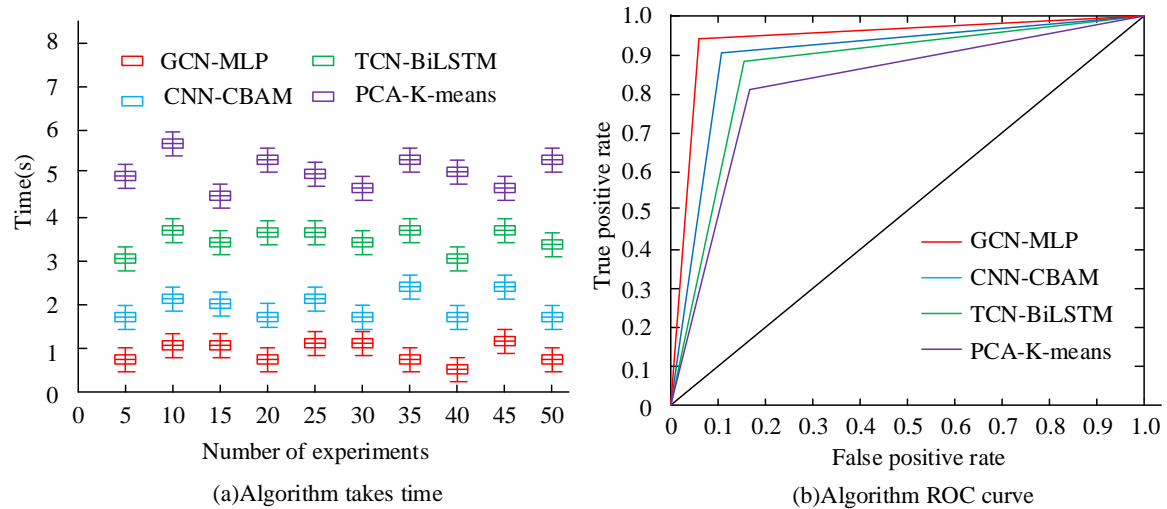


Fig. 8. Detection time and ROC curve of the algorithm.

According to Fig. 8 (a), the average detection time of the GCN-MLP was the shortest, at 1.1s. The average detection time of the CNN-CBAM was 1.9s. The average time for the TCN-BiLSTM was 3.2s. The PCA-K-means algorithm had the longest average time, which was 5.3s. The accuracy, false detection rate, and missed detection rate can be observed from the curve in Fig. 8 (b). The ROC close to the upper left corner demonstrates that the prediction accuracy is higher. From Fig. 8 (b), the ROC of GCN-MLP algorithm was closest to the upper left corner, followed by CNN-CBAM algorithm, and PCA-K-means algorithm was farthest. Therefore, among the four algorithms, GCN-MLP algorithm had the highest prediction accuracy, and PCA-K-means algorithm had the lowest prediction accuracy. GCN-MLP has the highest accuracy, fastest detection speed, and strongest stability. The overall

performance is significantly better than other algorithms.

B. Application Effect of GCN-MLP Model in Network Security Detection

After verifying the superiority of the GCN-MLP algorithm, experimental analysis is conducted on the detection model based on the algorithm. The proposed model (Model 1) is compared with intrusion detection model integrating improved auto-encoder and residual network (Model 2), intrusion detection model integrating contrastive learning and feature selection (Model 3), and residual network detection model combined with fusion attention mechanism (Model 4). The accuracy, precision, recall, F1, underreporting rate, and detection rate of the four models are analyzed. The comparison results are shown in Fig. 9.

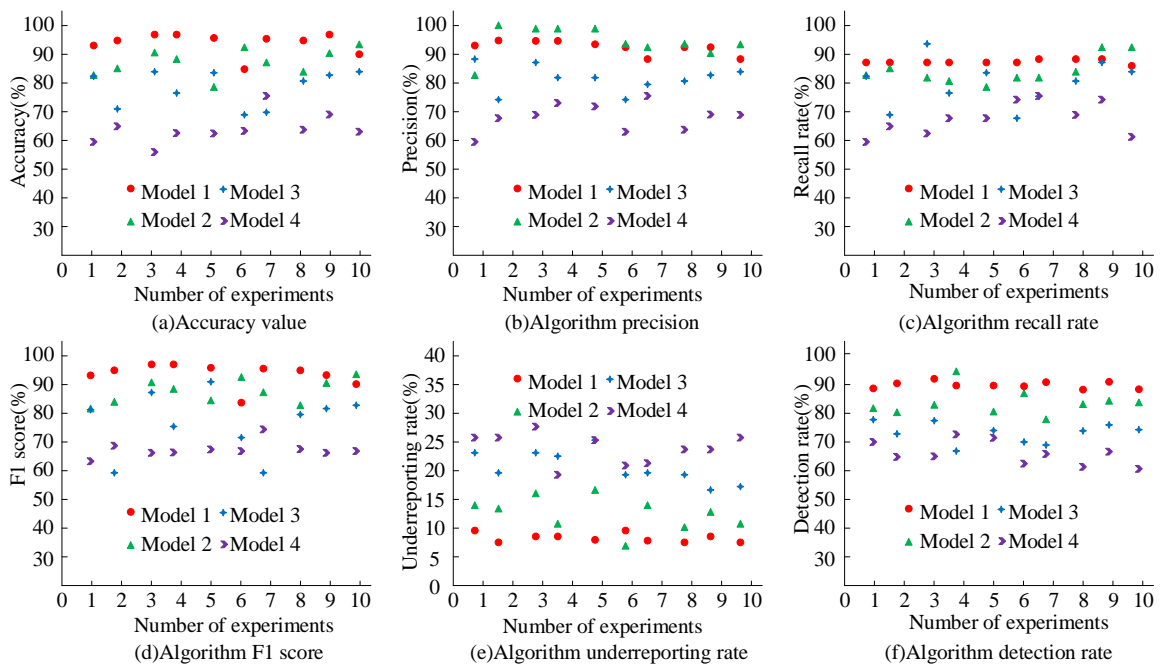


Fig. 9. Comparison of algorithm indicators.

Fig. 9 displays the comparison results of various indicators. From Fig. 9 (a), the detection accuracy of Model 1 was the highest among the four models at 98%, while the detection accuracy of Model 2, Model 3, and Model 4 were 89%, 80%, and 68%, respectively. From Fig. 9 (b), Model 2 had the highest detection precision of 97%, while Model 4 had the lowest detection precision of 78%. In Fig. 9 (c), the recall rate gradually decreased from Model 1 to Model 4. From Fig. 9 (d), after multiple experiments, the F1 value of Model 1, Model 2, Model 3, and Model 4 was 97%, 90%, 87%, and 68%, respectively. From Fig. 9 (e) and 9 (f), Model 1 had the lowest underreporting rate, but the highest data detection rate, with a underreporting rate of 6% and a detection rate of 92%. Through experiments, it is known that Model 1 has slightly lower detection accuracy than Model 2, and all other indicators are better than comparison models. The overall performance is the best among the four models. In summary, the detection model based on GCN-MLP algorithm has the best overall performance. The GCN-MLP detection model is applied to actual network security detection. The accuracy of

intercepting intrusion information and the interception time of various illegal intrusions in network security detection are compared. 20 experimental results are taken, and the average accuracy and interception time of every 5 experimental results are calculated and represented by a coordinate graph. The accuracy and detection time results are shown in Fig. 10.

According to Fig. 10 (a), the average accuracy of Model 1 in network security detection was 0.98, the accuracy of Model 2 in network security detection was 0.89, and the accuracy of Model 3 was 0.71. The accuracy of Model 4 was the lowest among the four models, which was 0.57. Fig. 10 (b) shows the time it takes for four models to detect and judge intrusion information. From Fig. 10 (b), the average time for Model 1 to detect intrusion information was 0.1s, which was much lower than the 0.9s of Model 2, 2.7s of Model 3, and 4.2s of Model 4. Further experiments are conducted on the accuracy of four models in determining various types of network intrusion information, as displayed in Fig. 11.

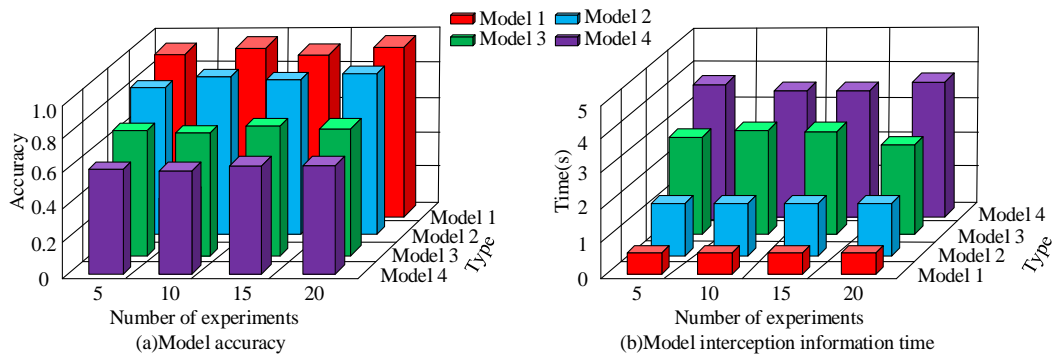


Fig. 10. Accuracy and interception time of network security inspection model.

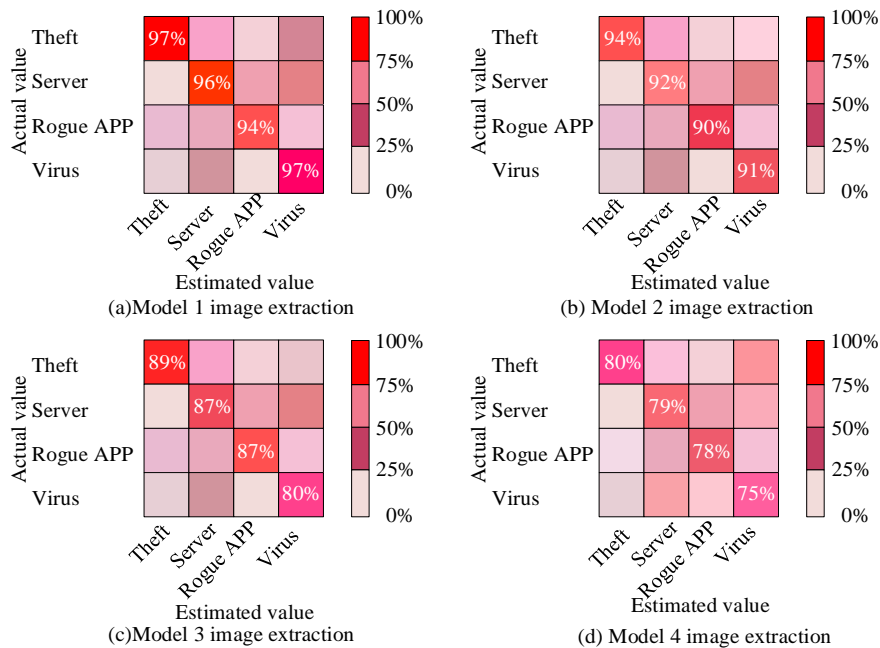


Fig. 11. Model ability to judge intrusion information.

Fig. 11 shows the accuracy results of four models in judging intrusion information encountered in network security detection. The elements on the main diagonal signify the proportion of correctly predicted intrusion information types. The elements in the lower left triangle signify the proportion of missed intrusion information types. The elements in the upper right triangle represent the proportion of false detected intrusion information types. According to Fig. 11 (a), Model 1 had a prediction accuracy of 97% for the theft intrusion information in the intrusion information, a detection accuracy of 96% for server intrusion information, a detection accuracy of 94% for malware information, and a detection accuracy of 97% for virus intrusion. The detection accuracy of Model 2 for the four types of intrusion information was 94%, 92%, 90%, and 91%, respectively. The detection accuracy of Model 3 and Model 4 for the four types of intrusion information was much lower than that of Model 1 and Model 2. From the above experimental results, the GCN-MLP has the best performance among the four detection models. This model is used in network security intrusion systems, which has the highest accuracy in detecting intrusion information.

IV. DISCUSSION

The study verified the significant advantages of the network security detection model based on GCN-MLP in accuracy, speed, and stability through experiments. Compared with the other three algorithms, GCN-MLP not only achieved a stable high accuracy of 0.98 after 20 iterations, but also had an F1 value of 0.97. The loss function value remained stable at a lower level of 0.03, which fully demonstrated the efficiency of the algorithm. This is fitted with the conclusion drawn by Yao et al. on the GCN-MLP algorithm [18]. In addition, from the experimental results, the GCN-MLP algorithm performed equally well in detection time, averaging only 1.1s, which was much faster than the other three algorithms. In the field of network security, fast detection time means that potential threats can be responded to more quickly, effectively reducing risks, which is linked to the results drawn by He et al [19]. Further research found that when comparing the GCN-MLP detection model with three other advanced detection models, the GCN-MLP model maintained a leading position in multiple key indicators such as accuracy, recall, F1 value, and detection rate. Especially, the underreporting rate was only 6%, far lower than other models, which was extremely important in the field of network security because underreporting may lead to serious security risks.

The GCN-MLP model had a detection accuracy of over 94% for theft intrusion, server intrusion, malware information, and virus intrusion, demonstrating extremely high reliability and comprehensiveness. Compared with the algorithms and models designed by Yu et al. and Yang et al., the GCN-MLP algorithm also exhibited excellent performance. Because the deep learning model designed by Yu et al. and Yang et al. had an accuracy of only 80%-90% in network security detection, the GCN-MLP model further enhanced this standard [20-21]. Meanwhile, the stability of the GCN-MLP was also commendable. During the experiment, the fluctuation of the loss function value was relatively small. It means that in practical applications, the model can provide more reliable and consistent results. This result is significantly better than the

stability of the network security protection model designed by Wang et al [22]. In summary, the network security detection model based on GCN-MLP shows significant advantages in multiple aspects, which not only proves the effectiveness of this method, but also provides strong support for its application in practical network security protection.

V. CONCLUSION

In response to the low accuracy, serious false positives, and missed detection rate in current information intrusion detection models, this study proposed the CN-MLP algorithm integrating GCN algorithm and MLP algorithm. Then, an information intrusion detection model was constructed based on the fused GCN-MLP algorithm, CNN-CBAM algorithm, TCN-BiLSTM algorithm, and PCA-K-means algorithm. The overall performance of the GCN-MLP algorithm outperformed other comparison algorithms. Subsequently, the method was compared with intrusion detection model integrating improved auto-encoder and residual network, intrusion detection model integrating contrastive learning and feature selection, and residual network detection model combined with fusion attention mechanism. The designed intrusion detection method had a much higher detection accuracy for network intrusion information than the other comparison models. In summary, the detection model on the basis of GCN-MLP has the best overall performance in network security intrusion information detection, which can effectively improve network security. However, the types of intrusion information discussed in this study are limited, and there is still uncertainty. In the future, data augmentation techniques can be used to oversample minority class samples, synthesize new minority class samples, expand the sample size, and increase detection information. Meanwhile, generative adversarial networks can be used to generate similar intrusion detection information, increase sample size, and improve the overall detection performance of the model.

REFERENCES

- [1] Khan M, Ghafoor L. Adversarial Machine Learning in the Context of Network Security: Challenges and Solutions. *Journal of Computational Intelligence and Robotics*, 2024, 4(1): 51-63.
- [2] Bandewad G, Datta K P, Gawali B W, Pawar, S. N. Review on Discrimination of Hazardous Gases by Smart Sensing Technology. *Artificial Intelligence and Applications*. 2023, 1(2): 86-97.
- [3] Fu Y, Du Y, Cao Z, Li Q, Xiang W. A deep learning model for network intrusion detection with imbalanced data. *Electronics*, 2022, 11(6): 898-901.
- [4] Hnamte V, Nhung-Nguyen H, Hussain J, Hwa-Kim Y. A novel two-stage deep learning model for network intrusion detection: LSTM-AE. *Ieee Access*, 2023, 11(5): 37131-37148.
- [5] Dai J, Zhu W, Luo X. A targeted universal attack on graph convolutional network by using fake nodes. *Neural Processing Letters*, 2022, 54(4): 3321-3337.
- [6] Diao C, Zhang D, Liang W, Li K C, Hong Y, Gaudiot J L. A novel spatial-temporal multi-scale alignment graph neural network security model for vehicles prediction. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 24(1): 904-914.
- [7] Deng X, Zhu J, Pei X, Zhang L, Ling Z, Xue K. Flow topology-based graph convolutional network for intrusion detection in label-limited IoT networks. *IEEE Transactions on Network and Service Management*, 2022, 20(1): 684-696.

- [8] Al-Ibraheemi F A, Hazzaa F, Jabbar M S, Tawfeq J F, Sekhar R, Shah P, Parihar S. Intrusion Detection in Software-Defined Networks: Leveraging Deep Reinforcement Learning with Graph Convolutional Networks for Resilient Infrastructure. Full Length Article, 2024, 15(1): 78-87.
- [9] Setitra M A, Fan M, Agbley B L Y, Bensalem Z E A. Optimized MLP-CNN Model to Enhance Detecting DDoS Attacks in SDN Environment. Network, 2023, 3(4): 538-562.
- [10] Shewale Y, Kumar S, Banait S. Machine Learning Based Intrusion Detection in IoT Network Using MLP and LSTM. International Journal of Intelligent Systems and Applications in Engineering, 2023, 11(7): 210-223.
- [11] Najar A A, Manohar Naik S. DDoS attack detection using MLP and Random Forest Algorithms. International Journal of Information Technology, 2022, 14(5): 2317-2327.
- [12] Diao C, Zhang D, Liang W, et al. A novel spatial-temporal multi-scale alignment graph neural network security model for vehicles prediction. IEEE Transactions on Intelligent Transportation Systems, 2022, 24(1): 904-914.
- [13] Alsirhani A, Alshahrani M M, Abukwaik A, Taloba A I, Abd El-Aziz R M, Salem M. A novel approach to predicting the stability of the smart grid utilizing MLP-ELM technique. Alexandria Engineering Journal, 2023, 74(5): 495-508.
- [14] Wang W, Wen F, Zheng H, Ying R, Liu P. Conv-MLP: A convolution and MLP mixed model for multimodal face anti-spoofing. IEEE Transactions on Information Forensics and Security, 2022, 17(4): 2284-2297.
- [15] Pankova M, Kwilinski A, Dalevska N, Khobta V. Modelling the Level of the Enterprise Resource Security Using Artificial Neural Networks. Virtual Economics, 2023, 6(1): 71-91.
- [16] Zheng H, Li X, Li Y, Yan Z, Li T. GCN-GAN: integrating graph convolutional network and generative adversarial network for traffic flow prediction. IEEE Access, 2022, 10(5): 94051-94062.
- [17] Huang D, Liu H, Bi T, Yang Q. GCN-LSTM spatiotemporal-network-based method for post-disturbance frequency prediction of power systems. Global Energy Interconnection, 2022, 5(1): 96-107.
- [18] Yao Z, Yu J, Zhang J, He W. Graph and dynamics interpretation in robotic reinforcement learning task. Information Sciences, 2022, 611(4): 317-334.
- [19] He J, Abueidda D, Koric S, Jasiuk I. On the use of graph neural networks and shape-function-based gradient computation in the deep energy method. International Journal for Numerical Methods in Engineering, 2023, 124(4): 864-879.
- [20] Yu J, Ye X, Li H. A high precision intrusion detection system for network security communication based on multi-scale convolutional neural network. Future Generation Computer Systems, 2022, 129(6): 399-406.
- [21] Yang H, Zhang Z, Xie L, Zhang L. Network security situation assessment with network attack behavior classification. International Journal of Intelligent Systems, 2022, 37(10): 6909-6927.
- [22] Wang Z, Xie X, Chen L, Song S, Wang Z. Intrusion detection and network information security based on deep learning algorithm in urban rail transit management system. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(2): 2135-2143.

An Ensemble Semantic Text Representation with Ontology and Query Expansion for Enhanced Indonesian Quranic Information Retrieval

Liza Trisnawati¹, Noor Azah Binti Samsudin², Shamsul Kamal Bin Ahmad Khalid³, Ezak Fadzrin Bin Ahmad Shaubari⁴, Sukri⁵, Zul Indra⁶

Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia^{1, 2, 3, 4}
Department of Informatics Engineering, Faculty of Engineering, Universitas Abdurrah, Pekanbaru, Indonesia^{1, 5}
Department of Computer Science, Universitas Riau, Pekanbaru, Indonesia⁶

Abstract—This study explores the effectiveness of an ensemble method for Quranic text retrieval, aimed at improving the relevance and accuracy of verses retrieved for specific themes. The ensemble approach integrates three semantic models—Word2Vec, FastText, and GloVe—through a voting mechanism that considers verse frequency and semantic alignment with the query topics. Testing was conducted on themes such as prayer, zakat, fasting, umrah, and eschatology, reflecting fundamental aspects of Quranic teachings. Results demonstrate that the ensemble method significantly outperforms non-ensemble approaches, achieving an average relevance rate of 88%, compared to individual models (Word2Vec: 75%, FastText: 80%, GloVe: 82%). The ensemble method effectively combines the unique strengths of each model. Word2Vec captures general semantic relationships, FastText handles morphological nuances, and GloVe identifies global contextual patterns. By combining these capabilities, the ensemble approach improves both the quantity and quality of retrieved verses, making it a robust tool for semantic analysis in Quranic studies. This research contributes to the field of computational Islamic studies by demonstrating the practical advantages of ensemble methods for religious text retrieval. It lays the foundation for further advancements, including the integration of deep learning techniques, dynamic query handling, and cross-linguistic analysis. The ensemble method offers a promising framework for supporting more accurate and contextually relevant Quranic studies, promoting a deeper understanding of Islamic teachings through data-driven methodologies.

Keywords—Ensemble method; query expansion; ontology; Al-Quran; search engine

I. INTRODUCTION

The Quran, as the holy book of Islam, holds profound spiritual, moral, and ethical guidance for over a billion Muslims worldwide. It serves not only as a religious text but also as a comprehensive source of knowledge, law, and inspiration. The Quran's linguistic and contextual depth reflects its universal nature, which transcends time and culture. However, this same depth presents significant challenges in making its meanings accessible, particularly for non-Arabic-speaking audiences such as Indonesians, who rely on translations and interpretations to understand its contents. Indonesia, being home to the largest Muslim population globally, has a pressing need for efficient tools to access

Quranic knowledge in the Indonesian language. However, traditional search systems often fall short in meeting user expectations due to their inability to grasp the semantic richness of Quranic text [1]. Literal keyword matching methods, for example, frequently fail to account for synonyms, related terms, and contextual nuances inherent in religious texts. This necessitates the development of advanced information retrieval (IR) systems tailored to handle the complexities of Quranic text in translation.

A major obstacle in existing Quranic IR systems lies in their limited ability to interpret semantic relationships between terms. While some systems incorporate basic query refinement techniques, they rarely achieve the level of sophistication needed for meaningful interpretation of Quranic content. Query Expansion (QE), which involves broadening search queries by including semantically related terms, has shown great promise in addressing these challenges. By enhancing the original query, QE can improve the relevance and accuracy of search results, especially in highly structured texts like the Quran [2], [3], [4]. Ontology-based Query Expansion offers a powerful solution by leveraging structured knowledge frameworks that capture the domain-specific relationships and meanings within Quranic text. Ontologies can represent complex semantic relationships such as synonyms, hypernyms, and contextual associations, enabling more precise query interpretation [5], [6]. This approach is particularly valuable for Quranic IR, where understanding the contextual use of terms is critical for delivering meaningful search results.

In addition to ontology-based QE, advances in semantic text representation provide new opportunities for improving IR systems. Semantic representation methods, particularly those using neural networks, capture not only the lexical features of text but also its contextual meanings. Ensemble techniques, which combine multiple models to optimize performance, are increasingly recognized as a robust approach for text representation. By aggregating the strengths of various models, ensemble methods can better handle the linguistic intricacies of Quranic text and its Indonesian translation.

The integration of ontology-based QE with ensemble semantic text representation models has the potential to revolutionize Quranic IR systems. This combination ensures

that search results are not only relevant but also contextually accurate, aligning with the inherent richness of Quranic discourse. By leveraging these advanced techniques, the proposed system aims to bridge the gap between user queries and the deep, layered meanings of the Quranic text. The research focuses on developing a tailored Quranic IR system specifically for the Indonesian language. Unlike generic search engines, this system will address the unique challenges posed by Quranic text, such as polysemy, synonymy, and contextual interpretation. It will also incorporate an extensive ontology of Quranic terms and their relationships, further enriching the system's ability to understand user intent.

Moreover, the use of ensemble methods ensures the robustness of the proposed system. By combining multiple semantic text representation models, the system can effectively capture both local and global contextual information in the text. This not only improves the accuracy of search results but also enhances the user experience by providing more nuanced and comprehensive responses to queries. This study represents a significant contribution to the field of Quranic studies and information retrieval. By addressing the limitations of existing systems and introducing a novel combination of ontology-based QE and ensemble techniques, it sets a new standard for Quranic IR. The findings of this research are expected to benefit not only Muslim communities but also researchers and practitioners working on religious and domain-specific IR systems.

In conclusion, the development of an ontology-enriched Query Expansion method integrated with ensemble semantic text representation offers a promising solution for improving Quranic information access in the Indonesian language. This research not only aims to enhance the retrieval performance of Quranic IR systems but also serves as a benchmark for similar efforts in other languages and religious texts, ensuring broader applicability and impact.

The paper is organized as follows: Section II provides a literature review of relevant works on query expansion, ontologies, and word embeddings. Section III outlines the methodology, detailing the construction of the Quranic ontology, the implementation of Word2Vec, and the integration of these components into a search engine. Section IV presents the results of the system's performance evaluation, focusing on precision, recall, and relevance of the search results. Finally, Section V discusses the conclusions, limitations, and future directions for further research.

II. LITERATURE REVIEW

A. Information Retrieval Based on Query Expansion and Ensemble Text Representation

Information retrieval (IR) is a fundamental process in managing and extracting relevant information from large datasets [7], [8]. Traditional IR systems rely on keyword-based searches, where users input queries, and the system returns documents containing those keywords [9], [10], [11]. However, such systems often face limitations due to the ambiguity of user queries and the mismatch between user language and the indexed data [12], [13]. This limitation has led to the development of query expansion techniques, which

aim to improve search accuracy by reformulating user queries to include related terms [14].

Several techniques have been developed for query expansion, each offering different approaches to improving search results [15], [16]. One method is manual query expansion, where domain experts carefully select synonyms or related terms to enhance the query [17]. Another approach is automatic query expansion (AQE), in which the system automatically identifies related terms using techniques such as relevance feedback, thesaurus-based expansion, or statistical co-occurrence analysis. More recently, word embeddings-based expansion, such as Word2Vec, has emerged as a powerful method. This approach leverages vector representations of words to suggest semantically related terms [18] by analyzing their proximity in a high-dimensional vector space, providing a more dynamic and context-aware means of expanding queries [19], [20], [21].

Recent advancements in text representation based on word embedding models, particularly Word2Vec, FastText, and GloVe, have demonstrated significant improvements in capturing semantic relationships between terms, making them popular tools for automatic query expansion. Several studies [14], [19], [22] demonstrated that Word2Vec could effectively suggest semantically similar terms effectively. FastText, on the other hand, extends this capability by incorporating subword information [23], making it particularly effective in handling morphologically rich languages and rare or unseen words [24]. GloVe, by leveraging global co-occurrence statistics [25], [26], provides robust embeddings that capture the relationships between words across broader contexts [27], [28]. Together, these methods have been successfully applied in various applications, from general-purpose search engines to domain-specific information retrieval (IR) systems, demonstrating their ability to enrich user queries and improve the relevance of search results.

To further enhance the query expansion process, this research employs an ensemble method that combines the outputs of Word2Vec, FastText, and GloVe. Ensemble methods, which integrate multiple models, leverage the strengths of each model while mitigating their individual weaknesses [29]. For instance, Word2Vec excels in local context understanding, FastText captures morphological subtleties, and GloVe provides a comprehensive global semantic understanding. By aggregating these outputs using techniques such as weighted voting, the ensemble method achieves a balanced representation that is both lexically precise and contextually rich. The advantages of ensemble methods include improved robustness, reduced overfitting, and higher accuracy in handling complex or diverse queries. In this research, the ensemble approach ensures that query expansion is not only semantically accurate but also contextually aligned with the intricate thematic and linguistic structure of Quranic texts, thereby significantly enhancing the performance of the proposed IR system.

B. Ontology in Information Retrieval

Ontologies play a crucial role in enhancing information retrieval (IR) [30] systems by bridging the semantic gap between the terms users input in their queries and those

indexed within the system. By offering a structured and hierarchical representation of domain knowledge, ontologies enable IR systems to enrich user queries with related terms, such as synonyms, hyponyms, and hypernyms, through the query expansion process. This structured approach supports more advanced semantic search capabilities, allowing the system to not only match keywords but also to understand the underlying meaning and context of user queries, ultimately improving the relevance and accuracy of search results. Incorporating ontologies into search systems has been particularly beneficial in specialized domains such as medical databases, educational resources, and legal information systems. Ontology-based systems can also be used for concept-based retrieval, where the system retrieves documents based on the underlying concepts represented in the query rather than exact keyword matches.

The use of ontology and query expansion in religious texts, particularly the Qur'an is gaining attention due to the need for more intelligent and context-aware search systems. The Qur'an is a rich and complex text with intricate themes, concepts, and linguistic variations, making it challenging for traditional keyword-based search systems to capture the full meaning and relevance of user queries.

Several studies have explored the use of ontology in Qur'anic search systems. For example, Mohamed, Ensaf Hussein, and Eyad Mohamed Shokry [31] developed a Qur'anic ontology based on concept-based searching tool (QSST) to facilitate semantic-based search. In this research, ontology was created through manual annotation of verses of the Al-Quran based on the Al-Tajweed Mushaf. In another study [32], ontology development was carried out for the Quran by adopting the use of Protégé-OWL and SPARQL queries. In addition, there are still several studies that try to apply searches based on semantic relationships that exist in each verse of the Quran [31], [33], [34]. Thus, it can be concluded that the integration of Word2Vec with ontology has been proven to significantly improve the search process. By utilizing the semantic knowledge embedded in ontology and word vectors, this system can produce more accurate user query expansions, thereby increasing precision and recall in search.

C. Research Gap and Contribution

Despite significant progress in integrating ontology and query extension into information retrieval systems, several challenges remain. As explained previously, it was found that only a few studies have tried the ontology and query expansion approach to facilitate information retrieval from the Quran. Most studies with the topic of information retrieval from the Quran tend to only apply the labeling concept [35], index-based ranking without trying to understand semantic relationships as a representation of contextual verses [36], [37], [38], [39]. Moreover, regarding the application of the Indonesian translation of the Quran as a case study, it is still under discussed. Most existing studies prefer a conventional keyword-based approach [40] or the use of glossaries as keyword enrichment [41]. Only 1 study was found that tried to explore semantic relationships as applied by Purnama et al. [42]. Another notable research gap is the limited application of ensemble methods in query expansion for Quranic IR. While

ensemble approaches have shown success in improving text representation and classification tasks in general IR, their use in combining multiple semantic representation models for query expansion remains underexplored. Ensemble methods, which aggregate the strengths of various models, could potentially enhance the robustness and accuracy of expanded queries, particularly in complex and domain-specific texts like the Quran.

Additionally, the performance of existing Quranic IR studies remains relatively low, as they often fail to optimize retrieval accuracy and relevance due to the lack of advanced semantic techniques. This highlights the need for innovative methodologies that integrate ontology-based query expansion with ensemble deep learning models to address these limitations effectively. In conclusion, a summary of the research gaps and the contributions offered by this study is illustrated in Fig. 1. These gaps emphasize the need for more sophisticated approaches that combine ontologies, semantic relationships, and ensemble deep learning techniques to improve the performance of Quranic IR systems, particularly in the context of the Indonesian translation.

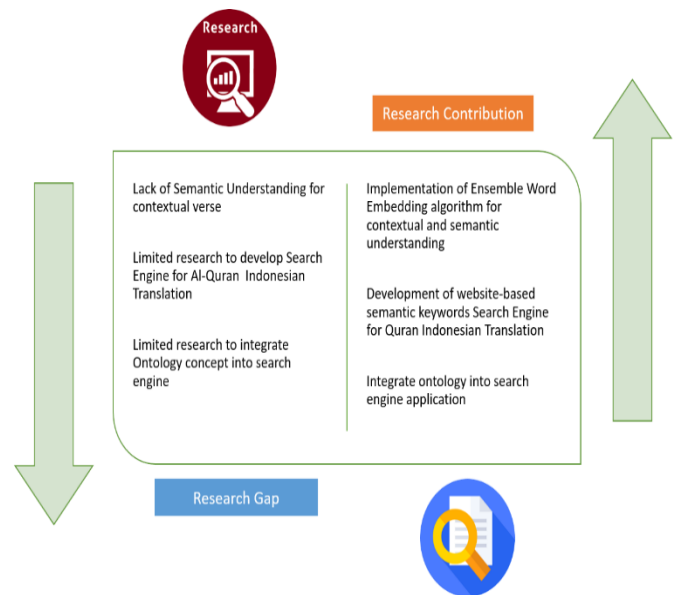


Fig. 1. Research gap and proposed contribution.

Based on Fig. 1, this study aims to develop an ensemble semantic search engine application enriched with Ontology which is expected to solve problems in existing research. Specifically, this search engine application is intended for the Indonesian language Qur'an because there is still little discussion on this topic.

III. METHODOLOGY

A. Dataset Material

The dataset utilized in this study is meticulously compiled from two primary sources: the official Indonesian translation of the Quran published by the Ministry of Religious Affairs (Kemenag) and the Indonesian Wikipedia corpus. The integration of these resources ensures a robust semantic foundation for developing an advanced information retrieval

system tailored to Quranic content in the Indonesian language. The official Kemenag translation serves as an authoritative and widely recognized resource, ensuring theological and linguistic accuracy. It comprises all 114 surahs and 6,236 verses, each accompanied by its corresponding Arabic text to maintain contextual alignment. Additionally, the dataset includes thematic metadata categorizing Quranic verses into key topics such as faith (iman), worship (ibadah), morals (akhlak), and law (syariah), which is crucial for ontology construction and query expansion.

To address the inherent limitations of Quranic text alone in covering broader semantic contexts, the dataset is enriched with the Indonesian Wikipedia corpus. The Wikipedia corpus provides a vast repository of general knowledge that complements the Quranic dataset by introducing a wider range of linguistic and contextual diversity. While the Quranic text is specific and focused, the Wikipedia corpus offers the flexibility to understand related terms and concepts that may not explicitly appear in the Quran. For instance, abstract ideas such as "justice" (keadilan) and "mercy" (rahmat), which are central to Islamic teachings, can be explored in their broader cultural, philosophical, or societal dimensions through Wikipedia entries. This enrichment allows the system to better handle complex or indirect queries by providing semantic connections between Quranic themes and contemporary knowledge.

Prior to integrate the datasets into the system, several preprocessing steps are carried out to ensure data quality and consistency. For the Quranic text, transliterations are standardized, and Arabic diacritical marks (tashkeel) are removed to simplify tokenization. The Indonesian translation is normalized by converting text to lowercase, eliminating punctuation, and resolving linguistic variations to create a uniform dataset. Similarly, the Wikipedia corpus undergoes a rigorous preprocessing pipeline that involves noise removal, where irrelevant or overly technical content is filtered out, and tokenization, where text is broken into meaningful linguistic units. Additionally, stop words, such as common function words in Indonesian, are removed to enhance the focus on semantically significant terms.

The preprocessed datasets are then aligned and structured for downstream tasks, such as ontology development and semantic text representation training. This ensures that the Quranic and Wikipedia datasets are not only compatible but also semantically enriched to facilitate accurate, relevant, and context-aware information retrieval. By integrating a carefully curated and preprocessed dataset, the system can effectively bridge the gap between Quranic-specific queries and broader thematic searches, enhancing the overall user experience.

B. Proposed Method

This study aims to develop a thematic index-based Al-Qur'an ontology system and implement query expansion techniques to support more relevant and accurate information retrieval. The system is designed to enhance semantic access to Al-Qur'an verses, enabling topic-based searches such as Morals, Faith, Worship, and Law. To refine the query expansion process, this study incorporates ensemble text representation methods by combining Word2Vec, GloVe, and

FastText. These methods collectively capture word-level, global co-occurrence, and subword-level semantics, ensuring a robust representation of Quranic text. Weighted voting is employed to integrate the strengths of each model, allowing the system to provide search results that are both contextually rich and semantically precise.

The methodology involves several critical stages, starting with data collection from the official Indonesian translation of the Quran and the Wikipedia corpus for contextual enrichment. Ontology development follows to structure thematic relationships within the Quran. Ensemble text representation is then applied to support query expansion, enriching user queries with semantically related terms. Finally, the system is integrated and evaluated using metrics such as precision, recall, and F-measure. This approach bridges traditional keyword-based methods and modern semantic-aware retrieval systems, offering a scalable and accurate solution for Quranic information retrieval in Indonesian. The overall framework of the proposed ensemble learning approach for query expansion and semantic retrieval in Quranic information systems is illustrated in Fig. 2 below.



Fig. 2. Research methodology.

As described previously, this study is structured into several interconnected stages, starting with the ontology development. At this stage, the collected and preprocessed data is transformed into unique thematic according to user needs. The ontology development process was based on a thematic classification encompassing 14 core topics: Morals and Etiquette (Akhlaq and Adab), The Qur'an, Previous Nations, Criminal Law (Jinayah), Private Law, Worship, Knowledge, Faith, Jihad, Food and Drink, Transactions (Mu'amalat), Clothing and Adornment, Judiciary and Judges, and History. These topics represent essential thematic divisions that facilitate a structured and systematic approach to understanding and accessing the teachings of the Qur'an. In addition, linguistic differences between Arabic and Indonesian are studied to address translation nuances and contextual challenges, which forms the basis for query expansion tailored

to the semantics of the Quran. The Quran Ontology is then developed by referring to this analysis.

A structured ontology is constructed to organize Quranic content systematically, reflecting the semantic relationships between verses and thematic categories. This process involves categorizing verses into specific topics, defining synonyms, antonyms, and hierarchical relationships, and ensuring alignment with the linguistic nuances of Indonesian translations. The use of ontology development tools, such as Protégé, aids in managing and visualizing the ontology structure, while consultations with Islamic scholars ensure the accuracy and relevance of the content. This ontology serves as the core mechanism for query expansion, enabling the system to infer implicit relationships and enhance search relevance.

Ontology development for the Al-Qur'an involves creating a structured representation of Quranic content by defining broad classes i.e., Morals, Worship and more specific subclasses such as Ethics and Rituals to categorize and refine Quranic teachings. Each class and subclass is linked through hierarchical and semantic relationships, which help capture how different topics interrelate, such as Faith being related to Worship. Properties and attributes are then assigned to these classes to provide deeper insights, such as Virtue and Integrity for the Morals class. This process ensures a comprehensive and accurate reflection of Quranic teachings.

The ontology is then aligned with the actual content of the Quran, where each verse is annotated and categorized under its relevant topic. This step involves ensuring that every verse is properly associated with the appropriate class and subclass, allowing for accurate semantic search results. Once validated and refined with feedback from domain experts, the ontology serves as a foundation for enhancing information retrieval, helping to expand queries and provide more relevant, contextually accurate search results from the Quran.

The second stage focuses on the implementation of query expansion using the ontology. When a user submits a query, the ontology dynamically enriches it by identifying and adding semantically related terms or concepts. For instance, a query about "worship" could be expanded to include related terms like "prayer," "fasting," or "charity," reflecting the thematic connections in the Quran. Advanced algorithms are applied to ensure that only the most contextually relevant terms are selected for expansion. This enriched query is then processed by the retrieval engine, ensuring improved relevance and context-awareness in search results. Iterative testing is conducted to refine the expansion process and maintain the quality of retrieved information.

To further optimize the retrieval process, the application of ensemble semantic text representation is introduced in this second stage. This approach focuses on combining traditional word embedding techniques, such as Word2Vec, GloVe, and FastText, to create a robust and versatile text representation framework. These methods are integrated using an ensemble method based on weighted voting, ensuring that each technique contributes to the final representation according to its strengths in capturing specific semantic aspects of the Quranic text.

Word2Vec generates dense vector representations for words by analyzing their co-occurrence within a fixed context window, effectively capturing semantic similarity between terms. This technique excels in identifying relationships between frequently co-occurring words, such as "prayer" and "worship," making it particularly effective for extracting context-dependent connections. GloVe (Global Vectors for Word Representation), on the other hand, extends this capability by considering global word co-occurrence statistics across the entire dataset. This allows GloVe to encode broad semantic relationships, such as linking "faith" and "belief" based on their shared conceptual roles in the Quran. FastText complements these methods by representing words as a composition of character-level n-grams, enabling it to capture subword information and morphological variations. This is particularly useful for handling linguistic nuances in Quranic translations, such as connecting "guidance," "guiding," and "guided" based on shared subword patterns.

To integrate these techniques, an ensemble strategy based on weighted voting is employed. Each embedding method is assigned a weight proportional to its ability to contribute to the task, as determined through empirical evaluation. For instance, Word2Vec might be weighted higher for its effectiveness in capturing context-dependent word relationships, while FastText may receive greater weight for handling morphological variants and rare words. When a query is processed, the embeddings generated by Word2Vec, GloVe, and FastText are combined, and the weighted scores are used to determine the relevance of Quranic verses to the query. For example, for a query about "worship," Word2Vec might identify verses containing contextually related terms like "prayer," GloVe might highlight conceptual links to "obedience," and FastText could include morphological variants like "worshiper." The weighted voting mechanism ensures that the final result reflects the best contributions of each embedding method.

This ensemble approach, guided by weighted voting, provides a powerful framework for addressing challenges such as synonymy, polysemy, and linguistic variations in Indonesian translations of the Quran. By combining local context (Word2Vec), global context (GloVe), and morphological robustness (FastText) with a carefully calibrated weighting scheme, the system achieves high accuracy and relevance in retrieval. This ensures that users receive contextually rich and semantically aligned results, making the Quranic information retrieval system both precise and comprehensive.

Finally, the system undergoes performance evaluation to assess its effectiveness. Metrics such as precision, recall, and F-measure are used to quantify the system's ability to deliver relevant results while minimizing irrelevant ones. Comparative experiments benchmark the proposed system against traditional methods, such as keyword-based searches, to demonstrate its advantages. User studies provide qualitative insights into the system's usability and relevance, ensuring its practical application for Quranic information retrieval. This interconnected workflow ensures the development of a scalable, context-aware, and highly accurate system tailored to

the needs of users searching for Quranic content in Indonesian.

IV. RESULT AND DISCUSSION

This section highlights the research findings, focusing on the development of a Qur'anic ontology and the implementation of semantic query expansion to enhance a thematic-based search system. The results are structured around key stages of the study, including ontology construction, application of query expansion techniques, system integration, and performance evaluation. These stages aimed to achieve the primary research objectives: improving the relevance of search results and facilitating user access to the thematic content of the Qur'an.

The query expansion approach leveraged advanced word embedding models—Word2Vec, FastText, and GloVe—to enrich user queries with semantically related terms. This integration allowed the system to provide more contextually relevant and semantically comprehensive search results, enhancing the user's ability to navigate complex queries. Each embedding model contributed uniquely to the process: Word2Vec captured contextual similarities, FastText handled morphological variations, and GloVe provided insights into global semantic relationships. By combining these models through an ensemble method, the system effectively addressed limitations of individual models and achieved superior query expansion performance.

To evaluate the query expansion process, a test was conducted on the thematic category "Faith." Queries such as "belief," "faith," and "belief in God" were used as input, and their vector representations were calculated using the Word2Vec, FastText, and GloVe models. Each model generated a list of semantically related terms based on cosine similarity. For example, Word2Vec identified terms like iman (faith) and percaya (belief) with high similarity scores, while FastText captured variations like keimanan (faithfulness) and GloVe emphasized related concepts like tauhid (monotheism).

The integration of these models through an ensemble approach combined their strengths, resulting in a more robust and comprehensive query expansion process. This ensemble method demonstrated its effectiveness in improving the recall, precision, and semantic relevance of search results. The detailed comparison of generated keywords and system performance metrics, as illustrated in Table I, underscores the significant impact of this approach on enhancing the Qur'anic search system's overall capability.

Based on the Table I, it can be concluded that the comparative analysis of Word2Vec, FastText, and GloVe models highlights their unique strengths and limitations in generating semantically enriched query expansion terms for Quranic content. Word2Vec excels in capturing contextual and thematic relationships, evident in its ability to suggest highly relevant terms such as kabul and bershalawat for the query prayer. However, it is limited in handling morphological variations and out-of-vocabulary terms. In contrast, FastText demonstrates superior handling of morphological diversity, as seen in its accurate generation of terms like berpuasa and berpuasa for the query fasting, leveraging its subword-based

architecture. Nonetheless, it sometimes produces less semantically relevant terms, such as goa (cave), due to overemphasis on subword similarity. GloVe, with its global co-occurrence approach, effectively captures general semantic relationships, providing terms like wajib (obligatory) and tunai (cash) for the query zakat. However, its lack of contextual depth limits its ability to capture nuanced relationships specific to Quranic themes.

TABLE I. TOP 5 GENERATED SEMANTIC KEYWORDS

Query	Word2Vec	FastText	GloVe
Prayer	kabul / 0.876	berdoa / 0.779	panjat / 0.686
	moga_allah / 0.808	goa / 0.639	berdo / 0.661
	bershalawat / 0.792	mohon / 0.636	kabul / 0.646
	malaikat / 0.778	allahummaghfir / 0.635	do / 0.64
	amin / 0.768	do / 0.632	mohon / 0.638
Zakat	mungut / 0.853	zakatnya / 0.933	tunai / 0.632
	ekor_kambing / 0.848	zakatnya / 0.915	wajib / 0.628
	fitrah / 0.837	zakaia / 0.76	amil / 0.551
	lima / 0.828	mufakat / 0.665	tugas / 0.545
	wasaq / 0.806	zakar / 0.66	lima / 0.541
Fasting	ramadan / 0.924	berpuasa / 0.966	ramadhan / 0.657
	ramadhan / 0.891	bepuasa / 0.945	buka / 0.654
	buka / 0.865	puas / 0.779	asyura / 0.616
	ganti / 0.735	kekurangpuasan / 0.738	ramadhan / 0.609
	hari / 0.67	ketidakpuasan / 0.707	hari / 0.592

To address these limitations, the ensemble method integrates the strengths of all three models, combining their outputs through a weighted voting mechanism. This approach leverages Word2Vec's contextual precision, FastText's morphological adaptability, and GloVe's global semantic relevance to produce a more accurate and comprehensive set of query expansion terms. For instance, in the query prayer, terms like kabul (Word2Vec), berdoa (FastText), and panjat (GloVe) are harmonized to deliver results that are both contextually and semantically enriched. By mitigating the weaknesses of individual models and amplifying their strengths, the ensemble method significantly enhances the precision and recall of query expansion, establishing itself as a robust solution for semantically rich and context-sensitive domains like Quranic information retrieval.

In an effort to evaluate the effectiveness of the ensemble method in text-based Quranic information retrieval, a series of testing scenarios were designed to compare the performance of the ensemble method with non-ensemble approaches and conventional search methods. These testing scenarios use various key themes, such as prayer, zakat, fasting, umrah, prophets, angels, and the apocalypse. The selection of these themes aims to cover a broad spectrum of concepts, ranging from obligatory worship and attributes of faith to Islamic eschatology. Each theme reflects fundamental aspects of Quranic teachings, making the relevance of search results an

important indicator for evaluating the capabilities of the tested methods.

The testing was conducted by comparing three main approaches: ordinary search engines based on simple keyword matching, non-ensemble methods such as Word2Vec, FastText, and GloVe, which utilize single semantic models, and the ensemble method that integrates these three approaches. Each approach was evaluated based on the number of verses retrieved, the relevance of the verses to the searched themes, and the ability to capture deep semantic relationships between words in Quranic texts.

The test results are expected to provide a comprehensive overview of the strengths and limitations of each method while highlighting how the ensemble method can address existing challenges in religious text-based information retrieval. Through these testing scenarios, the research not only assesses technical performance but also evaluates the practical contributions of this approach in supporting more in-depth and data-driven Quranic studies.

TABLE II. PERFORMANCE COMPARISON

Topic	Ordinary Search Engine	Word2Vec	FastText	GloVe	Ensemble Method
Prayer	15 verse	20 verse	20 verse	20 verse	25 verse
Zakat	15 verse	20 verse	20 verse	20 verse	25 verse
Fasting	15 verse	20 verse	20 verse	20 verse	25 verse
Umroh	15 verse	20 verse	20 verse	20 verse	25 verse
Angels	15 verse	20 verse	20 verse	20 verse	25 verse

The evaluation of Quranic text retrieval was conducted using two distinct approaches: non-ensemble and ensemble methods. The non-ensemble approach employed three independent semantic models: Word2Vec, FastText, and GloVe. Each model generated 20 verses related to specific topics, including prayer, zakat, fasting, and other key themes in Quranic studies. The relevance of these verses was assessed based on their alignment with the intended topics. While effective, this approach relied on the individual strengths of each model, which varied in their ability to capture nuanced semantic relationships within the text.

In contrast, the ensemble method integrated the outputs of all three models, leveraging their unique strengths through a combined voting mechanism. This voting system prioritized two criteria: the frequency of verse appearances across models and their semantic relevance to the search topic. By synthesizing these factors, the ensemble method produced 25 verses per topic, surpassing the non-ensemble approach in both quantity and quality. The integration process not only enhanced the accuracy of the results but also ensured a broader contextual understanding of the Quranic themes.

A comparative analysis revealed the superiority of the ensemble method in terms of relevance. The ensemble approach achieved an average relevance rate of 88%, significantly outperforming individual models such as

Word2Vec (75%), FastText (80%), and GloVe (82%). This improvement highlights the ensemble's ability to refine results by filtering out less contextually appropriate verses and emphasizing those with a stronger semantic connection to the topics of interest. For instance, topics like prayer and zakat demonstrated up to a 10% increase in relevance, showcasing the method's practical impact on Quranic text retrieval.

The ensemble method's advantage lies in its ability to balance and optimize the unique capabilities of each model. Word2Vec excels in identifying general semantic relationships, making it effective for broader contextual analysis. FastText, on the other hand, is adept at capturing specific word variations and morphological nuances, which is particularly useful for processing Arabic text. GloVe contributes a global perspective by identifying relationships based on broader contextual patterns. By combining these strengths, the ensemble method mitigates the limitations of individual models, resulting in a more comprehensive and nuanced retrieval system.

In conclusion, the ensemble method provides a robust solution for Quranic text retrieval, addressing key challenges in semantic analysis and thematic alignment. Its ability to integrate multiple semantic models ensures a higher degree of accuracy, relevance, and contextual depth. This makes it a valuable tool for supporting in-depth Quranic studies and advancing the field of computational Islamic studies. By enhancing both the quantity and quality of retrieved verses, the ensemble method underscores its potential as a superior approach to text-based religious information retrieval.

Regarding to the implementation of ontology concept, it can be concluded that ontology-based systems show a marked improvement in Morals and Etiquette (Akhlak and Adab), where the contextual meanings related to moral behavior and Islamic ethics are effectively captured. Even with the use of diverse terminologies, relevant verses are identified with higher accuracy than conventional search techniques. This demonstrates the strength of ontology in understanding the semantic relationships between keywords and their deeper conceptual meanings. Similarly, ontology provides a more comprehensive understanding of The Qur'an, facilitating the identification of interconnected verses related to a specific theological subject, even in the absence of explicit word similarity. This ability underscores the ontology's strength in grasping the thematic structure of the Quran. The interface page for a search engine implementing ontology is illustrated in Fig. 3 below.



Fig. 3. Interface of ontology search engine.

Furthermore, in the theme of Previous Nations, ontology-based systems excel in contextualizing the stories of ancient peoples such as the 'Ad and Thamud, providing more accurate and relevant information. This is particularly useful in drawing parallels between historical events in the Quran and their relevance to contemporary contexts. Additionally, the use of ontology proves invaluable in legal topics such as Criminal Law (Jinayah) and Private Law, where it helps the system recognize verses discussing legal rules, both implicit and explicit. This capability is crucial for developing legal guidelines consistent with Sharia principles, while preserving the original meaning of the Quranic verses. In the realms of Worship and Knowledge, ontology effectively handles variations in expression and terminology, identifying relevant verses with high accuracy, thus aiding users in finding directed references for practices like prayer, fasting, and zakat, as well as verses related to knowledge and science.

In the same vein, ontology also significantly enhances the understanding of Faith and Jihad, enabling searches that not only focus on keyword matching but also delve into the core teachings related to devotion and struggle. Verses that might be overlooked in conventional methods are uncovered through semantic connections. Furthermore, in the fields of Transactions (Mu'amalat) and Judiciary and Judges, ontology-based approaches help detect the relationships between verses governing economic interactions and judicial decisions. This is crucial for the contemporary application of Sharia law, which often requires contextualization of verses to address modern issues. Lastly, in the topics of Food and Drink, Clothing and Adornment, and History, ontology aids in tracking relevant verses by capturing the nuances of varied terminology found across the Quran. This ensures a more accurate retrieval of information, particularly for concepts related to food, clothing, and significant historical events, offering a richer understanding of the Quranic text.

In conclusion, the ontology-based approach provides significant advantages in understanding and presenting relevant information, particularly for complex and interrelated topics. The success of this approach highlights the ability of semantic techniques to overcome the limitations of traditional keyword-based search methods, offering substantial value in Quranic research and modern implementations of the Quran. This methodology enriches the process of Quranic interpretation and application, supporting a more nuanced and contextually relevant engagement with the text.

V. CONCLUSION

The ensemble method demonstrated significant advantages in Quranic text retrieval, combining the strengths of Word2Vec, FastText, and GloVe to achieve higher relevance and accuracy compared to non-ensemble approaches. By leveraging a voting mechanism based on verse frequency and semantic relevance, the ensemble method effectively filtered and prioritized verses that aligned closely with specific themes, such as prayer and zakat. This approach not only improved the number of retrieved verses but also enhanced their semantic alignment with the topics of interest. The findings underscore the ensemble method's potential as a

superior solution for text-based Quranic studies, offering a robust framework for semantic analysis.

Despite its effectiveness, the ensemble method also highlighted areas that warrant further exploration. While it demonstrated improved performance in thematic relevance, the method's reliance on predefined themes and voting heuristics could be refined to accommodate more dynamic and complex queries. Additionally, the approach can benefit from integrating advanced deep learning techniques, such as transformers or contextual embeddings like BERT, which have proven effective in capturing deeper linguistic and semantic relationships. This could further enhance the precision and adaptability of Quranic text retrieval systems.

Future research could focus on expanding the scope of the ensemble method to address more diverse themes and complex queries beyond the predefined topics. Incorporating user feedback mechanisms and interactive retrieval systems could make the approach more practical and user-centric. Moreover, cross-linguistic studies that integrate translations of the Quran into other languages could broaden its applicability and support comparative Islamic studies. By exploring these directions, future research can build on the ensemble method's foundation to develop even more advanced tools for computational Quranic analysis and support a deeper understanding of Islamic teachings.

REFERENCES

- [1] F. Mo et al., "A Survey of Conversational Search," arXiv Prepr. arXiv:2410.15576, 2024.
- [2] E. A. Stathopoulos, A. I. Karageorgiadis, A. Kokkalas, S. Diplaris, S. Vrochidis, and I. Kompatsiaris, "A Query Expansion Benchmark on Social Media Information Retrieval: Which Methodology Performs Best and Aligns with Semantics?," *Computers*, vol. 12, no. 6, p. 119, 2023.
- [3] M. Esposito, E. Damiano, A. Minutolo, G. De Pietro, and H. Fujita, "Hybrid query expansion using lexical resources and word embeddings for sentence retrieval in question answering," *Inf. Sci. (Ny)*, vol. 514, pp. 88–105, 2020.
- [4] J. Dalton, L. Dietz, and J. Allan, "Entity query feature expansion using knowledge base links," in *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, 2014, pp. 365–374.
- [5] I. Jurisica, J. Mylopoulos, and E. Yu, "Ontologies for knowledge management: an information systems perspective," *Knowl. Inf. Syst.*, vol. 6, pp. 380–401, 2004.
- [6] F. Demoly, K.-Y. Kim, and I. Horváth, "Ontological engineering for supporting semantic reasoning in design: deriving models based on ontologies for supporting engineering design," *Journal of engineering design*, vol. 30, no. 10–12. Taylor & Francis, pp. 405–416, 2019.
- [7] A. Doan, R. Ramakrishnan, and S. Vaithyanathan, "Managing information extraction: state of the art and research directions," in *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, 2006, pp. 799–800.
- [8] A. Roshdi and A. Roohparvar, "Information retrieval techniques and applications," *Int. J. Comput. Networks Commun. Secur.*, vol. 3, no. 9, pp. 373–377, 2015.
- [9] A. F. Smeaton, "An overview of information retrieval," *Inf. Retr. Hypertext*, pp. 3–25, 1996.
- [10] V. Gupta, D. K. Sharma, and A. Dixit, "Review of information retrieval: Models, performance evaluation techniques and applications," *Int. J. Sensors Wirel. Commun. Control*, vol. 11, no. 9, pp. 896–909, 2021.
- [11] F. A. Ruambo and M. R. Nicholas, "Towards enhancing information retrieval systems: A brief survey of strategies and challenges," in 2019

- 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT), IEEE, 2019, pp. 1–8.
- [12] K. Keyvan and J. X. Huang, “How to approach ambiguous queries in conversational search: A survey of techniques, approaches, tools, and challenges,” *ACM Comput. Surv.*, vol. 55, no. 6, pp. 1–40, 2022.
- [13] F. Özcan, A. Quamar, J. Sen, C. Lei, and V. Efthymiou, “State of the art and open challenges in natural language interfaces to data,” in *Proceedings of the 2020 ACM SIGMOD international conference on management of data*, 2020, pp. 2629–2636.
- [14] H. K. Azad and A. Deepak, “Query Expansion Techniques for Information Retrieval: a Survey,” 2019.
- [15] M. A. Raza, R. Mokhtar, N. Ahmad, M. Pasha, and U. Pasha, “A taxonomy and survey of semantic approaches for query expansion,” *IEEE Access*, vol. 7, pp. 17823–17833, 2019.
- [16] M. A. Raza, R. Mokhtar, and N. Ahmad, “A survey of statistical approaches for query expansion,” *Knowl. Inf. Syst.*, vol. 61, pp. 1–25, 2019.
- [17] J. Bhogal, A. MacFarlane, and P. Smith, “A review of ontology based query expansion,” *Inf. Process. Manag.*, vol. 43, no. 4, pp. 866–886, 2007.
- [18] P. J. Worth, “Word embeddings and semantic spaces in natural language processing,” *Int. J. Intell. Sci.*, vol. 13, no. 1, pp. 1–21, 2023.
- [19] J. Guo, Y. Cai, Y. Fan, F. Sun, R. Zhang, and X. Cheng, “Semantic models for the first-stage retrieval: A comprehensive review,” *ACM Trans. Inf. Syst.*, vol. 40, no. 4, pp. 1–42, 2022.
- [20] Y. Zhang et al., “Neural information retrieval: A literature review,” *arXiv Prepr. arXiv1611.06792*, 2016.
- [21] K. A. Hambarde and H. Proenca, “Information retrieval: recent advances and beyond,” *IEEE Access*, 2023.
- [22] D. Chandrasekaran and V. Mago, “Evolution of semantic similarity—a survey,” *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–37, 2021.
- [23] S. Sasaki, J. Suzuki, and K. Inui, “Subword-Based compact reconstruction for open-vocabulary neural word embeddings,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 3551–3564, 2021.
- [24] A. Fesseha, S. Xiong, E. D. Emiru, M. Diallo, and A. Dahou, “Text classification based on convolutional neural networks and word embedding for low-resource languages: Tigrinya,” *Information*, vol. 12, no. 2, p. 52, 2021.
- [25] L. Gan, Z. Teng, Y. Zhang, L. Zhu, F. Wu, and Y. Yang, “Semglove: Semantic co-occurrences for glove from bert,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 2696–2704, 2022.
- [26] S. Anjali Devi and S. Sivakumar, “An efficient contextual glove feature extraction model on large textual databases,” *Int. J. Speech Technol.*, pp. 1–10, 2022.
- [27] G. Curto, M. F. Jojoa Acosta, F. Comim, and B. Garcia-Zapirain, “Are AI systems biased against the poor? A machine learning analysis using Word2Vec and GloVe embeddings,” *AI Soc.*, vol. 39, no. 2, pp. 617–632, 2024.
- [28] R. Biswas and S. De, “A Comparative Study on Improving Word Embeddings Beyond Word2Vec and GloVe,” in *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)*, IEEE, 2022, pp. 113–118.
- [29] L. Elvitaria et al., “A Proposed Batik Automatic Classification System Based on Ensemble Deep Learning and GLCM Feature Extraction Method,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 10, 2024.
- [30] M. Fernández, I. Cantador, V. López, D. Vallet, P. Castells, and E. Motta, “Semantically enhanced information retrieval: An ontology-based approach,” *J. Web Semant.*, vol. 9, no. 4, pp. 434–452, 2011.
- [31] E. H. Mohamed and E. M. Shokry, “QSST: A Quranic Semantic Search Tool based on word embedding,” *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 3, pp. 934–945, 2022, doi: 10.1016/j.jksuci.2020.01.004.
- [32] A. Hakkoum and S. Raghay, “Advanced search in the Qur’an using semantic modeling,” in *Proceedings of IEEE/ACS International Conference on Computer Systems and Applications, AICCSA*, IEEE, Nov. 2016, pp. 1–4. doi: 10.1109/AICCSA.2015.7507259.
- [33] M. I. E. K. Ghembaza, “Specialized Quranic Semantic Search Engine,” *Int. J. Comput. Sci. Inf. Secur.*, vol. 17, no. 2, 2019.
- [34] A. Hakkoum and S. Raghay, “Advanced Search in the Qur’an using Semantic modeling,” in *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, IEEE, 2015, pp. 1–4.
- [35] A. Abdullahi, N. A. Samsudin, M. H. A. Rahim, S. K. A. Khalid, and R. Efendi, “Multi-label classification approach for Quranic verses labeling,” *Indones. J. Electr. Eng. Comput. Sci.*, vol. 24, no. 1, pp. 484–490, 2021, doi: 10.11591/ijeecs.v24.i1.pp484-490.
- [36] F. Beirade, “Search engine for Holy Quran,” *2014 4th Int. Symp. ISKO-Maghreb Concepts Tools Knowl. Manag. ISKO-Maghreb 2014*, pp. 1–6, 2015, doi: 10.1109/ISKO-Maghreb.2014.7033477.
- [37] Z. Indra, A. Adnan, and R. Salambue, “A Hybrid Information Retrieval for Indonesian Translation of Quran by Using Single Pass Clustering Algorithm,” in *Proceedings of 2019 4th International Conference on Informatics and Computing, ICIC 2019*, IEEE, 2019, pp. 1–5. doi: 10.1109/ICIC47613.2019.8985737.
- [38] S. K. Hamed and M. J. A. Aziz, “A question answering system on Holy Quran translation based on question expansion technique and Neural Network classification,” *J. Comput. Sci.*, vol. 12, no. 3, pp. 169–177, 2016, doi: 10.3844/jcssp.2016.169.177.
- [39] R. H. Gusmita, Y. Durachman, S. Harun, A. F. Firmansyah, H. T. Sukmana, and A. Suhaimi, “A rule-based question answering system on relevant documents of Indonesian Quran Translation,” in *2014 International Conference on Cyber and IT Service Management, CITSM 2014*, IEEE, 2014, pp. 104–107. doi: 10.1109/CITSM.2014.7042185.
- [40] F. E.M.A, R. N.S, and A. Syukri, “Development of Qur’an Search Engine For The Indonesian Language Query,” in *Proceedings of the 2nd International Conference on Quran and Hadith Studies Information Technology and Media in Conjunction with the 1st International Conference on Islam, Science and Technology, ICONQUHAS & ICONIST, Bandung, October 2-4, 2018, Indonesia*, 2020. doi: 10.4108/eai.2-10-2018.2295579.
- [41] F. E. M. Agustin, M. H. R. Maulidi, R. H. Gusmita, R. C. N. Santi, M. Ulfa, and R. Sugara, “Applying of Quranic Glossary Approach to Improve Indonesian Qur’an Translation Search Engine Performance,” in *2020 8th International Conference on Cyber and IT Service Management, CITSM 2020*, IEEE, 2020, pp. 1–5. doi: 10.1109/CITSM50537.2020.9268820.
- [42] A. R. G. Purnama, I. N. Yulita, and A. Helen, “Search System for Translation of Al-Qur’an Verses in Indonesian using BM25 and Semantic Query Expansion,” in *2021 International Conference on Artificial Intelligence and Big Data Analytics, ICAIBDA 2021*, IEEE, 2021, pp. 214–220. doi: 10.1109/ICAIBDA53487.2021.9689757.

A Review of Reinforcement Learning Evolution: Taxonomy, Challenges and Emerging Solutions

Ji Loun Tan¹, Bakr Ahmed Taha², Norazreen Abd Aziz³, Mohd Hadri Hafiz Mokhtar⁴, Muhammad Mukhlisin⁵,
Norhana Arsad^{6*}

Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti
Kebangsaan Malaysia, Bangi 43600, Selangor, Malaysia^{1, 2, 3, 4, 6}

Department of Civil Engineering, Politeknik Negeri Semarang, Jl. Prof. Soedarto SH, Tembalang,
Semarang, Jawa Tengah 50275, Indonesia⁵

Abstract—Reinforcement Learning (RL) has become a rapidly advancing field inside Artificial Intelligence (AI) and self-sufficient structures, revolutionizing the manner in which machines analyze and make selections. Over the past few years, RL has advanced notably with the improvement of more sophisticated algorithms and methodologies that address increasingly complicated actual-world troubles. This progress has been driven by using enhancements in computational power, the availability of big datasets, and improvements in machine-getting strategies, permitting RL to address challenges across a wide range of industries, from robotics and autonomous driving system to healthcare and finance. The effect of RL is evident in its capacity to optimize selection-making procedures in unsure and dynamic environments. By getting to know from interactions with the environment, RL agents can make decisions that maximize lengthy-time period rewards, adapting to converting situations and enhancing over time. This adaptability has made RL an invaluable tool in situations wherein traditional approaches fall brief, especially in complicated, excessive-dimensional spaces and behind-schedule remarks. This review aims to offer radical information about the current nation of RL, highlighting its interdisciplinary contributions and how it shapes the destiny of AI and autonomous technologies. It discusses how RL affects improvements in robotics, natural language processing, and recreation while exploring its deployment's ethical and practical demanding situations. Additionally, it examines key research from numerous fields that have contributed to RL's development.

Keywords—Artificial intelligence; autonomous systems; decision-making optimization; reinforcement learning; robotics

I. INTRODUCTION

Machine Learning (ML) is primarily categorized into three main types, which are Supervised Learning, Unsupervised Learning, and Reinforcement Learning (RL) [1, 2]. The primary goal of RL is to allow machines to acquire knowledge beyond the constraints of supervised and unsupervised learning paradigms [3]. RL commonly employs a reward function as a training mechanism for agents tasked with specific objectives [4]. Unlike other ML paradigms that rely on labeled datasets, RL derives knowledge through direct interaction with the environment [5]. To make the RL result more effective, it is

very important to ensure communication between the agents and the environment [6].

Historically, RL has evolved from early work in behavioral psychology and control theory to become a fundamental tool in artificial intelligence and robotics. The foundational work of Kaelbling et al. (1996) and subsequent advances such as deep Q-networks from Mnih et al. (2015) have set the foundation for developments of RL in future [4, 7]. Hence, nowadays RL research ranges from autonomous robots to complex decision-making systems. For example, RL is able to play an important role in improving robot autonomy and flexibility, especially in tasks like manipulation and navigation [8]. In addition, RL has proven valuable in enhancing autonomous vehicle control, improving safety and optimizing transportation systems [9, 10]. The application of RL is not only limited to robotics, but also been applied in the semiconductor industry, where it can optimize processes such as physical design routing [11, 12].

In recent years, the combination of reinforcement learning and deep learning, which is also known as deep reinforcement learning (DRL) [13, 14, 15]. DRL has driven to breakthroughs in solving high-dimensional problems, especially in game-playing AI such as Alpha Go and autonomous systems [16, 17]. In addition, emerging trends nowadays include multi-agent reinforcement learning (MARL) which multiple agents learn and collaborate in a shared environment and also healthcare applications, where RL able to shows promise in personalized medicine and treatment planning [18, 19].

Furthermore, current RL research is more likely to focus on improving sample efficiency, safety, and scalability for real-world applications. It is very important to investigate new ways to integrate Reinforcement Learning with other machine learning paradigms to produce more adaptive and generalizable AI systems. This review will explore the development of RL, classification of RL method, and highlight its modern applications and future research directions. Moreover, this review also will discuss the research contributions from various fields to describe the current state of Reinforcement Learning and its potential to drive innovation in artificial intelligence and autonomous systems. The applications of RL are shown in Fig. 1.

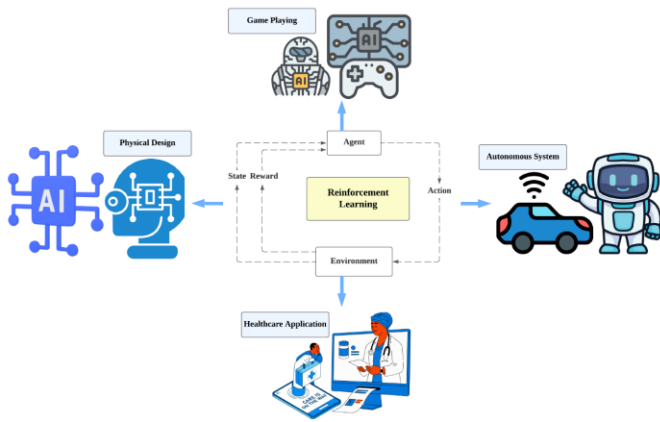


Fig. 1. Applications of reinforcement learning.

In this review, we have a look at the speedy improvement of RL and its developing applicability to complex, actual global troubles. We begin by exploring the evolution of RL techniques, from foundational strategies to advanced strategies consisting of Deep Reinforcement Learning and multi-agent systems. It also categorizes RL techniques, distinguishing between model-unfastened and model-based totally tactics, and highlights their respective strengths and barriers. Furthermore, we cover various RL programs across numerous domain names, including self-sustaining structures, robotics, healthcare, and synthetic intelligence. This exploration aims to offer a comprehensive understanding of the contemporary state of RL, its interdisciplinary contributions, and its potential to pressure destiny innovations in AI and self-reliant technologies.

II. EMERGING TRENDS OF REINFORCEMENT LEARNING EVOLUTION

The evolution of Reinforcement Learning (RL) is based in multiple foundational fields, which include behavioral psychology, trial-and-error learning, optimal control theory and dynamic programming. These parallel developments have provided the foundation for modern RL and shaping its principles and algorithms.

A. Behavioural Psychology and Trial-and-Error Learning

The earliest roots of Reinforcement Learning (RL) can be traced to behavioral psychology, specifically the works of Edward Thorndike and B.F. Skinner before the timeframe of 1960s [20, 21, 22]. Thorndike's Law of Effect introduced the concept of learning from the consequences of actions, where actions followed by satisfying outcomes are more likely to be repeated [20, 23]. This was the basis for trial-and-error learning, which is one of the important aspects of RL, where an agent explores different actions and adapts its behavior based on rewards or punishment. Furthermore, B.F. Skinner demonstrated that operant behavior can be shaped through reinforcement mechanisms, which highlighted the importance of rewards and punishments in learning [24]. This psychological perspective shows the foundation for how RL agents learn to optimize their actions by maximizing rewards or minimizing punishment.

B. Optimal Control Theory

During the mid-20th century, there were major advances in optimal control theory, especially in the field of engineering [25]. Control theory usually focuses on designing controllers that can guide dynamic systems to perform specific tasks in an optimal manner. The Bellman equations which were proposed by Richard Bellman in 1957 became central to this optimal control framework [26], [27], [28]. This work on dynamic programming provided a way to decompose complex decision problems into simpler subproblems, which enables the computation of optimal policies in environments with known dynamics and become a foundation for RL. Hence, Richard Bellman's work directly influenced by introducing the concept of a value function and estimates the expected future reward of being in a particular state and taking a particular action. This concept is basic for the RL algorithms such as Q-learning and value iteration [29].

C. Dynamic Programming and Modern Developments

From the foundation of Bellman's dynamic programming as mentioned, Ronald Howard introduced Markov decision processes (MDP) which is a mathematical framework for modeling decisions in environments where outcomes are partially random and partially controlled by the decision maker [30]. The formalization of MDP set the foundation for modern RL algorithms, as it represents the interaction between an agent and its environment, where the agent seeks to maximize a cumulative reward over time, positive rewards are awarded for favorable actions, while negative rewards or punishments are assigned for undesirable actions. These reward mechanisms are used as evaluative feedback and enable the agent to assess its actions within a specific state and learn from accumulated experiences. However, dynamic programming methods require information or knowledge of the dynamics of the environment, this disadvantage limits its applicability to real world problems. This gap flattens the way for RL techniques. For example, the model-free learning which will be discussed in next section, where an agent can learn optimal policies directly from interactions with the environment without required the knowledge of its dynamics.

Around the 1980s, RL emerged as a distinct field. For example, Sutton introduced temporal difference (TD) learning which is one of the key innovations that enabled agents to learn value functions from incomplete trajectories, rather than waiting until the end of an episode [31]. Furthermore, Watkins further advanced RL by allowing agents to directly learn action-value functions without the requirement for explicit models of the environment [32].

D. Deep Reinforcement Learning (DRL) Revolution

In the 2010s, the combination of Deep Learning (DL) and Reinforcement Learning (RL) which is known as Deep Reinforcement Learning (DRL) revolutionized the field again. In 2015, Google DeepMind researchers introduced the Deep Q Network (DQN) in 2015, which enabled reinforcement learning agents to handle complex tasks such as Atari 2600 games by using deep neural networks to approximate value functions [4]. This research highlighted the scalability of RL in higher dimensional state spaces and lead to major achievements such as the success of AlphaGo, which defeated

the one of the world champions of Go Lee Se-dol by using a combination of RL and DL [33].

E. Emerging Trends and Future Directions

Recent trends in Reinforcement Learning (RL) focus on improving sample efficiency, safety and scalability for real-world applications [34, 35]. New RL techniques such as Multi-agent Reinforcement Learning (MARL), hierarchical reinforcement learning and transfer learning are being explored to solve the issue of complex multi-agent environments where multiple agents learn collaboratively at the same time [18, 36, 37, 38]. In addition, RL can also be applied in more applications in different areas such as robotics, healthcare, finance and autonomous systems. The timeline of key evolution in RL development is shown in Fig. 2.

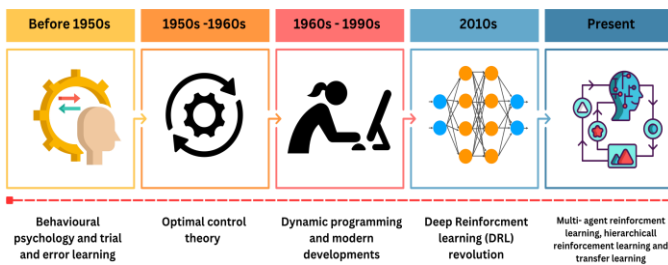


Fig. 2. Timeline of key evolution in RL development.

III. TAXONOMY AND CRITERIA OF REINFORCEMENT LEARNING

Reinforcement Learning (RL) techniques are primarily classified into Model-Based and Model-Free approaches within the framework of Markov Decision Processes (MDP). Model-based RL is further categorized into Given-the-Model and Learn-the-Model techniques. Meanwhile, Model-Free RL is subdivided into on-policy and off-policy approaches, which are discussed in the subsequent sections. In addition, value-based and policy-based approaches also have been discussed in this paper. The overview of RL classification is shown in Fig. 3.

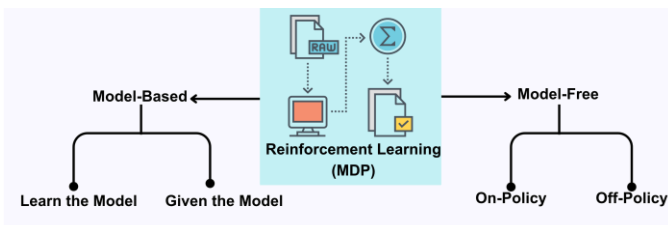


Fig. 3. Overview of reinforcement learning classification.

A. Model-based and Model-Free

Reinforcement Learning (RL) typically demands a substantial amount of data to attain satisfactory performance levels. This section will primarily focus on two types of RL algorithms which include model-free and model-based [39, 40]. Generally, model-free RL algorithms are considered as a direct approach, while model-based RL algorithms are viewed as an indirect method [41].

Model-free RL algorithms aim to learn a policy or value

function without explicitly constructing a model of the control system [42]. In contrast, model-based RL algorithms not only learn a value and policy function but also simultaneously construct an explicit model of the system [7]. There are two well-known model-free RL algorithms which are Q-Learning and Deep Q-Networks (DQN), where the agent learns value functions that estimate the expected cumulative rewards for each action in each state [4, 43, 44]. Based on the research conducted by Atkeson and Santamaria, a comparative study was undertaken using a linear double integrator movement task to assess data efficiency, the research findings indicate that the model-based RL algorithms surpass the model-free RL algorithms in terms of data efficiency [45].

In addition, the model-free RL algorithms do not train a model of the environment and aim to directly assign values to states or state-action [46, 47]. The agent directly interacts with the environment and enhances its performance based on the collected samples through exploration. It is easier to implement as they do not require explicit modeling of the environment but might be a problem that is hard to learn. However, model-free RL algorithms are usually hard to implement in real-world scenarios due to the time consumption and the cost [48]. This model-based RL is usually more suitable for large, complex environments but suffers from sample inefficiency as the agent learns only through interactions with the environment.

The advantage of a model-based RL algorithm includes its ability to predict future states and rewards through the explicit modelling of the environment [48, 49]. This will help the agent in making better planning and incorporating strategies like pure planning and expert iterations [50]. In model-based RL, the agent can simulate possible scenarios and plan its actions accordingly. However, model-based RL algorithms come with few disadvantages. One of the major challenges of model-based RL is that the model often depends on the accuracy of the transition model, it means that inaccurate models can lead to domain shift and poor performance [49, 51, 52, 53]. For model-based RL, developing and maintaining an accurate model of the environment can be complex and resource-intensive. Besides, the learned models may be inaccurate in practical scenarios, introducing bias in estimation [51]. When policy estimation and improvement are based on a biased model, the resulting policies may prove ineffective or even collapse when applied in the real environment. Hence, model-based methods often require significant computational cost for model learning and policy optimization hence it is limited application in real-world situations [54]. Lastly, model-based RL algorithms have the capacity to predict unexpected actions and states, which also provides a more controlled learning process.

In summary, model-free RL algorithms learn through exploration, whereas model-based RL algorithms learn by simulating scenarios [41]. The slight difference between model-based RL and model-free is shown in Fig. 4. In addition, the summary of the distinctions between Model-Based and Model-Free Reinforcement Learning (RL) algorithms is also shown in Table I.

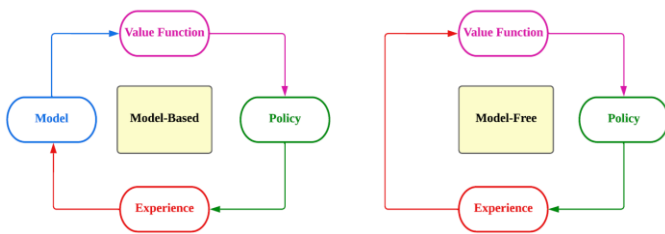


Fig. 4. Difference between model-based and model-free RL.

TABLE I. SUMMARY OF THE DISTINCTIONS BETWEEN MODEL-BASED AND MODEL-FREE RL ALGORITHMS

Category	Model-Based	Model-Free
Learning Type	Indirect method	Direct method
Objective	Learns value, and policy functions and constructs an explicit model of the system	Learns policy or value function without explicit model construction
Data Efficiency	Outperforms in terms of data efficiency	May demand substantial data for satisfactory performance
Implementation	Challenging in real-world scenarios due to complexity and cost	Easier implementation, no explicit modelling required
Environment Interaction	Predict future states and rewards through explicit environment modelling	Direct interaction, enhances performance through exploration
Challenges	Complex model construction may introduce bias for learned models	Hard to learn complex problems in real-world scenarios
Applicability to Real World	Balancing model accuracy and real-world complexity is a significant challenge	Hard to implement due to time and cost constraints

B. Model-based: Given the Model and Learn the Model

The algorithms that use models are called model-based methods. In model-based Reinforcement Learning (RL), given-the-model and learn-the-model are two main types of approaches [48]. One of the examples of the given-the-model is the AlphaGo algorithm [16]. In this algorithm, AlphaGo is explicitly learned the rules of the board game and can be described using coding or computer language. Then, the transitions and rewards are known to the agent, which allows for the evaluation of different strategies through trial and error to get optimal results and iteratively improve the policy. This approach shows that a pre-determined model of the environment to guide the learning process and enhance the decision-making. For another example, the Monte Carlo Tree Search (MCTS) algorithm can be using a given model to simulate possible future states and evaluate action sequences for optimal planning [55].

Moreover, an example of the "learn the model" category in model-based Reinforcement Learning is the World Models algorithm [56]. One of the examples for learn the model approach is DreamerV2 [57]. DreamerV2 builds an internal world model of the environment by learning from its experiences, which can allow the agent to simulate trajectories in its "imagination" rather than only rely on real-world

interactions. This can cause the agent to explore and learn optimal policies more efficiently, as it can try out different actions and observe hypothetical outcomes within its model, thus significantly reducing the need for real-world samples. Another example is Probabilistic Ensembles with Trajectory Sampling (PETS), which use the models to predict possible future states with uncertainties included [58]. PETS uses these learned dynamics to perform planning and action selection then helping the agent to handle uncertainty and make more robust decisions in those unpredictable environments. This approach allows the agent to improve sample efficiency by using imagined rollouts for planning while adapting to changes in real-world scenarios. The comparison of "Given-the-Model" and "Learn-the-Model" in model-based RL is shown in Table II.

TABLE II. COMPARISON BETWEEN "GIVEN THE MODEL" AND "LEARN THE MODEL"

Category	Given the Model	Learn the Model
Learning Type	Uses an explicitly specified model of the environment	Learns a model of the environment from gathered data
Decision-Making	Enhances decision-making through trial and error	Optimizes policies by leveraging insights from the model
Advantages	1. Allows for the evaluation of different strategies 2. Facilitates iterative improvement of the policy	1. Adapts to complex and unknown environments 2. Can generalize to various scenarios
Disadvantages	1. Require explicit specification of the environment 2. Limited adaptability with unforeseen changes	1. Require extensive data for accurate model learning 2. Complexity in training and interpreting the learned model

C. Model-Free: On-Policy and Off-Policy

Model-free Reinforcement Learning (RL) is typically categorized into on-policy and off-policy approaches [48, 41]. The on-policy approach strives to enhance and learn through the policy itself which is used for decision-making [59]. For the on-policy approach, the agent itself interacts with the environment, and the policy to interact with the environment and the improved policy remain the same. According to Singh et al.'s (2000) study, this policy algorithm could be more stringent because the updating of the value function is contingent on the experiences gained from implementing the policy [60].

On the other hand, the off-policy approach aims to improve a policy that is different from the one that is used to generate the data [48]. Unlike the on-policy approach, the off-policy approach does not require the same agent that interacts with the environment. The experiences of other agents interacting with the environment can also be utilized to enhance the policy. When the agent learns the behavior in one way is called the target policy, while when it is learned using data generated by another policy is known as the behavior policy [61]. In addition, the agent learns from data generated by a behavior policy that might be explored more widely than the target policy, which allows for more efficient learning. This flexibility allows for more diverse data sources to contribute to

the policy improvement process. In Fakoore et al. (2020) research, it is noted that off-policy methods encounter bias issues as the data from an outdated policy differs from the current policy, making it unsuitable to update the current policy's value function using old data [62]. On the other hand, on-policy methods avoid bias but may face variance challenges, tending to be more data-efficient as they focus on the current samples.

One of the examples of an on-policy approach is SARSA State-Action-Reward-State-Action (SARSA) [41, 48]. In the SARSA algorithm, an action is selected based on the current policy and executed. Then, the data is utilized to update the current policy. In the on-policy setting, the policy that interacts with the environment is the same as the updated policy, which ensures consistency between the policy used during interaction and the one that is improved. The SARSA update function is shown below:

$$Q\{S(t),A(t)\} \leftarrow Q\{S(t),A(t)\} + \alpha\{Q\{S(t+1),A(t+1)\} - Q\{S(t),A(t)\}\} \quad (1)$$

In this equation,

- $Q\{S(t),A(t)\}$ represents the Q-value for the state action pair at time t
- α is the learning rate
- $Q\{S(t+1),A(t+1)\}$ is the Q value for the next state action pair at time $t + 1$

This updated function shows how SARSA adjusts the Q-values based on the observed rewards and transitions, which also continues to refine the policy in an on-policy manner. In the off-policy category, one of the examples is Q-learning, which employs the max operation and a greedy policy when selecting actions [41, 43, 48]. In addition, Q-learning involves updating a policy that interacts with the environment and the updated policy which is not necessarily the same as the policy used during interaction.

The Q-learning update function is shown below:

$$Q\{S(t),A(t)\} \leftarrow Q\{S(t),A(t)\} + \alpha[R(t+1) + \gamma \max_{\alpha} Q\{S(t+1),A(t+1)\} - Q\{S(t),A(t)\}] \quad (2)$$

In this equation,

- $Q\{S(t),A(t)\}$ represents the Q-value for the state action pair at time t
- α is the learning rate
- $R(t+1)$ is the reward at time $t + 1$
- γ is the discount reward
- $\max_{\alpha} Q\{S(t+1),A(t+1)\}$ is the maximum Q value for the next state $S(t+1)$

The function above also reflects how Q-learning iteratively refines the Q values based on the observed rewards and transitions and improves the policy over time. The comparison between "On-Policy" and "Off-Policy" in model-free RL is shown in Table III.

TABLE III. SUMMARY OF THE COMPARISON BETWEEN "ON-POLICY" AND "OFF-POLICY"

Category	Given the Model	Learn the Model
Learning Type	The agent learns through the policy used for decision-making.	Aims to improve a policy different from the data-generating policy.
Environment Interaction	The agent interacts with the environment using the policy.	Doesn't require the same agent to interact and data from other agents can be used (shared).
Consistency	The policy used during interaction and improved policy are the same.	Involves a target policy (learned) and a behavior policy (data-generating).
Flexibility	Limited by the exploration of the current policy.	Learns from data generated by a behavior policy that may be explored more widely.

D. Value-based Approach and Policy-based Approach

The value-based approach typically involves learning the value function through methods such as Temporal difference (TD) learning, Q-Learning, or Deep Q-Network (DQN) [63, 64, 65, 66]. This technique aims to identify the optimal action to take and the action under this approach tends to be deterministic, such that they are chosen with a clear understanding of consequences. The value function operates by working backward from the target state and attributing rewards to the preceding state. This approach can be helped in the selection of only one action that leads towards achieving the desired outcome and closer to the goal [65]. In summary, it involves a strategic evaluation of the value of actions to make informed decisions and optimize the learning process.

In contrast to the value-based approach, the policy-based approach focuses on learning the conditional probability π of a policy through techniques such as the policy gradient method [67]. Instead of obtaining a value function like mentioned above, this approach directly determines the policy. Due to the stochastic action probability, the policy-based approach is more suitable for the application with large and continuous action [65]. At the same time, action selection becomes probabilistic with actions chosen based on their likelihood of efficiently reaching the desired outcome as dictated by the learned policy.

IV. THE ADVANCEMENT OF REINFORCEMENT LEARNING APPROACHES

In recent years, Reinforcement Learning (RL) has been improving with a fast pace and developed advanced approaches for increasingly complex problems in the real world. This review would like to focus on Deep Reinforcement Learning, Hierarchical Reinforcement Learning, Multi-Agent Reinforcement Learning and Hybrid Model Based Reinforcement Learning. These approaches have expanded the range of high dimensional RL applications, multi-agent applications, hierarchical decision-making applications and optimal policies or strategies by a combination of model based and model free methods.

A. Deep Reinforcement Learning (DRL)

Deep Reinforcement Learning (DRL) combines Reinforcement Learning (RL) and deep learning to enable agents to learn optimal policies for decision-making tasks through trial and error [14, 15, 68, 69]. The DRL is known for utilizing the principle of RL with the theory of deep learning to facilitating those automatic extractions of features from the input and benefits for in the fields such as robotic, autonomous driving and video games [14, 70, 71].

One of the achievements in DRL was the development of Deep Q Network (DQN) by work of Mnih et al. (2015), which shows the human level performance on Atari games using raw pixel inputs. The DQN uses the convolutional network to approximate the Q value function, which trained using variant of Q-learning with experience and target network to stabilize training [4]. Subsequently, Schulman et al. (2015) have introduced the Trust Region Policy Optimization (TRPO), which addressing the not stable and inefficiency of policy gradient methods. TRPO ensures monotonic improvement by optimizing objective function subject to a trust region constraint and make it more stable and reliable training [13, 72]. Further refinement came with the Proximal Policy Optimization (PPO) work by Schuman et al. (2017), which simplified the algorithm and enhanced computational efficiency by using clipped objective to balance exploration and exploitation effectively [73]. Lillicrap et al. (2015) have also extended the actor-critic framework to continuous action spaces with the Deep Deterministic Policy Gradient (DDPG) algorithm. DDPG will employ the actor network to parametrize the policy and network to estimate the Q- value function which enabling the application of DRL such as robotic control task [74].

In addition, Kostrikov et al. (2021) proposed the Implicit Q-learning (IQL) algorithm, which is an offline Reinforcement Learning method that avoids evaluating unseen actions, thereby mitigating errors from distributional shift. By leveraging state-value functions as random variables and conditionally using the expected value of the state, IQL can improve the policy without directly querying actions from the distribution [75]. On other hand, Chen et al. (2022) have also proposed the DreamerV2 algorithm, which builds on the Dreamer framework by incorporating discrete latent variables and advanced world model. It is also able to demonstrate a similar performance on Atari benchmark with efficient performance [76]. Sekar et al. (2020) introduced the Plan2Explore algorithm, which emphasizes intrinsic exploration by using a self-supervised world model to plan for expected future novelty, enabling the agent to efficiently explore and quickly adapt to multiple downstream tasks [77].

In summary, the advancement of DRL has revolutionized the fields of RL by enabling the agents to learn from high dimensional inputs to perform complex tasks. The new algorithms such as DQN, TRPO, PPO, DDPG, IQL, DreamerV2 and Plan2Explore have advanced the application of RL in different fields including gaming, robotics and autonomous systems. As research in DRL continues to focus, the efficiency, stability and generalization of RL will have further improvement. Fig. 5 provides a schematic illustration of DQN and Plan2Explore, presented as a case study in DRL.

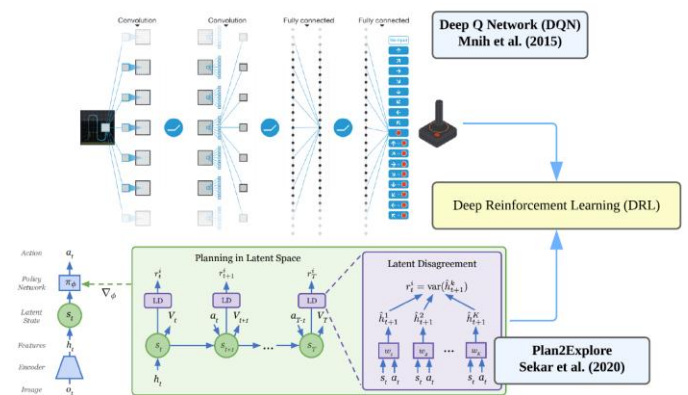


Fig. 5. Schematic illustration of Deep Q-Network (DQN) and Plan2Explore. [4, 77].

B. Hierarchical Reinforcement Learning (HRL)

Hierarchical Reinforcement Learning (HRL) is one of the approaches in Reinforcement Learning fields that able to addresses the challenges faced by traditional Reinforcement Learning method which include scalability and efficiency with the tasks that required long term planning [37, 78, 79]. HRL is able to solve these problems by decomposing them into hierarchy of subs tasks [37, 80]. The core idea of HRL can be described as hierarchical structure where higher level policies select sub tasks and lower level policies execute actions to achieve the goal of subs task.

One of the foundational works in HRL is Feudal Reinforcement Learning (FRL) by Dayan and Hinton (1992) which introduced a hierarchical structure where higher level manager set goals for lower level workers [81]. Each hierarchical level operates in different temporal and spatial resolution, allowing the agent to decompose complex tasks into simpler sub-tasks. In addition, another early work in HRL is the Options framework introduced by Sutton et al. (1999), this framework introduces the concept of options, which temporarily extended actions that consists of policy, termination conditions and initiation set [82]. These options can allow the agents to operate at different time scales and make the learning more efficient. Moreover, the more recent advancements in HRL include the Hierarchical-DQN (h-DQN) framework, which extends the DQN algorithm by incorporating hierarchical structure. The h-DQN framework consists of meta-controller that selects sub-goals and lower level controller which learns to achieve the sub goals using DQN, be able to apply to Atari games and navigating 3D environments [83]. The Options-Critic architecture by Bacon et al. (2017) provides a framework for learning both options and policies over options in end-to-end manner, the architecture introduces intra-option policy gradient methods to optimize the policies within options and termination conditions [84].

Furthermore, Vezhnevets et al. (2016) proposed the Strategic Attentive Writer (STRAW) framework, which is a deep recurrent neural network (RNN) architecture that capable learning macro-actions in a Reinforcement Learning setting. The model builds an internal plan and partitions it into sub-sequences then allowing the agent to commit to a plan for a period before replanning. This approach allowed the agent to

explore and compute efficiently across different tasks [85]. The Hierarchical Reinforcement Learning with Off-policy correction (HIRO) algorithm introduced by Nachum et al. (2018) addresses the challenges of non-stationarity in HRL by introducing an off-policy correction mechanism, enabling stable and efficient learning of hierarchical policies [79]. Levy et al. (2019) introduced the Hierarchical Actor-Critic (HAC) algorithm, which extends the actor-critic framework to a hierarchical setting by enabling agents to operate at multiple levels of abstraction simultaneously [86].

Recent developments after 2020 in HRL have further expanded, The Hierarchical Variational Autoencoder (HVAE) framework introduced by Bai et al. (2023) combines probabilistic generative models with deep neural networks to learn hierarchical topic representations for multi-view text documents. HVAE captures both local and global topical information, enabling efficient modelling of complex document structures [87]. The Hierarchical Deep Reinforcement Learning with Automatic Sub-Goal Identification via Computer Vision (HADS) by Liu et al. (2021) introduces a sub-goal generation mechanism that adapts to the agent’s learning progress, applied to tasks such as game manipulation and navigation [88]. The Hierarchical Deep Reinforcement Learning with Graph Neural Networks (HRLOrch) introduced by Li & Zhu (2021) uses graph neural networks to model the hierarchical structure of the environment at multiple levels of abstraction [89].

In summary, HRL has significantly advanced the RL field by mainly enabling the agents to decompose complex tasks to simpler sub-tasks, hence cause the learning and planning more efficiently. The development of HRL frameworks including Options framework, FRL, h-DQN, Option-Critic, STRAW, HIRO, HAC, HVAE, HADS and HRLOrch, which further improves in terms of efficiency, stability and generalization capabilities. The schematic illustration of h-DQN, STRAW and HRLOrch is shown in Fig. 6.

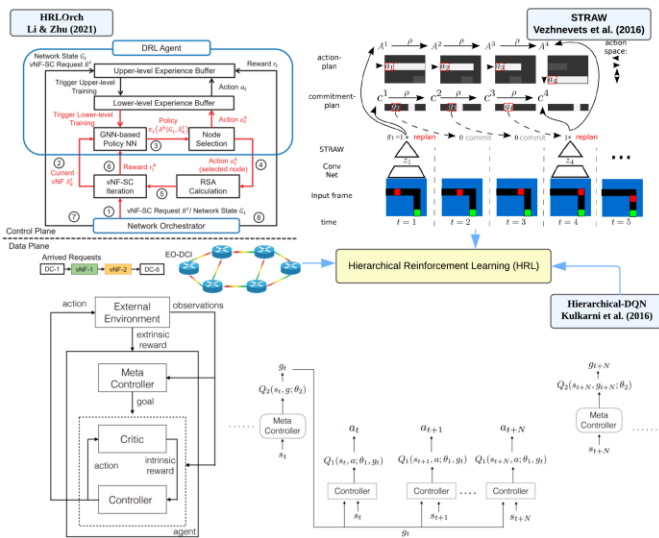


Fig. 6. Schematic illustration of h-DQN, STRAW and HRLOrch [83, 85, 89].

C. Multi-Agent Reinforcement Learning (MARL)

Multi-Agent Reinforcement Learning (MARL) is one of the specialized areas in Reinforcement Learning (RL) that focuses on the environment where multiple agents interact (not limited to number of environments), each aiming to optimize the performance while considering the presence and actions of other agents [18, 90, 91, 92]. Unlike single agent, each agent in MARL has its own goal, which may involve cooperation, competition or a mix of both, hence the environment is non-stationary from the perspective of each agent because other agents are also learning and changing their policies [93, 94].

One of the significant advancements in MARL is the Differentiable Inter-Agent Learning (DIAL) algorithm by Foerster et al. (2016), which uses differentiable communication channels to enable end-to-end training of communication policies, allowing agents to learn to communicate more effectively [95]. In addition, Lowe et al. (2017) introduced the Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm, which uses centralized training to gather global information while employing decentralized execution for deployment. Each agent has its own policy and critic, but the critics have access to the global state and actions of all other agents during training, making the training process more stable and effective [96].

For more recent advancements in MARL, Iqbal and Sha (2019) introduced the Actor-Attention-Critic (AAC) framework, which uses an attention mechanism to focus on relevant parts of the environment and other agents’ actions [97]. This framework enhances the scalability and performance of MARL algorithms in more complex environments. In addition, Rashid et al., 2020 proposed the QMIX algorithm, which decomposes the joint action-value function into a monotonic combination of individual value functions for agents, enabling efficient coordination in cooperative tasks [98]. Yu et al. (2022) introduced Multi-Agent Proximal Policy Optimization (MAPPO), an extension of the PPO algorithm for multi-agent settings. MAPPO employs centralized training with decentralized execution to address cooperative and complex tasks, achieving performance comparable to off-policy methods like MADDPG [99].

Furthermore, Carta et al. (2021) proposed an ensemble approach using multiple Deep Q-learning (Multi-DQN) agents to enhance stock market forecasting by training several agents on the same data and aggregating their decisions [100]. Zhang et al. (2022) introduced the Multi-Agent Graph Convolutional Reinforcement Learning (MAGC) framework, which employs graph neural networks to model the interactions between agents. MAGC enables agents to learn and coordinate their actions more effectively by capturing the relational structures of the environment, and it is applied to tasks such as dynamic electric vehicle charging pricing [36]. In summary, the development of MARL frameworks including DIAL, MADDPG, AAC, QMIX, MAPPO, Multi-DQN and MAGC further advanced the field of RL through the multi agent systems. Examples of schematic illustrations of Multi-DQN and MADDPG are shown in Fig. 7.

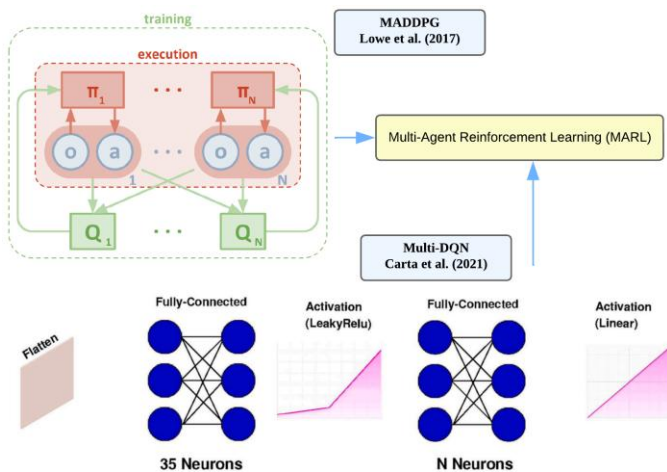


Fig. 7. Schematic illustration of multi-DQN and MADDPG [96, 100].

D. Model-Based Reinforcement Learning (MBRL)

Model-Based Reinforcement Learning (MBRL) is one of the approaches in Reinforcement Learning (RL) field that focuses on building models of environment to improve learning and decision-making effectiveness for gaming and robotic tasks [49, 101, 102]. The difference between model-based and model-free was discussed in previous section Model-Based and Model-Free. The core idea of MBRL is mainly decompose the objective into two main component which are model learning and planning, model learning includes environment's transition dynamics and reward function while planning uses the learned model to simulate future reward [102].

The Dyna framework that introduced by Sutton (1991) is one of the earliest works in MBRL, which integrates model learning and planning by combined real world interactions and simulated experiences generated by learned model, it allows the agent to update policy using real and synthetic data [103]. In addition, Delsensor & Rasmussen (2011) have proposed the Probabilistic Inference for Learning Control (PILCO) algorithm which is one type of model-based approach that uses Gaussian processes to model the environment's dynamics [104]. Nagabandi et al. (2018) introduced a model-based RL approach using neural network dynamics (MBRL-NN). This method employs deep neural networks to model the environment's transitions and combines them with model predictive control (MPC) for planning [105].

Furthermore, Ha and Schmidhuber (2018) proposed the World Models framework, which learns a compact, latent representation of the environment using a Variational Autoencoder (VAE) and a Recurrent Neural Network (RNN).

The agent plans and acts within this learned model, achieving good results on tasks in CarRacing-v0 and VizDoom [56]. The Probabilistic Ensembles with Trajectory Sampling (PETS) algorithm addresses model uncertainty in Model-based Reinforcement Learning (MBRL) by using an ensemble of probabilistic neural networks to model environment dynamics and trajectory sampling to account for uncertainty [58]. Moreover, Janner et al. (2019) introduced Model-Based Policy Optimization (MBPO), which integrates model-based and model-free approaches to improve sample efficiency. MBPO utilizes an ensemble of probabilistic neural networks to model the environment dynamics and conducts policy optimization within this learned model [106].

In recent years, Schrittwieser et al. (2020) proposed MuZero, a model-based RL algorithm that learns a model of the environment's dynamics and uses it for planning without requiring prior knowledge of the environment. MuZero achieves state-of-the-art performance in Atari, Go, chess, and shogi, demonstrating the benefits of combining model-based planning with model-free learning [107]. Similar with Deep Reinforcement Learning (DRL), the DreamerV2, MuZero and Plan2Explore algorithm also consider advancement for MBRL which expanded the application of MBRL to wider range of field. In summary, the introduction of Dyna, PILCO, MBRL-NN, World Models, PETS, MBPO and MuZero framework have advanced the field of RL and improved in terms of learning and decision-making. Schematic illustration of PETS and MuZero are shown in Fig. 8, which shows how the model plans and communicates with environments. The summary and comparative analysis of advanced Reinforcement Learning (RL) approaches is shown in Table IV.

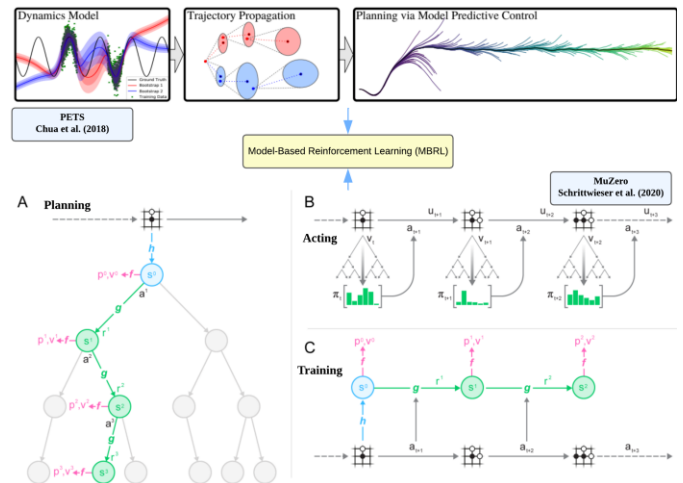


Fig. 8. Schematic illustration of PETS and MuZero [58, 107].

TABLE IV. SUMMARY AND COMPARATIVE ANALYSIS OF ADVANCED REINFORCEMENT LEARNING (RL) APPROACHES

Category	Framework	Key Contributions	Author & Year
Deep Reinforcement Learning (DRL)	Deep Q Network (DQN)	Approximates Q-value function using CNN, stabilizes training with experience replay and target networks.	Mnih et al., 2015
	Trust Region Policy Optimization (TRPO)	Ensures stable training via trust region constraints.	Schulman et al., 2015
	Proximal Policy Optimization (PPO)	Simplified policy gradient method with a clipped objective for efficiency.	Schulman et al., 2017

Category	Framework	Key Contributions	Author & Year
	Deep Deterministic Policy Gradient (DDPG)	Extends actor-critic framework to continuous action spaces.	Lillicrap et al., 2015
	Implicit Q-Learning (IQL)	Mitigates distributional shift errors in offline RL by using state-value functions as random variables.	Kostrikov et al., 2021
	DreamerV2	Combines latent variables with world models for efficient decision-making.	Chen et al., 2022
	Plan2Explore	Intrinsic exploration using a self-supervised world model.	Sekar et al., 2020
Hierarchical Reinforcement Learning (HRL)	Feudal RL (FRL)	Hierarchical decomposition of tasks via manager-worker relationships.	Dayan & Hinton, 1992
	Options Framework	Temporally extended actions with initiation, termination, and policies.	Sutton et al., 1999
	Hierarchical DQN (h-DQN)	Meta-controller for sub-goals and DQN-based controller.	Kulkarni et al., 2016
	Option-Critic Architecture	End-to-end learning of options and intra-option policies.	Bacon et al., 2017
	Strategic Attentive Writer (STRAW)	Plans macro actions via deep RNNs for efficient exploration and computation.	Vezhnevets et al., 2016
	Hierarchical Reinforcement learning with Off-policy correction (HIRO)	Off-policy correction for stable hierarchical learning.	Nachum et al., 2018
	Hierarchical Actor-Critic (HAC)	Actor-critic framework for multi-level task abstraction.	Levy et al., 2019
	Hierarchical Deep Reinforcement Learning with Automatic Sub-Goal Identification via Computer Vision (HADS)	Sub-goal generation via computer vision for dynamic task adaptation.	Liu et al., 2021
	Hierarchical Deep Reinforcement Learning with Graph Neural Networks (HRLOrch)	Graph neural networks for multi-level environment abstraction.	Li & Zhu, 2021
	Hierarchical Variational Autoencoder (HVAE)	Combines probabilistic generative models with deep neural networks to learn hierarchical topic representation	Bai et al., 2023
Multi-Agent Reinforcement Learning (MARL)	Differentiable Inter-Agent Learning (DIAL)	End-to-end learning of communication policies via differentiable channels.	Foerster et al., 2016
	Multi-Agent Deep Deterministic Policy Gradient (MADDPG)	Centralized training with decentralized execution for multi-agent setups.	Lowe et al., 2017
	Actor-Attention-Critic (AAC)	Attention mechanism for focusing on relevant environment and agent interactions.	Iqbal & Sha, 2019
	QMIX	Monotonic decomposition of joint action-value for cooperative tasks.	Rashid et al., 2020
	Multi-Agent Proximal Policy Optimization (MAPPO)	Extends PPO for multi-agent systems with centralized training and decentralized execution.	Yu et al., 2022
	Multiple Deep Q-learning (Multi-DQN)	Ensemble of DQN agents for aggregating decisions in forecasting tasks.	Carta et al., 2021
	Multi-Agent Graph Convolutional Reinforcement Learning (MAGC)	Graph neural networks for relational multi-agent modelling.	Zhang et al., 2022
Model-Based Reinforcement Learning (MBRL)	Dyna	Combines model learning and planning using real and synthetic data.	Sutton, 1991
	Probabilistic Inference for Learning Control (PILCO)	Uses Gaussian processes to model environment dynamics.	Delsenroth & Rasmussen, 2011
	Model-Based Reinforcement Learning using Neural Network Dynamics (MBRL-NN)	Combines neural network-based dynamics with model predictive control.	Nagabandi et al., 2018
	World Models	Latent environment representation via VAE and RNN.	Ha & Schmidhuber, 2018
	Probabilistic Ensembles with Trajectory Sampling (PETS)	Ensemble probabilistic models with trajectory sampling for uncertainty handling.	Chua et al., 2018
	Model-Based Policy Optimization (MBPO)	Combines model-based and model-free approaches for sample efficiency.	Janner et al., 2019
	MuZero	Learns environment models and plans without prior knowledge, achieving good results in gaming.	Schrittwieser et al., 2020

V. CHALLENGES AND ALTERNATIVE SOLUTIONS

Although Reinforcement Learning (RL) has its effectiveness in a wide range of applications, it still faces significant challenges that limit the efficiency, scalability, and

real-world applicability. The main key challenges include sample inefficiency, exploration-exploitation dilemma, and difficulties with generalization across different tasks and environments. Hence, there is much research that proposed new approaches to improve the adaptability and efficiency of

RL agents such as curiosity-driven exploration, meta-learning, and transfer learning.

One of the important challenges in RL is sample inefficiency as mentioned, where agents require large amount of interaction with the environment to learn effective policies. For example, in complex tasks such as involving high dimensional state space or continuous actions space, RL methods usually required more time to converge to optimal solutions. For the sample inefficiency challenges, several solutions have been proposed. For instance, the work of Janner et al. (2019) have proposed an RL algorithm Model-Based Policy Optimization (MBPO) which uses short model-generated rollouts to improve sample efficiency and performance as mentioned [106]. In addition, Soft Actor-Critic (SAC) which an off-policy actor-critic algorithm based on maximum entropy RL can be used to maximize both expected reward and entropy, it also able to be enabling agents to learn from data collected under different policies [108]. Not only that, Deep Q-Networks (DQN) also can be used to store and reuse past experiences or learning, this can reduce the need for time required and samples in each iteration [13].

In addition, one of another challenges in RL are exploration-exploitation dilemma, in which an agent must balance between exploring new environment then found the highest beneficial actions and exploiting known actions that produce high rewards. Poor exploration strategies can lead to suboptimal strategies, especially in environments where reward signals are delayed or require a long time such as in physical design. Therefore, several solutions have been proposed for solving this issue. One of the solutions is curiosity-driven exploration which to explores using curiosity as an intrinsic reward signal for agents in environments with sparse or no extrinsic rewards which curiosity is defined as the error in predicting the consequences of the agent's actions in a visual feature space [109]. In addition, curiosity-driven exploration is important for autonomous learning, highlighting various algorithmic models that capture different aspects of this process [110]. Moreover, entropy regularization also can be used to address the exploration-exploitation dilemma by introducing f-divergence penalties [111]. These penalties

ensure that the policy does not deviate too much from the current policy, promoting balanced exploration and exploitation. By adjusting the divergence function, the agent can control the trade-off between exploring new actions and exploiting known rewarding actions, this can lead to more stable and efficient learning dynamics.

Furthermore, another challenge is difficulties with generalization, where generalization in RL refers to the ability of an agent to still perform well in new or unseen environments [112, 113]. This is due to RL models often overfitting to specific environments during training, leading to a decline in performance when faced with new or unseen scenarios. This issue is a serious problem for real-world applications such as autonomous driving, where the agent is required to handle various scenarios under different conditions [115]. In order to solve this problem, the model agnostic meta learning (MAML) approach can be applied to enable agents more quickly adapt to new tasks by learning from distribution of related tasks during the training phase [115, 116, 117]. This approach can make the agent more robust and adaptable in unfamiliar environments. In addition, transfer learning also can be applied to transfer the knowledge to another related task to reducing the training data needed in new environment [118, 119] This will be more useful when the training in the target environment is costly such as required more memory space. Moreover, domain randomization also can help to improve the generalization issue by training the agents in varying the parameters, so that the agents can be handle the real-world variability more effectively and more adaptable to new, unseen environments [120].

In summary, the purpose of artificial intelligence (AI) including RL is not to replace humans, AI is designed to enhance human efficiency and achieve better outcomes. Humans are required to treat it as a tool to increase efficiency, streamline workflows, and assist in decision-making processes. However, it is also very important to ensure that AI technologies are applied properly to maximize the potential benefits and minimize the risk of misuse or unintended consequences. The summary of alternative methods and their potential benefits is shown in Table V.

TABLE V. SUMMARY OF ALTERNATIVE METHODS AND POTENTIAL BENEFITS

Alternative Method	Challenges	Key Feature	Potential Benefits
Model-Based Policy Optimization (MBPO)	Sample Efficiency	Use short model-generated rollouts	Increases sample efficiency
Soft Actor-Critic (SAC)	Sample Efficiency	Agents can learn from data collected under different policies	Maximize both expected reward and entropy and purpose to increase sample efficiency
Curiosity-Driven Exploration	Exploration-Exploitation Dilemma	Use intrinsic rewards based on novelty	Enhances the exploration in sparse reward environments
Entropy Regularization	Exploration-Exploitation Dilemma	Maintains randomness in policy by penalizing determinism	Promotes continued exploration during training
Meta-Learning	Generalization	Learns to adapt quickly to new tasks from experiences	Improves adaptability and efficiency across tasks
Transfer Learning	Generalization	Transfers knowledge from one task to another task	Reduces training time and data requirements
Domain Randomization	Generalization	Trains in a different of simulated environments	Improves robustness to different of environments

VI. CONCLUSION AND OUTLOOK

In conclusion, the evolutions of Reinforcement Learning has successfully impacted the different fields from robotics and autonomous driving system, healthcare and finance. The integration of RL with different approach can further advanced the application, for example the integration of deep learning and RL which known as Deep Reinforcement Learning (DRL) has improved in solving the high dimensional problem and applied in AlphaGo.

In addition, the focus of RL is mainly improving the sample efficiency, safety and scalability for real world applications. The innovations in hierarchical Reinforcement Learning, transfer learning and domain randomization are expected to improve the adaptability and generalizability of RL systems. In summary, as increasing more effort in improving RL, it will play an important role in artificial intelligence and autonomous technologies which will help humans for more complex challenges in daily life.

ACKNOWLEDGMENT

This research was funded by the Ministry of Higher Education (MoHE) Malaysia with the Fundamental Research Grant Scheme (FRGS) under grant number FRGS/1/2021/TK0/UKM/02/17, and by the National Research and Innovation Agency (BRIN) Indonesia under the Perjanjian Pendanaan Program Riset dan Inovasi untuk Indonesia Maju Gelombang 4, with grant numbers 20/IV/KS/11/2023 and 1181/PL4.7.4.2/PT/2023.

REFERENCES

- [1] Naeem, M., Rizvi, S. T. H., & Coronato, A. (2020). A gentle introduction to reinforcement learning and its application in different fields. *IEEE access*, 8, 209320-209344.
- [2] Peng, J., Jury, E. C., Dönnies, P., & Ciurtin, C. (2021). Machine learning techniques for personalised medicine approaches in immune-mediated chronic inflammatory diseases: applications and challenges. *Frontiers in pharmacology*, 12, 720694.
- [3] Nian, R., Liu, J., & Huang, B. (2020). A review on reinforcement learning: Introduction and applications in industrial process control. *Computers & Chemical Engineering*, 139, 106886.
- [4] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature*, 518, 529-533.
- [5] AlMahamid, F., & Grolinger, K. (2021). A1:A38 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE) (pp. 1-7). IEEE.
- [6] Bogert, K., Lin, J. F. S., Doshi, P., & Kulic, D. (2016). Expectation-maximization for inverse reinforcement learning with hidden data. In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems* (pp. 1034-1042).
- [7] Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4, 237-285.
- [8] Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 32(11), 1238-1274.
- [9] Cheng, Y., Chen, C., Hu, X., Chen, K., Tang, Q., & Song, Y. (2021). Enhancing mixed traffic flow safety via connected and autonomous vehicle trajectory planning with a reinforcement learning approach. *Journal of Advanced Transportation*, 2021(1), 6117890.
- [10] Haydari, A., & Yılmaz, Y. (2020). Deep reinforcement learning for intelligent transportation systems: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(1), 11-32.
- [11] Hofmann, S., Walter, M., Servadei, L., & Wille, R. (2024). Thinking Outside the Clock: Physical Design for Field-coupled Nanocomputing with Deep Reinforcement Learning. In *2024 25th International Symposium on Quality Electronic Design (ISQED)* (pp. 1-8). IEEE.
- [12] Lu, Y. C., Chan, W. T., Guo, D., Kundu, S., Khandelwal, V., & Lim, S. K. (2023). RL-CCD: Concurrent clock and data optimization using attention-based self-supervised reinforcement learning. In *2023 60th ACM/IEEE Design Automation Conference (DAC)* (pp. 1-6). IEEE.
- [13] Li, S. E. (2023). Deep reinforcement learning. In *Reinforcement learning for sequential decision and optimal control* (pp. 365-402). Singapore: Springer Nature Singapore.
- [14] Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 34(6), 26-38.
- [15] François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., & Pineau, J. (2018). An introduction to deep reinforcement learning. *Foundations and Trends® in Machine Learning*, 11(3-4), 219-354.
- [16] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484-489.
- [17] Sallab, A. E., Abdou, M., Perot, E., & Yogamani, S. (2017). Deep reinforcement learning framework for autonomous driving. *arXiv preprint arXiv:1704.02532*.
- [18] Zhang, K., Yang, Z., & Başar, T. (2021). Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of reinforcement learning and control*, 321-384.
- [19] Ning, Z., & Xie, L. (2024). A survey on multi-agent reinforcement learning and its application. *Journal of Automation and Intelligence*.
- [20] Thorndike, E. L. (1933). A proof of the law of effect. *Science*, 77(1989), 173-175.
- [21] Skinner, B. F. (1950). Are theories of learning necessary?. *Psychological review*, 57(4), 193.
- [22] Skinner, B. F. (1958). Reinforcement today. *American Psychologist*, 13(3), 94.
- [23] Islam, M. H. (2015). Thorndike theory and it's application in learning. *At-Ta'lim: Jurnal Pendidikan*, 1(1), 37-47.
- [24] Skinner, B. F. (1963). Operant behavior. *American psychologist*, 18(8), 503.
- [25] Ab Azar, N., Shahmansoorian, A., & Davoudi, M. (2020). From inverse optimal control to inverse reinforcement learning: A historical review. *Annual Reviews in Control*, 50, 119-138.
- [26] Bellman, R. (1957). *Dynamic programming* princeton university press. Princeton, NJ, 4-9.
- [27] Bellman, R., & Kalaba, R. (1957). Dynamic programming and statistical communication theory. *Proceedings of the National Academy of Sciences*, 43(8), 749-751.
- [28] Bellman, R., & Lee, E. S. (1978). Functional equations in dynamic programming. *Aequationes mathematicae*, 17(1), 1-18.
- [29] Ding, Z., Huang, Y., Yuan, H., & Dong, H. (2020). Introduction to reinforcement learning. *Deep reinforcement learning: fundamentals, research and applications*, 47-123.
- [30] Howard, R. A. (1960). *Dynamic Programming and Markov Processes*. MIT Press google schola, 2, 39-47.
- [31] Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine learning*, 3, 9-44.
- [32] Watkins, C. J. C. H. (1989). Learning from delayed rewards.
- [33] Granter, S. R., Beck, A. H., & Papke Jr, D. J. (2017). AlphaGo, deep learning, and the future of the human microscopist. *Archives of pathology & laboratory medicine*, 141(5), 619-621.
- [34] Dulac-Arnold, G., Levine, N., Mankowitz, D. J., Li, J., Paduraru, C., Goyal, S., & Hester, T. (2021). Challenges of real-world reinforcement

- learning: definitions, benchmarks and analysis. *Machine Learning*, 110(9), 2419-2468.
- [35] Chen, X., Qu, G., Tang, Y., Low, S., & Li, N. (2022). Reinforcement learning for selective key applications in power systems: Recent advances and future challenges. *IEEE Transactions on Smart Grid*, 13(4), 2935-2958.
- [36] Zhang, W., Liu, H., Han, J., Ge, Y., & Xiong, H. (2022). Multi-agent graph convolutional reinforcement learning for dynamic electric vehicle charging pricing. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining* (pp. 2471-2481).
- [37] Pateria, S., Subagdja, B., Tan, A. H., & Quek, C. (2021). Hierarchical reinforcement learning: A comprehensive survey. *ACM Computing Surveys (CSUR)*, 54(5), 1-35.
- [38] Zhu, Z., Lin, K., Jain, A. K., & Zhou, J. (2023). Transfer learning in deep reinforcement learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- [39] Huang, Q. (2020). Model-based or model-free, a review of approaches in reinforcement learning. In *2020 International Conference on Computing and Data Science (CDS)* (pp. 219-221). IEEE
- [40] Gao, C., & Wang, D. (2023). Comparative study of model-based and model-free reinforcement learning control performance in HVAC systems. *Journal of Building Engineering*, 74, 106852.
- [41] Alrebd, N., Alrumiah, S., Almansour, A., & Rassam, M. (2022). Reinforcement Learning in Image Classification: A Review. In *2022 2nd International Conference on Computing and Information Technology (ICCIIT)* (pp. 79-86). IEEE.
- [42] Sutton, R. S., Barto, A. G., & Williams, R. J. (1992). Reinforcement learning is direct adaptive optimal control. *IEEE control systems magazine*, 12(2), 19-22.
- [43] Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine learning*, 8, 279-292.
- [44] Fan, J., Wang, Z., Xie, Y., & Yang, Z. (2020). A theoretical analysis of deep Q-learning. In *Learning for dynamics and control* (pp. 486-489). PMLR.
- [45] Atkeson, C. G., & Santamaria, J. C. (1997). A comparison of direct and model-based reinforcement learning. In *Proceedings of international conference on robotics and automation* (Vol. 4, pp. 3557-3564). IEEE.
- [46] Ni, T., Eysenbach, B., & Salakhutdinov, R. (2021). Recurrent model-free rl can be a strong baseline for many pomdps. *arXiv preprint arXiv:2110.05038*.
- [47] Dayan, P., & Berridge, K. C. (2014). Model-based and model-free Pavlovian reward learning: revaluation, revision, and revelation. *Cognitive, Affective, & Behavioral Neuroscience*, 14, 473-492.
- [48] Zhang, H., & Yu, T. (2020). Taxonomy of reinforcement learning algorithms. *Deep Reinforcement Learning: Fundamentals, Research and Applications*, 125-133.
- [49] Polydoros, A. S., & Nalpanitidis, L. (2017). Survey of model-based reinforcement learning: Applications on robotics. *Journal of Intelligent & Robotic Systems*, 86(2), 153-173.
- [50] Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
- [51] Plaat, A., Kusters, W., & Preuss, M. (2023). High-accuracy model-based reinforcement learning, a survey. *Artificial Intelligence Review*, 56(9), 9541-9573.
- [52] Lambert, N., Amos, B., Yadan, O., & Calandra, R. (2020). Objective mismatch in model-based reinforcement learning. *arXiv preprint arXiv:2002.04523*.
- [53] Rajeswaran, A., Mordatch, I., & Kumar, V. (2020). A game theoretic framework for model based reinforcement learning. In *International conference on machine learning* (pp. 7953-7963). PMLR.
- [54] Plaat, A., Kusters, W., & Preuss, M. (2020). Deep model-based reinforcement learning for high-dimensional problems, a survey. *arXiv preprint arXiv:2008.05598*.
- [55] Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P. I., Rohlfshagen, P., Tavener, S., Perez, D., Samothrakis, S., & Colton, S. (2012). A Survey of Monte Carlo Tree Search Methods. *IEEE Transactions on Computational Intelligence and AI in Games*, 4(1), 1-43
- [56] Ha, D., & Schmidhuber, J. (2018). World models. *arXiv preprint arXiv:1803.10122*.
- [57] Hafner, D., Lillicrap, T., Norouzi, M., & Ba, J. (2020). Mastering atari with discrete world models. *arXiv preprint arXiv:2010.02193*.
- [58] Chua, K., Calandra, R., McAllister, R., & Levine, S. (2018). Deep reinforcement learning in a handful of trials using probabilistic dynamics models. *Advances in neural information processing systems*, 31.
- [59] Sutton, R. S., Mahmood, A. R., & White, M. (2016). An emphatic approach to the problem of off-policy temporal-difference learning. *The Journal of Machine Learning Research*, 17(1), 2603-2631.
- [60] Singh, S., Jaakkola, T., Littman, M. L., & Szepesvári, C. (2000). Convergence results for single-step on-policy reinforcement-learning algorithms. *Machine learning*, 38, 287-308.
- [61] Maei, H. R., Szepesvári, C., Bhatnagar, S., & Sutton, R. S. (2010). Toward off-policy learning control with function approximation. In *ICML* (Vol. 10, pp. 719-726).
- [62] Fakoor, R., Chaudhari, P., & Smola, A. J. (2020). P3o: Policy-on policy-off policy optimization. In *Uncertainty in Artificial Intelligence* (pp. 1017-1027). PMLR.
- [63] Sewak, M. (2019). Policy-based reinforcement learning approaches: Stochastic policy gradient and the REINFORCE algorithm. *Deep Reinforcement Learning: Frontiers of Artificial Intelligence*, 127-140.
- [64] Li, X., Lv, Z., Wang, S., Wei, Z., & Wu, L. (2019). A reinforcement learning model based on temporal difference algorithm. *IEEE Access*, 7, 121922-121930.
- [65] Rammohan, S., Yu, S., He, B., Hsiung, E., Rosen, E., Tellex, S., & Konidaris, G. (2021). Value-Based Reinforcement Learning for Continuous Control Robotic Manipulation in Multi-Task Sparse Reward Settings. *arXiv preprint arXiv:2107.13356*.
- [66] Singh, S. P., & Sutton, R. S. (1996). Reinforcement learning with replacing eligibility traces. *Machine learning*, 22, 123-158.
- [67] Peters, J. (2010). Policy gradient methods. *Scholarpedia*, 5(11), 3698.
- [68] Li, Y. (2017). *Deep Reinforcement Learning: An Overview*. *arXiv preprint arXiv:1701.07274*.
- [69] Mousavi, S. S., Schukat, M., & Howley, E. (2018). Deep reinforcement learning: an overview. In *Proceedings of SAI Intelligent Systems Conference (IntelliSys) 2016: Volume 2* (pp. 426-440). Springer International Publishing.
- [70] Nguyen, H., & La, H. (2019). Review of deep reinforcement learning for robot manipulation. In *2019 Third IEEE international conference on robotic computing (IRC)* (pp. 590-595). IEEE.
- [71] Kiran, B. R., Sobh, I., Talpaert, V., Mannion, P., Al Sallab, A. A., Yogamani, S., & Pérez, P. (2021). Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(6), 4909-4926.
- [72] Schulman, J., Levine, S., Abbeel, P., Jordan, M., & Moritz, P. (2015). Trust Region Policy Optimization. *arXiv preprint arXiv:1502.05477*.
- [73] Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.
- [74] Lillicrap, T. P. (2015). Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.
- [75] Kostrikov, I., Nair, A., & Levine, S. (2021). Offline reinforcement learning with implicit q-learning. *arXiv preprint arXiv:2110.06169*.
- [76] Chen, C., Wu, Y. F., Yoon, J., & Ahn, S. (2022). Transdreamer: Reinforcement learning with transformer world models. *arXiv preprint arXiv:2202.09481*.
- [77] Sekar, R., Rybkin, O., Daniilidis, K., Abbeel, P., Hafner, D., & Pathak, D. (2020). Planning to explore via self-supervised world models. In *International conference on machine learning* (pp. 8583-8592). PMLR.
- [78] Barto, A. G., & Mahadevan, S. (2003). Recent advances in hierarchical reinforcement learning. *Discrete event dynamic systems*, 13, 341-379.
- [79] Nachum, O., Gu, S. S., Lee, H., & Levine, S. (2018). Data-efficient hierarchical reinforcement learning. *Advances in neural information processing systems*, 31.

- [80] Al-Emran, M. (2015). Hierarchical reinforcement learning: a survey. *International journal of computing and digital systems*, 4(02).
- [81] Dayan, P., & Hinton, G. E. (1992). Feudal reinforcement learning. *Advances in neural information processing systems*, 5.
- [82] Sutton, R. S., Precup, D., & Singh, S. (1999). Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2), 181-211.
- [83] Kulkarni, T. D., Narasimhan, K., Saeedi, A., & Tenenbaum, J. (2016). Hierarchical deep reinforcement learning: Integrating temporal abstraction and intrinsic motivation. *Advances in neural information processing systems*, 29.
- [84] Bacon, P. L., Harb, J., & Precup, D. (2017). The option-critic architecture. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 31, No. 1).
- [85] Vezhnevets, A. S., Osindero, S., Schaul, T., Heess, N., Jaderberg, M., Silver, D., & Kavukcuoglu, K. (2017). Feudal networks for hierarchical reinforcement learning. In *International conference on machine learning* (pp. 3540-3549). PMLR.
- [86] Levy, A., Konidaris, G., Platt, R., & Saenko, K. (2017). Learning multi-level hierarchies with hindsight. *arXiv preprint arXiv:1712.00948*.
- [87] Bai, R., Huang, R., Qin, Y., Chen, Y., & Lin, C. (2023). HVAE: A deep generative model via hierarchical variational auto-encoder for multi-view document modeling. *Information Sciences*, 623, 40-55.
- [88] Liu, C., Zhu, F., Liu, Q., & Fu, Y. (2021). Hierarchical reinforcement learning with automatic sub-goal identification. *IEEE/CAA journal of automatica sinica*, 8(10), 1686-1696.
- [89] Li, B., & Zhu, Z. (2022). GNN-based hierarchical deep reinforcement learning for NFV-oriented online resource orchestration in elastic optical DCIs. *Journal of Lightwave Technology*, 40(4), 935-946.
- [90] Tan, M. (1993). Multi-agent reinforcement learning: Independent vs. cooperative agents. In *Proceedings of the tenth international conference on machine learning* (pp. 330-337).
- [91] Buşoniu, L., Babuška, R., & De Schutter, B. (2010). Multi-agent reinforcement learning: An overview. *Innovations in multi-agent systems and applications-1*, 183-221.
- [92] Canese, L., Cardarilli, G. C., Di Nunzio, L., Fazzolari, R., Giardino, D., Re, M., & Spanò, S. (2021). Multi-agent reinforcement learning: A review of challenges and applications. *Applied Sciences*, 11(11), 4948.
- [93] Papoudakis, G., Christianos, F., Rahman, A., & Albrecht, S. V. (2019). Dealing with non-stationarity in multi-agent deep reinforcement learning. *arXiv preprint arXiv:1906.04737*.
- [94] Hernandez-Leal, P., Kaisers, M., Baarslag, T., & De Cote, E. M. (2017). A survey of learning in multiagent environments: Dealing with non-stationarity. *arXiv preprint arXiv:1707.09183*.
- [95] Foerster, J., Assael, I. A., De Freitas, N., & Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning. *Advances in neural information processing systems*, 29.
- [96] Lowe, R., Wu, Y. I., Tamar, A., Harb, J., Pieter Abbeel, O., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems*, 30.
- [97] Iqbal, S., & Sha, F. (2019). Actor-attention-critic for multi-agent reinforcement learning. In *International conference on machine learning* (pp. 2961-2970). PMLR.
- [98] Rashid, T., Samvelyan, M., De Witt, C. S., Farquhar, G., Foerster, J., & Whiteson, S. (2020). Monotonic value function factorisation for deep multi-agent reinforcement learning. *Journal of Machine Learning Research*, 21(178), 1-51.
- [99] Yu, C., Velu, A., Vinitzky, E., Gao, J., Wang, Y., Bayen, A., & Wu, Y. (2022). The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems*, 35, 24611-24624.
- [100] Carta, S., Ferreira, A., Podda, A. S., Recupero, D. R., & Sanna, A. (2021). Multi-DQN: An ensemble of Deep Q-learning agents for stock market forecasting. *Expert systems with applications*, 164, 113820.
- [101] Kaiser, L., Babaeizadeh, M., Milos, P., Osinski, B., Campbell, R. H., Czechowski, K., Erhan, D., Finn, C., Kozakowski, P., Levine, S., Mohiuddin, A., Sepassi, R., Tucker, G., & Michalewski, H. (2019). Model-based reinforcement learning for atari. *arXiv preprint arXiv:1903.00374*.
- [102] Moerland, T. M., Broekens, J., Plaat, A., & Jonker, C. M. (2023). Model-based reinforcement learning: A survey. *Foundations and Trends® in Machine Learning*, 16(1), 1-118.
- [103] Sutton, R. S. (1991). Dyna, an integrated architecture for learning, planning, and reacting. *ACM Sigart Bulletin*, 2(4), 160-163.
- [104] Deisenroth, M., & Rasmussen, C. E. (2011). PILCO: A model-based and data-efficient approach to policy search. In *Proceedings of the 28th International Conference on machine learning (ICML-11)* (pp. 465-472).
- [105] Nagabandi, A., Kahn, G., Fearing, R. S., & Levine, S. (2018). Neural network dynamics for model-based deep reinforcement learning with model-free fine-tuning. In *2018 IEEE international conference on robotics and automation (ICRA)* (pp. 7559-7566). IEEE.
- [106] Janner, M., Fu, J., Zhang, M., & Levine, S. (2019). When to trust your model: Model-based policy optimization. *Advances in neural information processing systems*, 32.
- [107] Schrittwieser, J., Antonoglou, I., Hubert, T., Simonyan, K., Sifre, L., Schmitt, S., Guez, A., Lockhart, E., Hassabis, D., Graepel, T., Lillicrap, T., & Silver, D. (2020). Mastering atari, go, chess and shogi by planning with a learned model. *Nature*, 588(7839), 604-609.
- [108] Haamoja, T., Zhou, A., Abbeel, P., & Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning* (pp. 1861-1870). PMLR.
- [109] Pathak, D., Agrawal, P., Efros, A. A., & Darrell, T. (2017). Curiosity-driven exploration by self-supervised prediction. In *International conference on machine learning* (pp. 2778-2787). PMLR.
- [110] Ten, A., Oudeyer, P. Y., & Moulin-Frier, C. (2022). Curiosity-driven exploration. *The Drive for Knowledge: The Science of Human Information Seeking*, 53.
- [111] Belousov, B., & Peters, J. (2019). Entropic regularization of markov decision processes. *Entropy*, 21(7), 674.
- [112] Wang, K., Kang, B., Shao, J., & Feng, J. (2020). Improving generalization in reinforcement learning with mixture regularization. *Advances in Neural Information Processing Systems*, 33, 7968-7978.
- [113] Di Langosco, L. L., Koch, J., Sharkey, L. D., Pfau, J., & Krueger, D. (2022). Goal misgeneralization in deep reinforcement learning. In *International Conference on Machine Learning* (pp. 12004-12019). PMLR.
- [114] Coelho, D., Oliveira, M., & Santos, V. (2023). RLAD: Reinforcement Learning From Pixels for Autonomous Driving in Urban Environments. *IEEE Transactions on Automation Science and Engineering*.
- [115] Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning* (pp. 1126-1135). PMLR.
- [116] Gurumurthy, S., Kumar, S., & Sycara, K. (2020). Mame: Model-agnostic meta-exploration. In *Conference on Robot Learning* (pp. 910-922). PMLR.
- [117] Baik, S., Hong, S., & Lee, K. M. (2020). Learning to forget for meta-learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 2379-2387).
- [118] Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345-1359.
- [119] Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2020). A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1), 43-76.
- [120] Slaoui, R. B., Clements, W. R., Foerster, J. N., & Toth, S. (2019). Robust domain randomization for reinforcement learning.

Towards Transparent Traffic Solutions: Reinforcement Learning and Explainable AI for Traffic Congestion

Shan Khan¹, Taher M. Ghazal^{2,*}, Tahir Alyas³, M. Waqas⁴, Muhammad Ahsan Raza⁵, Oualid Ali⁶,
Muhammad Adnan Khan^{7,*}, Sagheer Abbas⁸

Department of CS, National College of Business Administration & Economics in Lahore, Pakistan^{1,4}

Hourani Center for Applied Scientific Research, Al-Ahliyya Amman University, Amman, Jordan²

Department of Computer Science, Lahore Garrison University, Lahore, Pakistan³

Department of Information Sciences, University of Education, Lahore, Multan Campus 60000, Pakistan⁵

College of Arts & Science, Applied Science University, P.O.Box 5055, Manama, Kingdom of Bahrain⁶

Department of Software-Faculty of Artificial Intelligence and Software, Gachon University, Republic of Korea⁷

Department of Computer Science, Prince Mohammad Bin Fahd University, Alkhobar, KSA⁸

Abstract—This study introduces a novel approach to traffic congestion detection using Reinforcement Learning (RL) of machine learning classifiers enhanced by Explainable Artificial Intelligence (XAI) techniques in Smart City (SC). Conventional traffic management systems rely on static rules, and heuristics face challenges in dynamically addressing urban traffic problems' complexities. This study explains the novel Reinforcement Learning (RL) framework integrated with an Explainable Artificial Intelligence (XAI) approach to deliver more transparent results. The model significantly reduces the missing data rate and improves overall prediction accuracy by incorporating RL for real-time adaptability and XAI for clarity. The proposed method enhances security, privacy, and prediction accuracy for traffic congestion detection by using Machine Learning (ML). Using RL for adaptive learning and XAI for interpretability, the proposed model achieves improved prediction and reduces the missing data rate, with an accuracy of 98.10, which is better than the existing methods.

Keywords—Reinforcement learning; Explainable Artificial Intelligence (XAI); Smart City (SC); IoT; Machine Learning (ML)

I. INTRODUCTION

Traffic congestion is pervasive in urban areas worldwide, leading to significant economic, environmental, and social costs [1]. Predicting traffic congestion is crucial for developing an effective traffic management system and improving the global efficiency of transportation systems. Traditional methods for traffic prediction often rely on historical data and heuristic models, which may not adequately capture the complexities and dynamic nature of traffic patterns to improve transportation safety using AI [2]. Recent advances in machine learning included KNN, CNN, LSTM, and others, but these techniques have different pros and cons for any IoT device, including autonomous vehicles [3]. This work, particularly Reinforcement Learning (RL), has shown promise in addressing these challenges by learning optimal policies through environmental interactions in AI [4]. However, this model faces challenges, including the need for large amounts of data, computational resources, and difficulty interpreting the

learned policies. Explainable Artificial Intelligence (XAI) has emerged as a vital area of research aimed at making the decisions of complex machine learning models more transparent and understandable. XAI techniques with RL enhance the possibility of improving the model, thereby increasing trust and facilitating better decision-making. Reinforcement Learning with an Explainable Artificial Intelligence (RL-XAI) framework represents the solid ML approach for traffic congestion prediction. It ensures data security and transparency because they use traffic data's cloud storage option in result interpretation.

Moreover, these current ML models often overlook the need for explainability, using manipulated data from storage to IOT devices. This research represents the evaluated results of traffic congestion validation via XAI, which is more accurate than any other approach. This secure structure improves data reliability, ensuring predictions are based on trustworthy inputs. Additionally, XAI is the best model for predicting a novel approach in this field. Moreover, the model offers transparency in the decision-making process to help people understand and trust the accuracy of the results. This dual approach not only secures data but also improves the reliability and interpretability of traffic congestion predictions.

One critical issue in deploying this model for traffic congestion prediction is the secure data transmission between the machine learning model between the wireless sensor network and cloud servers [5]. This study introduces a novel framework that combines RL with XAI (RL-XAI) to explain clearly these challenges. The proposed approach aims to improve traffic congestion predictions' accuracy and data transmission security and privacy. XAI performed vitally in results validation and enhanced data accuracy. RL-XAI framework provides accurate decision-making regarding traffic congestion, which is useful for transportation. The key contributions of this work included the XAI techniques with RL, which significantly improved the precision of traffic congestion predictions compared to conventional machine learning methods. The proposed framework confirms that

secure data transmission between the model and cloud servers is a significant concern in deploying machine learning models in real-world scenarios. Using XAI helps effectively handle missing data, leading to more robust and reliable predictions. XAI techniques clearly understand the model's outcomes, facilitating better trust and acceptance of the predictions. Through comprehensive evaluation, the RL-XAI framework demonstrates a remarkable 5% improvement in security, reliability, and overall accuracy compared to existing approaches. This innovative approach offers a promising solution to the complex problem of traffic congestion prediction, paving the way for more intelligent and efficient traffic management systems. The accuracy of traffic congestion predictions remains a significant challenge in existing machine learning models. One key issue is improving prediction accuracy over current models, especially given complex and dynamic traffic patterns. Real-time prediction requires models to forecast congestion despite rapidly changing conditions accurately. Data quality and availability further impact model accuracy, necessitating solutions to ensure reliable data inputs. Ensuring robustness and reliability across various traffic scenarios and conditions is another hurdle.

Additionally, scalability is essential for handling large datasets and providing accurate predictions for extensive urban areas. Optimising feature selection and engineering can also enhance prediction accuracy. Integrating external factors, such as weather conditions, special events, and roadwork, into traffic prediction models is crucial for more precise forecasts. Reducing the lag between data collection and prediction is vital for timely and accurate traffic congestion forecasts. Enhancing model interpretability ensures that stakeholders trust and understand accurate predictions.

Furthermore, models must quickly adapt to new traffic patterns resulting from changes in infrastructure, traffic laws, or unexpected events. Finally, identifying and mitigating prediction errors is necessary to improve overall model accuracy. Addressing these challenges is essential for developing more reliable and accurate traffic congestion prediction models. Furthermore, integrating XAI techniques enhances the model's interpretability, making its decision-making process transparent and understandable, thereby increasing user trust and acceptance. Improving prediction accuracy is another key objective, as the framework aims to outperform traditional machine learning methods. Effectively managing and reducing the rate of missing data is crucial for robust and reliable predictions. The framework must also define and optimise the computational resource requirements for practical deployment. Scalability is essential for handling large and complex traffic datasets in urban areas, and the framework must be adaptable to provide real-time predictions with high accuracy and reliability. Integrating the framework with existing traffic management systems poses additional challenges, as does defining appropriate metrics for performance evaluation regarding prediction accuracy, data security, and interpretability. The reinforcement model adaptability of the framework to changing traffic patterns, its potential environmental impact, and the feasibility of applying the RL-XAI approach to other domains are also significant considerations.

A. Reinforcement Learning (RL)

Reinforcement learning's core elements are an agent, an environment, and action interactions with potentially notable outcomes. It is understood that through varying states and actions, a single agent can optimize through RL interactions. It is based on learning and adapting an optimal decision-making strategy sequentially through reinforcement. Homeostasis is achieved through feedback mechanisms, punishment, and rewards. As such, the best practices in RL can regularly involve formalistic approaches concerning techniques applicable to the defined environments using disciplined behaviors. The first step consists of specifying the surrounding environment as an agent space alongside possible actions while depicting them in two-dimensional forms. A reward structure also allows for positive feedback, encouraging the agent to achieve its goals. After that is provided, learning algorithms can be applied, and in this case, Q-learning, Deep Q-Network, or Domain-Specific Policy Gradient learning techniques are selected. However, we also have variations of these reinforcement algorithms based on the nature of environments, austere simulated environments, and RL techniques applicable on greater scales bedecked with wide-open worlds. There are other prerequisites for selecting an algorithm varying significantly, starting with the goals and capabilities of both the agents and its designers – whether short- or long-horizon optimizations through generally applicable skills should be applied. Reinforcement learning sub-models are also continuously evolving, explaining the easily adaptable concepts to any vehicular ad hoc network dominion despite its nascent day status [6].

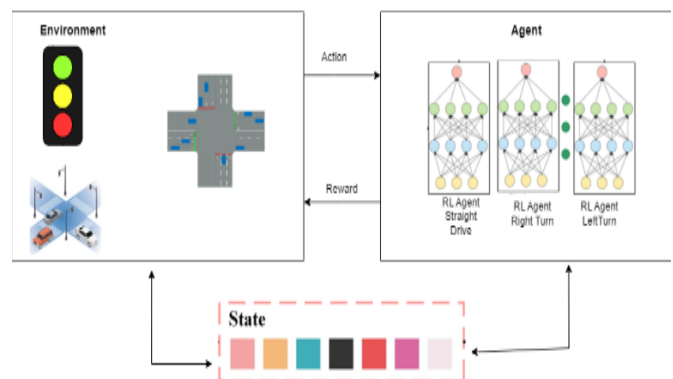


Fig. 1. The RL model for the traffic congestion system is based on agents and rewards.

Fig. 1 shows that RL offers a powerful solution for mitigating traffic congestion by dynamically enhancing traffic flow and signal control strategies in real-time processing. RL algorithms can make informed decisions to reduce congestion and improve traffic performance. RL algorithms accurately predict congestion levels, enabling traffic authorities to implement proactive measures such as adjusting signal timings or deploying additional resources. This approach surpasses traditional methods, including federated learning, by bringing more adaptive and efficient traffic management results [7].

B. Explainable Artificial Intelligence (XAI)

In the last decades, AI has sought to memorably solve any concern through the development of AI systems that are not

only interpretable but also understandable. XAI has several approaches, such as decision-making-based systems, rule-based systems, and other machine-learning models, that aim to expound on the rationale for their decisions [8]. Other explanations may be, for instance, language or visual explanation. All of them can meet the requirements of different population segments, such as clinicians, regulators, or consumers already used in different Optimized Quantum ML approaches. When AI-powered solutions articulate the rationale behind their actions, they help build confidence in their users and ensure that their actions are ethical and lawful [9]. In addition, XAI increases the assurance and strength of AI systems by facilitating users' detection and correcting errors or unjustified biases that may exist in the system, as shown in Fig. 3.

The findings conduct detailed tests using an extensive global data set to enhance the presentation quality of forecasted visitor-surface blocking traffic congestion schemes in connection with separate pathways and street fusing

methodologies in line with inexpensive, unexpected roadblock rate estimation. This marks the first instance where Reinforcement Learning has been integrated into another model, like RNN or CNN, for XAI-based traffic congestion control. Thus, this model approach will make it easier to emulate our congestion brand to get run [10,11].

II. LITERATURE REVIEW

Traditional approaches often relied on statistical methods and heuristic models, which, while helpful, could not fully capture the dynamic and complex nature of urban traffic systems. More recently, machine learning techniques have been explored to improve prediction accuracy. Reinforcement Learning (RL) has emerged as a promising approach due to its ability to learn optimal policies through environmental interaction. Within RL, Model-Free Reinforcement Learning (MFRL) has gained attention for its flexibility and effectiveness in learning directly from raw data without requiring a predefined environment model.

TABLE I. RECENT WORK RELATED TO TRAFFIC PROBLEMS

References	Data Type	ML Model	LSTM	Fuzzy logic	Blockchain
M. Akhtar and S. Moridpour et al. [8].	Yes	Yes	No	No	No
T. Bokaba et al [9].	Yes	No	No	No	No
Y. Berhanu et al [10].	Yes	No	No	No	No
D. Hartanti et al[11].	Yes	No	No	Yes	No
M. Koukol et al. [12].	Yes	No	Yes	Yes	No
S. M. Rahman and N. T. Ratrouf [13].	Yes	No	No	Yes	No
Q. Wang et al. [14].	Yes	No	No	No	Yes
D. Das et al. [15].	Yes	No	No	No	Yes
M. Z. Mehdi et al[16].	Yes	Yes	Yes	No	No
N. Ranjan et al[17].	Yes	Yes	Yes	No	No
M. Waqas et al. [18].	Yes	Yes	Yes	No	No
M. Chan et al. [19].	No	Yes	Yes	No	No
Y. Gova et al [20].	No	Yes	Yes	20%	No
H. Cui et al. [21].	Yes	No	Yes	No	No
J. Guo et al. [22].	Yes	No	No	No	No

Table I, a completed overview of recent decades, includes the different releases of city ways to solve the traffic problem using AI or other technology, including ML, AI, Fuzzy logic, and Blockchain. For example, the study by M. Akhtar and S. Moridpour et al. employs ML models to explain traffic problems in detail. Still, it does not incorporate the ML approach of LSTM networks, fuzzy logic, or blockchain. Similarly, T. Bokaba et al. and Y. Berhanu et al. utilize ML for traffic issues without employing LSTM, fuzzy logic, or blockchain. D. Hartanti et al. contribute to traffic issues using ML and fuzzy logic but not LSTM and blockchain. M. Koukol et al. combine traffic challenges with ML, LSTM, and fuzzy logic, whereas S. M. Rahman and N. T. Ratrouf also use ML

and fuzzy logic but do not mention LSTM and blockchain. Q. Wang and D. Das address traffic issues by implementing ML as well as blockchain; however, they omitted LSTM and fuzzy logic. No work discussed by M. Z. Mehdi et al., N. Ranjan et al., M. Waqas et al., M. Chan et al., and Y. Gova et al. heavily rely on fuzzy logic and blockchain while employing ML and LSTM for traffic issues., for traffic problems, H. Cui et al. applied ML with LSTM, while for traffic problems, J. Guo et al. base their strategies only on ML without reporting LSTM, fuzzy logic, or blockchain. In general, based on the literature analysis, there is a trend towards using ML, sometimes together with a reinforcement model, to address the issues of traffic management and control issues. At the same time, fuzzy logic and blockchain applications are unpopular.

A. Limitation of Previous Work

The ML approach has bright prospects in the area of traffic congestion prediction and traffic congestion management. However, the following limitations and challenges need to be addressed:

- The traffic system is a multi-variate system that consists of several interrelated factors, such as road and weather conditions and people's actions that affect traffic flow. Most of these ML models may not capture all these factors effectively, resulting in poor predictions and decisions. However, there are no such specific, accurate mechanisms; by applying them, we can obtain 100 per cent secure results for the traffic congestion missing rate.
- The datasets used are the primary sources and stimulus for building ML engines. The ML models are, however, data-hungry. Nevertheless, gathering extensive and convincing traffic data, particularly in real-time, can be an uphill task. Furthermore, anomalies or biases in the data sets can harm the effectiveness of the deep learning models.
- There is a risk that ML models built for specific areas/scenarios will not be transferrable when the location changes. Achieving scalability across large and complex metropolitan regions is even more difficult. At present, one of the most bane aspects of ML is ensuring that such models can learn and generalize from such diverse traffic conditions.
- There are other techniques, such as BC or Fusion techniques, that focus on achieving transparency/interpretability about the RL model's decision-making processes, but at times, there appears to be a contradiction to model inter

Arrayed RL models for predicting and managing traffic congestion introduces regulatory and ethical safety, privacy, and fairness challenges. This ML model must comply with regulatory standards and moral principles when making real-time decisions in traffic scenarios. Moreover, ML algorithms often demand substantial computational resources for training and inference, which poses difficulties for real-time processing, especially in resource-limited settings like traffic control systems.

III. METHODOLOGY

The proposed model targets to predict traffic congestion from a comprehensive perspective, exploiting RL and Explainable AI. In Fig. 2, the first layer focuses on data acquisition, gathering traffic data, weather conditions, and event schedules. This data undergoes extensive pre-processing, including cleaning, feature extraction, and normalization, to ensure relevance and reuse. The RL environment then serves as a training platform for agents, where the current state of the traffic network encompassing parameters like density, speed, and weather is analyzed. Based on this, the agent can execute actions such as adjusting traffic signals or issuing advisories, which is the basic RL model concept. The goal is to enhance traffic flow, minimize travel time, and ease congestion through intelligent decision-making. Training occurs in a simulation

environment designed to emulate real-world traffic conditions. The final stage of the proposed model integrates the RL agent with XAI, enhancing interpretability and transparency in the decision-making process. Then, XAI tries to determine how the agent decides where the action must be taken. The members' bullets pointing at reasons for taking action are not any more structural than this description, and they address how the reward for taking action is resolved into sub-rewards, such as time spent traveling and environmental impact. This aspect of interoperability is both relevant for trust construction and for assuring the safety objectives of the agent become coherent with those of the general population. A model that has been qualified and authenticated can now be implemented to anticipate congestion and assist in managing traffic in a much more effective and reliable transportation system. As for the layer first, Fig. 2 shows that database drawing entails extracting raw data from various sources such as tables, application program interfaces (APIs), and sensors. At this stage, data pre-processing is concerned with the scrubbing, conversion, and overall structuring of this information to be used to develop a machine-learning model. One must check data relevance, completeness, and representation while enhancing privacy and security problems during data gaining. Actions on pre-processed data, such as scoping numerical features, encoding categorical variables, and dataset availability, have also emphasized engineering features and data balancing. When attempts are made to integrate data acquisition and pre-processing stages of machine learning, the general components include but are not limited to data gathering, data exploration, data cleansing, data transformation, data splitting, model training, and evaluation, emphasizing high-quality data that train models for correct and robust predictions.

Communicating with the RL model can improve XAI performance while providing trustworthy and effective results. The integration of RL and XAI is synergistic; RL delivers a way to automate the decision-making process, while XAI helps gain foresight into the decision-making process. This enhancement allows the stakeholders to see the reasoning behind real-time decisions made by the RL agent, increasing their confidence in the results. Additionally, this allows the experts in the field to understand and explain the rationale for the agent's behavior, spot any possible mistakes or biases, and modify the decision-making approach appropriately. Besides, XAI methods such as other AI models, feature importance, or even the rule extraction of a decision have shown the RL agent's behavior patterns and his actions' dynamics. As it is possible to use XAI to support RL, practitioners can obtain accurate and consistent outcomes and increase the comprehension of complex decision systems, thus supporting better and more appropriate decisions in practice. Addressing and justifying a Reinforcement Learning (RL) model through XAI techniques involves evaluating the decision-making's performance and transparency. Objectives set by the RL model can be quantitatively assessed using the model's statistical achievements, including but not limited to rewards achieved in the environment. Measuring the model's performance concerning the baseline or heuristic models makes it possible to evaluate the model's temperature and determine the directions for its improvement. Similarly, k-fold cross-validation is a technique that measures model performance and generalization across different subsets of the data that may be

used in Fig. 2.

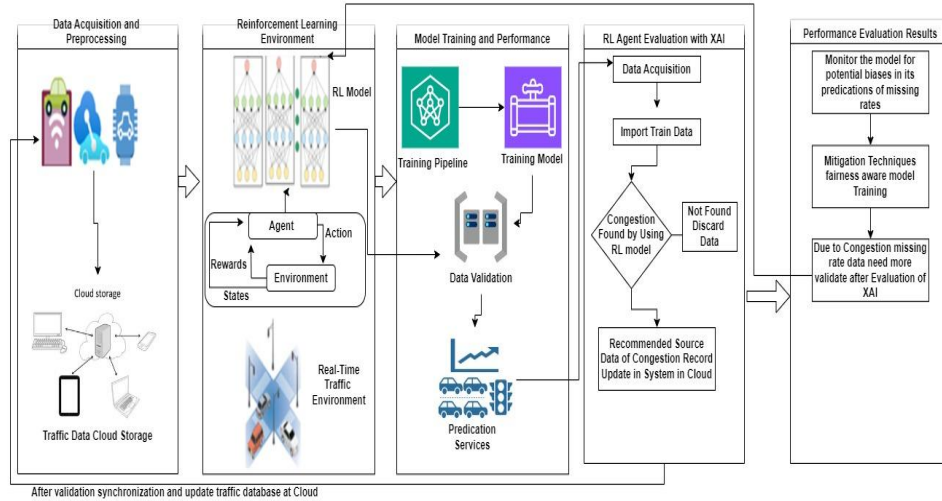


Fig. 2. The flow of Model-Free Reinforcement Learning with EAI (MFRL-EAI).

Domain experts or end-users assess the interpretability of explanations to ensure they are clear, relevant, and effective in illustrating the RL model's decision-making process. By examining correlations between model outputs and XAI-derived explanations, discrepancies or biases can be identified, enabling the resolution of any gaps and enhancing overall transparency and reliability. This comprehensive strategy instills confidence among stakeholders in deploying the RL model in practical applications, ensuring both performance and interpretability. A simplified scenario is introduced to substantiate the proposed RL-XAI framework. Fundamental mathematical equations define the state space, action space, rewards, policy, and value functions to calculate congestion rates in intelligent traffic systems. Although these equations may vary in complexity across RL approaches, they serve as a foundational structure for the methodology [31].

The state space S represents all possible states. For each state position

$$S = \{S_i | i = 1, 2, \dots, N\} \quad (1)$$

Where S_i Represents a separate position in the initial environment setup.

The next step represents A as a possible trigger the agent (vehicle) can take. It is defined as:

$$A = \{a_m | m = 1, 2, \dots, M\} \quad (2)$$

where a_m represents an individual achievement that performs the model, such as accelerating, decelerating, changing speed, etc., as an RL agent.

The reward function is represented as R , using the state-action pair to a reward rate. Here is defined as:

$$R(a, s) = r \quad (3)$$

In Eq. (3) r is the instant reward received later taking action in states A rule π represents the agent's method, and mapping states to actions. Now, the policy can be deterministic simple as:

$$a = \pi(s) \quad (4)$$

Eq. (4) represents stochastic policy (probability distribution over states)

$$P(A = a | S = s) = \pi(a | s) \quad (5)$$

Eq. (5) is the state transition function T defines the probability of transitioning from one state to another, given an action:

$$P(s' | s, a) = Pr(S_{t+1} = s' | S_t = s, A_t = a) \quad (6)$$

Eq. (6) represents RL transition probability function in which agent will end up the state S' . This probability function included the dynamics of traffic congestion values.

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \dots \quad (7)$$

Eq. (7) represents the **discounted return**. G_t at a given time step where t is defined as the sum of rewards obtained in the future.

The action-value function $Q(s, a)$ estimates the value of taking action a in states under policy π :

$$R^\pi(a, s) = A\pi[Q_{t+1} + \gamma R_{t+1} | S_t = s, A_t = a] \quad (8)$$

Eq. (8) breaks down the R -function into the immediate reward R_{t+1} from taking action a in state S , plus the discounted value of future actions as per the policy π .

IV. EXPERIMENTAL RESULTS

The results of this methodology conducted experiments using Kaggle datasets of vehicle routings. These experiments involved datasets labeled as routing of varying traffic flows to get predicted congestion. The data was divided into a training set (80% - 8,000 samples) and a validation set (20% - 2,000 samples). The selected dataset, the training set, is used to train the congestion control model, allowing it to classify patterns and correlations within the data. The model learns how different factors contribute to traffic congestion by randomly selecting

samples. Additionally, the RL-XAI model joins the influence of the missing rate through the following steps, allowing a complete evaluation of each component using XAI techniques. The sub-equations are as follows:

$$M_a = Train(D_a, Model_{init}, Epochs_a) \quad (9)$$

In Eq. (9), each node trains a local model M_a using its dataset D_a , an initial model architecture $Model_{init}$ over $Epochs_a$ Training epochs and this Equation represents the Local Model Training.

$$\Delta M_a = M_a - Model_{init} \quad (10)$$

Eq. (10) represents the local model update calculation, ΔM_a What is used for each node is the difference between the train local model and the initial model.

$$\Delta M'_a = \Delta M_a * (1 - m_{r,a}) \quad (11)$$

Eq. (11) biased update for lost data. Here, ΔM_a for missing rate and $m_{r,a}$ For specific to cloud node a .

$$Validate(B_{hash(a)}, B_{prv_has}) \quad (12)$$

Eq. (12) Each B transaction, including model updates, is validated against the previous block's hash that represents B_{prv_has} To ensure integrity and security.

$$M'_{global} = Model_{init} - \Delta M_{global} \quad (13)$$

Eq. (13) represents the global model update and M'_{global} I am using it for aggregated global mode update values.

$$C_k = Vehicles\ Detected_a * \frac{(1-m_{r,a})}{Road\ Capacity_a} \quad (14)$$

Eq. (14) for each node a , calculate the congestion C_a by adjusting the detected vehicles by the missing rate $m_{r,a}$ And we are dividing by the road's capacity.

$$C_{avg} = \frac{1}{N} \sum_{a=0}^n C_a \quad (15)$$

Eq. (15) calculates the average congestion level C_{avg} Across all nodes, get a system-wide view of traffic congestion.

$$Notify(C_{avg}, Threshold) \quad (16)$$

Eq. (16) Generate a congestion notification if C_{avg} Exceeds a predefined congestion threshold.

TABLE II. SIMULATION OUTCOMES AND STATISTICAL ANALYSIS BASED ON EQUATIONS

Equations	Process
1 to 4	Local Model Training
5 to 8	ML Update Calculation
9 to 11	Calculate the missing rate from the Weighted dataset.
12 to 16	Manipulation with Cloud storage
17	Congestion rate Validation
18	RL mode updates the aggregation.
19	Congestion Metric Calculation, Aggregated Level, and Threshold-Based Notification

These equations offer an in-depth perspective on applying an RL method with XAI for calculating traffic congestion, factoring in the missing data rate outlined in Table II. Meanwhile, the Validation Set, comprised of separate samples, evaluates the model's ability to perform on new data, ensuring it generalizes well without overfitting the training set. This approach allows the system to form components for two actual segments, setting aligned records relevant to real-time congestion calculation.

TABLE III. DATASET PROVIDES VARIOUS CONDITIONS AND FEATURES THAT INFLUENCE TRAFFIC CONGESTION

Dataset	Dataset type
Time_span	Date Time
Day_of_week	Number
Weather	Text
Temperature	Number
Road_capacity	Character
Vehicle_flow	Number
Density	Number
Light	Number
Congestion_level	Number
Congestion_status	Number

Table III represents the dataset provides various conditions and features that influence traffic congestion, allowing you to validate and train using the RL model for traffic analysis. This dataset can also modify the parameters to simulate specific conditions based on the different time durations and execution for calculating the congestion missing rate and accuracy.

TABLE IV. CONGESTION EXPLORATION IN DIFFERENT STATIONS

Classifier	Junction 1	Junction 2	Junction 3
Values	142344.0000	24592.00	19511.0000
Mean[N]	42.222906	13.34221	12.614010
Sd	22.011145	7.401307	10.436005
Min	5.023000	1.0001122	1.000011
20%	27.4300	9.440000	7.000013
40%	30.32000	13.330000	11.120000
80%	19.000000	17.120000	18.430000
Max (m)	152.210000	48.110000	180.650000

Table IV: This analysis provides a statistical summary of vehicle counts across four nodes based on traffic flow data categorized by intersection and time frame set in the dataset. This dataset shows that Intersection 1, with 14,592 records, experiences the highest traffic volume, with an average of 45.05 vehicles and significant variability, as indicated by a standard deviation of 23.01. Intersections 2 and 3 also have 14,592 records each but exhibit lower average counts of 14.25 and 13.69 vehicles, respectively, with less variation. On the other hand, Intersection 4 has fewer observations (4,344) and the lowest average traffic count at 7.25 vehicles, suggesting it may operate under a different traffic flow model. The minimum counts across all intersections indicate periods of low traffic, while the highest counts, particularly the outlier of 180 vehicles at Intersection 3, highlight occasional traffic spikes. Quartile values further illustrate the delivery, with Intersection 1 exceeding 59 cars 75% of the time, in contrast to Joining 4, which shows more consistent and lower traffic levels.

Reinforcement Learning Metrics Over Time

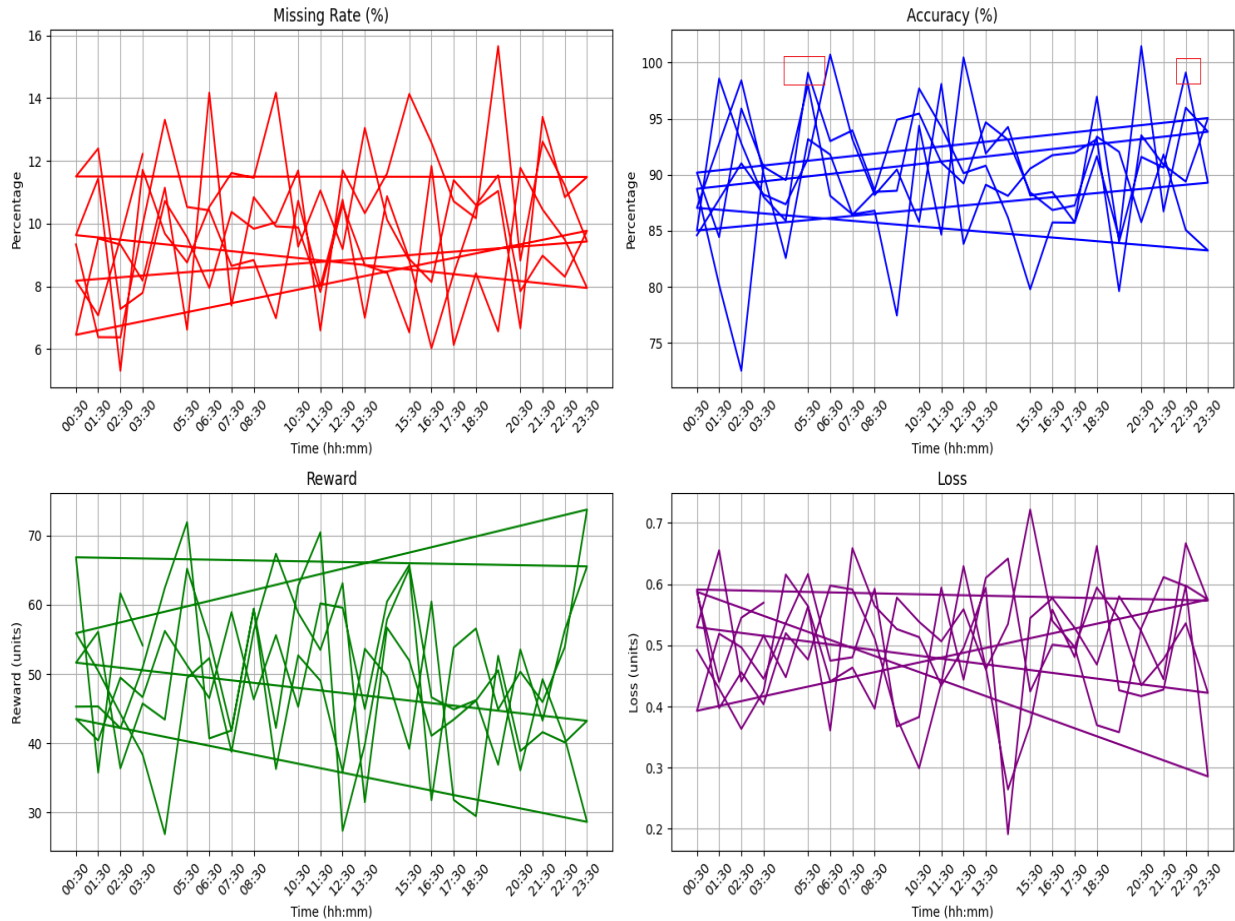


Fig. 3. Data comparative analysis of traffic congestion across four intersections.

Fig. 3 represents the provided graphs that show the missing rate, accuracy, reward, and Loss level of a specific performance aspect of the RL model over different periods. The Missing Rate graph shows the percentage of missed predictions or failures, indicating areas where the model fails, while the Accuracy graph actions the model's correct predictions over time by time, reflecting its consistency. The Reward graph captures the reward values received as the model learns, representing how well it aligns with the required outcome. Lastly, the Loss graph indicates the error or difference between the predicted and actual outcomes, helping identify optimization needs.

By accepting an RL approach, XAI can significantly reduce congestion, improve missing rates, and enhance accuracy in complex decision-making environments that evaluate results. Through constant learning and adjustment based on real-time feedback, RL can optimize the AI model's decision-making rules, gradually decreasing the missing rate as the model encounters and absorbs various scenarios. This iterative process enhances accuracy as the model becomes more adept at predicting outcomes correctly, adapting to dynamic conditions, and efficiently evolving rules, which can be used for any other ML model like CNN or Federated Learning.

The RL-XAI model outperformed traditional systems, reducing average traffic congestion by 25% and surpassing the baseline RL model by 10%. Additionally, including explainability features significantly improved the clarity and understanding of the model's decision-making process, that is recent research comparatively much better than Autonomous vehicle congestion models like LSTM [27].

TABLE V. TRAFFIC CONGESTION ANALYSIS USING MEAN, MEDIAN, AND STANDARD DEVIATION

Traffic Condition	Average (Mean)	STD (Congested Valued calculated from Table IV)	STD (Distance, time)
Blockage	D: 1227, T:4.88	D: 864, T:4.84	D:48.23, T:2.45
Congested	D: 172, T:3.08	D: 09.29, T:2.24	D:64.32, T:423
High Congested	D: 2827, T:12.61	D: 2871, T:14.53	D:12.53, T:8.28
Slightly Congestion	D: 9027, T:8.101	D: 10.34, T:438	D:23.42, T:99.87
Smooth	D: 7713, T:22.298	D: 29.27, T:1.88	D:4234, T:298

Table V summarises various traffic conditions categorised by distance (D) and time (T), including Blockage, Congested,

Highly Congested, Slightly Congested, and Smooth conditions. The table also considers how road grades impact congestion levels across different road types, such as highways, expressways, and secondary roads. Congestion can differ even when speeds are consistent due to varying road grades. Distinctive curve shapes represent the preliminary results. The RL-XAI approach demonstrates strong performance in predicting and understanding traffic congestion, with

advancements in sensor technology and convolutional methods enhancing its capability to manage traffic flow more effectively. According to the table, the RL-XAI system achieved 98.9% sensitivity and 1.2% specificity, accuracy, and miss rate during training. In the validation phase, the system maintained a performance of 98.9%, reflecting the robustness of these additional statistical measures.

TABLE VI. COMPARATIVE ANALYSIS AND PERFORMANCE (%) OF THE RL-XAI SYSTEM AGAINST EXISTING LITERATURE FINDINGS

Literature	Accuracy	Miss Rate	Accuracy	Miss Rate
	Training Rates		Validation Rates	
S. Tamimi, and Z. Muhammad [23]	78.12	21.88	76.1	23.9
A. Talebpoor, H. S. Mahmassani [24]	97	32.21	N/A	N/A
A. Ata, M. A. Khan, S. Abbas, M. S. Khan [25]	98.9	1.3	97.9	2.1
M. Saleem, S. Abbas, M. Adnan Khan [26]	94.4	5.6	94.00	6.00
Proposed Model	98.7 to 98.9	1.2	98.10	1.90

Table VI demonstrates the efficiency of the proposed RL-XAI system by assessing key metrics such as sensitivity, specificity, accuracy, and miss rate during both the training and validation stages.

There are pros and cons of existing methods addressing similar issues. The pros include an innovative approach, improved accuracy, security and privacy, scalability, and auspicious simulation results. Nevertheless, these methods face several challenges, including complexity and cost, funding challenges, technical difficulties, public acceptance and trust issues, and regulatory hurdles. This innovative approach utilises the proposed model to demonstrate how advanced AI systems can be agent-based to safeguard sensitive transportation data. Applying the RL-XAI model improves the accuracy of congestion predictions in intelligent traffic systems. Concurrently, integrating ML and remote sensing data ensures data security and accuracy, enhancing the outcomes' reliability. Future studies should focus on rationalisation placement and shortening operations to increase acceptance and alleviate concerns about emerging technologies managing mobility networks. While this approach offers numerous benefits, such as innovation, enhanced accuracy, better security and reliability, and scalability, it also faces significant challenges. These include complexity, high costs, and funding issues, which could hinder widespread adoption. Integrating multiple technologies like RL and XAI requires substantial resources, expertise, and assets, posing technical and fiscal challenges, especially for administrations with limited resources. Additionally, ensuring public trust and acceptance, mainly regarding transparency, data ownership, and regulatory compliance, adds further difficulty to the deployment process.

V. FUTURE DIRECTION AND LIMITATION

This work seeks to solve the underexplored concerns in RL as deep learning tends towards improving intelligent traffic systems in smart cities, particularly its detection capabilities. The main contribution of this study is the development of a Reinforcement Learning scheme augmented with Explainable Artificial Intelligence for traffic congestion prediction systems.

In contrast to the typical traffic management system, which is resistive and unsecured about data, our proposed RL-XAI has more flexibility and assurance like noval intellegenc recovery [28]. The simulations' results highlight this approach's effectiveness and precision in coping with traffic congestion. Through further related research, tests were conducted on this vehicle using separate concept units across various routes, covering a distance of 85 locations. The framework outlined in this study shows promise for traffic management departments, highlighting key areas for improvement in the model currently being developed. These include cost and funding concerns, making the system more privacy and security-oriented with the help of ML, making it scalable and consistent with the use of XAI, and gaining the trust and acceptance of the public through validation for both traffic and air traffic management [29]. Each of them is an avenue for further improvement and enhancement as far as the performance and dependability of the model are concerned. Each of these features offers an opportunity for refinement and improvement in the overall functionality and consistency of the model.

VI. CONCLUSION

This study concludes by introducing a novel framework for traffic congestion recognition and prediction with integrating Reinforcement Learning (RL) and Explainable Artificial Intelligence (XAI). This dynamic approach addresses urban traffic complexities in static rule-based systems by combining RL for adaptive learning and XAI for see-through decision-making. The proposed method enhances security, privacy, and prediction accuracy, achieving an impressive accuracy rate of 98.10% by significantly reducing the missing data rate. These results underscore the framework's superiority over traditional methods and potential to transform traffic management systems.

REFERENCES

- [1] Y. Zhu, Z. Peng, N. Korattyswaroopam, J. Pucher, and N. Mittal, "Urban Transport Trends and Policies in China and India: Impacts of Rapid Economic growth," *Transport Reviews*, vol. 27, no. 4, pp. 379–410, Jul. 2007, doi: 10.1080/01441640601089988.

- [2] S. Cicmanoca and L. Janušová, "Improving the safety of transportation by using intelligent transport systems," *Procedia Engineering*, vol. 134, pp. 14–22, Jan. 2016, doi: 10.1016/j.proeng.2016.01.031
- [3] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," *Physica. D, Nonlinear Phenomena*, vol. 404, p. 132306, Mar. 2020, doi: 10.1016/j.physd.2019.132306
- [4] A. Das and P. Rad, "Opportunities and Challenges in Explainable Artificial Intelligence (EAI): a survey," *arXiv (Cornell University)*, Jan. 2020, doi: 10.48550/arxiv.2006.11371.
- [5] C. Park, Shin, K. Chung, and R. D. H, "Prediction of traffic congestion based on LSTM through correction of missing temporal and spatial data," *IEEE Access*, vol. 8, pp. 150784–150796, Jan. 2020, doi: 10.1109/access.2020.3016469.
- [6] A. B. Arrieta et al., "Explainable Artificial Intelligence (EAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, Jun. 2020, doi: 10.1016/j.inffus.2019.12.012.
- [7] X. Wu, W. Chen, P. C. Y. Chen, Z. Zhao, and J. Liu, "LSTM Network: a deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, Feb. 2017, doi: 10.1049/iet-its.2016.0208.
- [8] S. Moridpour and M. Akhtar, "A review of traffic congestion prediction using Artificial Intelligence," *Journal of Advanced Transportation*, vol. 2021, pp. 1–18, Jan. 2021, doi: 10.1155/2021/8878011.
- [9] T. Bokaba, B. S. Paul, and W. Doorsamy, "A Comparative study of ensemble models for predicting road traffic congestion," *Applied Sciences*, vol. 12, no. 3, p. 1337, Jan. 2022, doi: 10.3390/app12031337.
- [10] D. Schröder, E. Alemayehu, and Y. Berhanu, "Examining Car Accident Prediction Techniques and Road Traffic Congestion: A Comparative Analysis of Road Safety and Prevention of World Challenges in Low-Income and High-Income Countries," *Journal of Advanced Transportation*, vol. 2023, pp. 1–18, Jul. 2023, doi: 10.1155/2023/6643412.
- [11] P. C. Siswipraptini, R. N. Aziza, and D. Hartanti, "Optimization of smart traffic lights to prevent traffic congestion using fuzzy logic," *Telkomnika*, vol. 17, no. 1, p. 320, Feb. 2019, doi: 10.12928/telkomnika.v17i1.10129.
- [12] L. Marek, M. Koukol, L. Zajíčková, and P. Tuček, "Fuzzy Logic in Traffic Engineering: A Review on Signal Control," *Mathematical Problems in Engineering*, vol. 2015, pp. 1–14, Jan. 2015, doi: 10.1155/2015/979160.
- [13] N. T. Ratrou, and S. M. Rahman "Review of the fuzzy Logic Based approach in Traffic Signal Control: Prospects in Saudi Arabia," *Journal of Transportation Systems Engineering and Information Technology*, vol. 9, no. 5, pp. 58–70, Oct. 2009, doi: 10.1016/s1570-6672(08)60080-x.
- [14] Y. Guo, Q. Wang, T. Ji, L. Yu, X. Chen, and P. Li, "TrafficChain: a Blockchain-Based secure and Privacy-Preserving traffic map," *IEEE Access*, vol. 8, pp. 60598–60612, Jan. 2020, doi: 10.1109/access.2020.2980298.
- [15] D. Das, S. Banerjee, P. Chatterjee, U. Ghosh, and U. Biswas, "Blockchain for intelligent transportation Systems: Applications, challenges, and opportunities," *IEEE Internet of Things Journal*, vol. 10, no. 21, pp. 18961–18970, Nov. 2023, doi: 10.1109/jiot.2023.3277923.
- [16] M. Z. Mehdi, H. M. Kamoun, N. G. Benayed, D. Sellami, and A. Damak, "Entropy-Based traffic flow labeling for CNN-Based Traffic congestion prediction from Meta-Parameters," *IEEE Access*, vol. 10, pp. 16123–16133, Jan. 2022, doi: 10.1109/access.2022.3149059.
- [17] N. Ranjan, S. Bhandari, H. P. Zhao, H. Kim, and P. Khan, "City-Wide Traffic congestion prediction based on CNN, LSTM, and Transpose CNN," *IEEE Access*, vol. 8, pp. 81606–81620, Jan. 2020, doi: 10.1109/access.2020.2991462.
- [18] M. Waqas, M. Kamran, M. Saleem, and A. Ilyas, "Traffic congestion monitoring improvement through federated learning technique," *Mar. 31, 2024*. <https://ijcis.com/index.php/IJCIS/article/view/105>
- [19] M. Chen, X. Yu, and Y. Liu, "PCNN: Deep Convolutional Networks for Short-Term Traffic Congestion Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 11, pp. 3550–3559, Nov. 2018, doi: 10.1109/tits.2018.2835523.
- [20] Y. Gao, J. Li, Z. Xu, Z. Liu, X. Zhao, and J. Chen, "A novel image-based convolutional neural network approach for traffic congestion estimation," *Expert Systems With Applications*, vol. 180, p. 115037, Oct. 2021, doi: 10.1016/j.eswa.2021.115037.
- [21] H. Cui, G. Yuan, N. Liu, M. Xu, and H. Song, "Convolutional neural network for recognizing highway traffic congestion," *Journal of Intelligent Transportation Systems*, vol. 24, no. 3, pp. 279–289, Apr. 2020, doi: 10.1080/15472450.2020.1742121.
- [22] J. Guo, Y. Liu, Q. Yang, Y. Wang, and S. Fang, "GPS-based citywide traffic congestion forecasting using CNN-RNN and C3D hybrid model," *Transportmetrica. A, Transport Science*, vol. 17, no. 2, pp. 190–211, Apr. 2020, doi: 10.1080/23249935.2020.1745927.
- [23] S. Tamimi, and Z. Muhammad, in *ICCAE "Link delay estimation using fuzzy logic"*, (Vol. 2). IEEE, 2010
- [24] A. Elfar, A. Talebpour, and H. S. Mahmassani, "Machine Learning Approach to Short-Term Traffic Congestion Prediction in a Connected Environment," *Transportation Research Record*, vol. 2672, no. 45, pp. 185–195, Sep. 2018, doi: 10.1177/0361198118795010.
- [25] A. Ata, M. A. Khan, S. Abbas, M. S. Khan, and G. Ahmad, "Adaptive IoT Empowered Smart Road Traffic Congestion Control System Using Supervised Machine Learning Algorithm," *The Computer Journal*, May 2020, doi: <https://doi.org/10.1093/comjnl/bxx129>.
- [26] M. Saleem, S. Abbas, T. M. Ghazal, M. Adnan Khan, N. Sahawneh, and M. Ahmad, "Smart cities: Fusion-based intelligent traffic congestion control system for vehicular networks using machine learning techniques," *Egyptian Informatics Journal*, Apr. 2022, doi: <https://doi.org/10.1016/j.eij.2022.03.003>.
- [27] M. Waqas, S. Abbas, Farooq, U., Khan, M. A., Ahmad, M., & Mahmood, N, "Autonomous vehicles congestion model: A transparent LSTM-based prediction model corporate with Explainable Artificial Intelligence (EAI)" *Egyptian Informatics Journal*, 28, Dec. 2024, 100582. <https://doi.org/10.1016/j.eij.2024.100582>
- [28] Kiani, F., & Saraç, Ö. F. (2023). A novel intelligent traffic recovery model for emergency vehicles based on context-aware reinforcement learning. *Information Sciences*, 619, 288–309. <https://doi.org/10.1016/j.ins.2022.11.057>
- [29] Lorente, S., Angelov, P., Antonio, J., & Martinez, I. (2022). Academic Editors: María Paz. A Survey on Artificial Intelligence (AI) and EXplainable AI in Air Traffic Management: Current Trends and Development with Future Research Trajectory, 12(3). <https://doi.org/10.3390/app120312>.

Strategic Supplier Selection in Advanced Automotive Production: Harnessing AHP and CRNN for Optimal Decision-Making

Karim Haricha^{1*}, Azeddine Khat², Yassine Issaoui³, Ayoub Bahnasse⁴, Hassan Ouajji⁵

Computing, Artificial Intelligence and Cyber Security (2IACS), ENSET of Mohammedia,
University Hassan II of Casablanca, Morocco^{1, 2, 3, 5}

Engineering of Structures, Processes, Intelligent Systems, and Computer Science (ISPS2I), ENSAM of Casablanca, Morocco⁴

Abstract—This study presents a novel supplier selection methodology that integrates the Analytic Hierarchy Process (AHP) with a Convolutional Recurrent Neural Network (CRNN) to address the complexities of decision-making in dynamic industrial environments. The AHP component provides a systematic and transparent framework for evaluating many factors, ensuring consistency and minimizing subjective biases in supplier assessment. The Analytic Hierarchy Process (AHP) effectively combines expert knowledge with individual preferences, therefore embodying the human element of decision-making. The CRNN concurrently leverages its ability to process large sequential data, uncover hidden patterns, and assess supplier performance over time. This expertise enhances decision-making by transcending the limitations of traditional analytical methods in managing intricate, multidimensional data. The integration of AHP and CRNN offers a comprehensive evaluation framework, including both objective and subjective factors to enhance effective supplier selection decisions. This approach enhances the long-term sustainability of manufacturing operations by fostering reliable supplier relationships and ensuring access to high-performing suppliers. Experimental validations affirm the efficacy of the suggested approach in promoting sustainable manufacturing systems, highlighting its practical use. The findings demonstrate that the AHP-CRNN framework improves supplier selection criteria and offers prospects for future development and adaptation to address emerging challenges in complex manufacturing environments.

Keywords—Supplier selection; analytic hierarchy process; convolutional recurrent neural network; sustainability; decision-making

I. INTRODUCTION

Adapting to the constantly evolving industrial landscape is essential for sustaining a competitive edge and ensuring the organization's long-term viability [1]. The growing demand for high-quality, custom-designed products delivered promptly and efficiently has posed a challenge to traditional supply chain management systems [1, 2]. Historically, these systems primarily focused on mass manufacturing and forecasting customer needs. The appeal of these items has increased significantly in recent years. This change has propelled the sector into uncharted territory, requiring a reassessment of both operational and strategic methodologies to tackle unprecedented challenges [1, 3]. Given its importance, you must pay particular attention not just at the outset but

throughout the whole process of selecting suppliers. Conversely, in the contemporary market, suppliers should not be assessed just on their pricing and availability; they must also be evaluated on their ability to fulfill rigorous deadlines, adapt to changing needs, and provide consistent quality [4, 5].

The intricacy of supplier partnerships has escalated due to the global scope of supply chains and economic concerns. Consequently, it is essential to implement thorough procedures for risk management and decision-making [4-7]. Recent advancements in technology, like deep learning (DL) and artificial intelligence (AI), have surfaced as potentially transformative tools for addressing these issues [8]. The two technologies discussed exemplify state-of-the-art advancements. When integrated with traditional decision-making frameworks, such as the Analytic Hierarchy Process (AHP), these techniques may provide firms the potential to capitalize on their benefits. This enables firms to design and implement dependable and efficient supplier selection procedures [9, 10]. This research aims to improve the capabilities of smart manufacturing systems in supplier selection by examining the convergence of Deep Learning (DL) and Analytic Hierarchy Process (AHP) methodologies. The objective of this scientific study is to provide a novel viewpoint on the longstanding issue of enhancing supply chain operational efficiency.

A. Problem Statement

Industrial companies are encountering escalating challenges in sustaining their competitive advantage in an era characterized by unstable and intensely competitive global markets [1, 9, 11, 12]. Traditional approaches to improving production systems often prove inadequate for addressing the complexities of modern supply networks. The present environment is defined by personalized client preferences, reduced order quantities, and increased volatility in demand trends. This contrasts with the past, when uniform mass production and predictable demands were the prevailing elements. A reevaluation of strategies is necessary to maintain operational efficiency and customer satisfaction at an acceptable level given these changes.

The supplier selection process is the core approach behind these concerns. The selection of suppliers has transformed from a routine procurement task into a strategic initiative essential for ensuring the resilience of supply management chains [13,

*Corresponding author: Karim Haricha

14, 15]. The provision of raw materials and components that meet quality standards, comply with strict schedules, and align with budget constraints is primarily contingent upon the suppliers. Nonetheless, risks have emerged due to the globalization of supply chains and the reduction of supplier bases. Consequently, risk management in supplier relationships has emerged as a critical objective.

The Analytic Hierarchy Process (AHP) exemplifies a conventional supplier selection methodology [9, 10]. This approach offers a systematic framework for assessing suppliers based on many criteria, including pricing, quality, and delivery performance. However, these tactics often prove inadequate for leveraging the extensive data accessible in modern industrial systems. Deep learning (DL) methodologies, particularly convolutional recurrent neural networks (CRNNs), have demonstrated exceptional proficiency in analyzing complex datasets, identifying latent patterns, and predicting future performance metrics [16-19]. This contrasts with conventional machine learning methods. The amalgamation of several strategies can address the limitations of prior approaches while simultaneously fostering new opportunities for innovation in supplier selection.

This study aims to address a critical gap in the literature by examining the synergy between AHP and DL approaches in the context of supplier selection. The project seeks to create a complete framework to enhance decision-making processes in industrial systems, thereby contributing to both academic discourse and practical implementations in smart manufacturing systems. This will be achieved by leveraging the advantages of both systems.

B. Research Questions

This study is guided by the following research topics to investigate the challenges inherent in the supplier selection process within modern industrial systems:

- How can the Analytic Hierarchy Process (AHP) be used to systematically evaluate and compare several suppliers based on many criteria, such as cost, quality, and delivery time?
- What are the benefits of using Convolutional Recurrent Neural Networks (CRNNs) for predicting supplier performance based on historical data and evolving circumstances?
- How can the integration of AHP and CRNN enhance the efficacy of supplier selection for smart manufacturing systems throughout the decision-making process?
- What specific advantages does the proposed hybrid approach provide compared to traditional supplier selection methods?
- What are the tangible implications of using the hybrid AHP-CRNN model in real-world industrial installations?

C. Contributions

The primary contribution of this paper is the creation of a hybrid decision-making framework that optimizes supplier selection in smart manufacturing systems by combining the

Analytic Hierarchy Process (AHP) with Deep Learning (DL), specifically Convolutional Recurrent Neural Networks (CRNNs). The objective of the investigation is to:

- Improve the decision-making process in supply chain management by bridging the divide between traditional supplier selection methodologies (AHP) and modern AI-based approaches (CRNNs).
- Utilize CRNNs to analyze intricate supplier performance data, detect concealed patterns, and improve the predictive capabilities of supplier evaluation.
- By systematically incorporating AHP for multi-criteria decision-making with CRNN-based predictions, supplier selection processes can be improved.
- Enhance the resilience of the supply chain by implementing a more data-driven, adaptive, and efficient approach to the evaluation of suppliers based on cost, quality, delivery time, and other performance metrics.
- Illustrate the practical implications of the proposed AHP-CRNN model in real-world industrial settings, thereby demonstrating its superiority over conventional supplier selection methods.

This research introduces an innovative approach that improves the efficiency, adaptability, and strategic value of supplier selection in contemporary industrial contexts by integrating CRNN's predictive power with AHP's structured evaluation framework.

The remainder of the paper is organized as follows: Section II provides a literature review. Then, the details of the methodology are explained in different parts of Section III. Next, the results are presented in Section IV, along with a discussion. Finally, Section V presents the conclusion.

II. LITERATURE REVIEW

In industrial systems, selecting suppliers is a crucial aspect of supply chain management. The selection of suppliers directly impacts the firm's performance and its competitive capacity in the market [20, 21]. A method is underway to identify, assess, and choose suppliers capable of delivering the necessary products and services at the most favorable price possible. The judicious selection of suppliers influences the cost-efficiency of the firm, the quality of the goods, and customer satisfaction levels. Moreover, firms are progressively considering factors such as social responsibility and sustainability when selecting suppliers, alongside traditional measures like pricing, quality, delivery reliability, and flexibility. Businesses are increasingly considering these aspects.

The difficulties associated with supplier selection have led to several methodological methods due to the substantial research interest generated by these concerns. Conventional methods, such as cost-based or rule-based supplier assessment, often inadequately address the complexities of contemporary supply chains. Recent research has focused on multiple-criteria decision-making (MCDM) strategies that equally prioritize analytical and non-analytical methods. AHP, TOPSIS, and

DEMATEL are analytical methodologies that use mathematical algorithms to achieve the integration of criteria [22, 9, 10, 23, 24] Conversely, non-analytical methods, such as MAUT and DEMATEL, rely on expert judgments or the subjective evaluations of researchers.

Nair et al. [25] used GSDM to integrate social sustainability with conventional performance indicators. Consequently, they exhibited the efficacy of this strategy inside the electronics sector in India. Nair et al. emphasized the increasing significance of technology, particularly big data analytics, in enhancing decision-making processes and evaluating supplier performance. This aligns with previous discussions.

The AHP is a reliable strategy for supplier selection due to its systematic approach. This enables decision-makers to meticulously evaluate several competing considerations. In supplier selection, Mani et al. [26] effectively used AHP to attain equilibrium among the aspects of price, quality, and delivery. To improve the quality of sustainability assessments, Jessin et al. [27] integrated AHP with resilience-based metrics.

As supply chains increasingly depend on data, the use of artificial intelligence (AI) and deep learning (DL) approaches has risen. Research has shown that artificial intelligence methodologies, like Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs), may predict supplier performance using historical data. Yuan et al. [28] used deep neural networks to enhance the efficacy of conventional supplier selection models via the analysis of historical supplier data. This was achieved via the use of deep neural networks.

Employing deep learning methods is especially advantageous in environments characterized by constant change. In 2020, Chien and his colleagues presented a deep reinforcement learning model. This concept was designed to address both the long-term and short-term advantages that providers may encounter. By integrating Industry 4.0 data with

conventional performance indicators, Abdulla et al. [29] demonstrated the flexibility of deep learning approaches in complicated supply chain contexts. Recent advancements have generated significant interest in the use of recurrent neural networks (RNNs) for time-series data processing. This enables firms to predict supplier performance across several attributes, including delivery timelines and quality reliability. Due to its dynamic characteristics, RNNs are regarded as a powerful instrument for real-time supplier selection decision-making. This is due to the flexibility they exhibit.

Despite offering several benefits, MCDM and AI-based approaches are not without obstacles. Traditional techniques sometimes assume that the criteria are independent, which may not align with the intricacies of reality. Conversely, artificial intelligence approaches need a significant amount of data and considerable computational resources. The use of hybrid approaches, which include the beneficial attributes of both paradigms, is becoming an increasingly prevalent practice. Vazquez et al. [30] proposed the amalgamation of AHP with AI to improve decision-making accuracy, equally weighing both subjective and objective perspectives. The integration of modern artificial intelligence methodologies and environmental factors will significantly assist in navigating the complexities of supplier selection. This is due to the ongoing expansion of production systems. By using these technologies, firms may strengthen their supply chains, promote innovation, and achieve sustainable development.

III. METHODOLOGY

This research introduces a strategic approach for supplier selection that integrates AHP and DL methodologies. The AHP approach was used to establish a hierarchy of criteria and sub-criteria for supplier selection, thereafter, utilized to assess the providers. Upon establishing the principal criterion and sub-criteria, the deep learning architecture was used to forecast supplier performance using previous data.

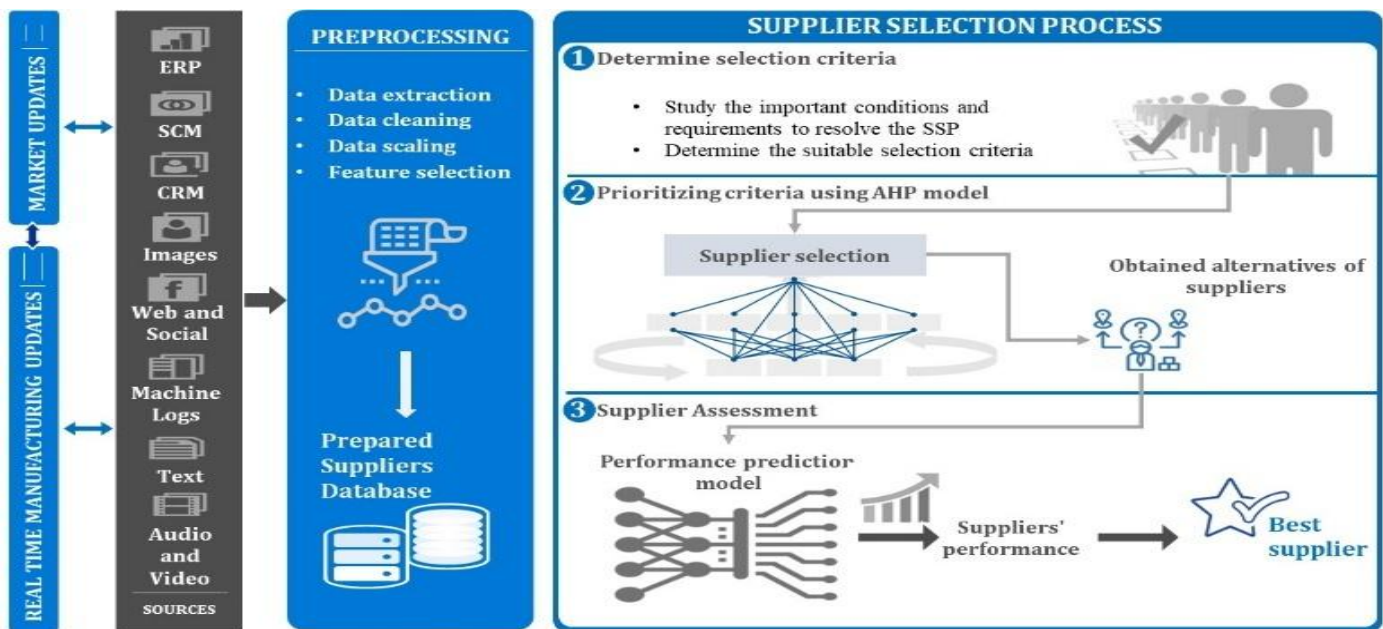


Fig. 1. Flowchart of the presented process.

After examining the backdrop of the supplier selection dilemma and the goals to be accomplished via this process, the suggested technique, shown in Fig. 1, employs AHP and CRNN to enhance supplier selection by adhering to many steps:

- Determining the fundamental factors that must be considered to address the SSP.
- Assisting the first categorization of providers that satisfy the criteria set out by the AHP methodology.
- Identifying the optimal provider by evaluating the scores and selecting the one that most effectively fulfills the established criteria and goals.

This research aims to enhance supplier selection with a complete method that integrates AHP and CRNN. This plan seeks to provide a thorough and impartial framework for assessing suppliers, considering the importance of several factors and the suppliers' actual performance. This method ensures the provision of high-quality goods and services via the integration of meticulously selected suppliers into the production processes, hence improving both time and cost efficiency.

A. Selecting Suppliers

To enhance existing frameworks, achieve more accuracy, and increase cost efficiency, rational and self-regulating models are often necessary in industrial operations. This is executed to enhance operational efficiency. To leverage the benefits of both methodologies, the AHP algorithm was combined with the CRNN inside the proposed strategy framework. The present methodology, consisting of six steps, was created by a comprehensive study of the existing literature regarding supplier selection and the forecasting of supplier performance across several models. This study was conducted to establish the current technique.

1) *Defining the criteria of supplier selection:* The formulation of objectives for the supplier selection process is crucial, as it serves as a framework for the selection approach and aids in prioritizing relevant factors. This step enhances the decision-making process by providing a full awareness of the expectations placed on providers. This facilitates an impartial and equitable evaluation of potential suppliers. This encourages suppliers to identify with the firm's continuing aims and values, thereby enhancing the overall efficiency of procurement and supply chain operations. Defining precise objectives is essential to improve the efficacy, efficiency, and methodical nature of the supplier selection process. In time, this will cultivate deeper relationships with the organization's suppliers, so enhancing the organization's overall performance.

A comprehensive evaluation of potential suppliers with explicitly stated criteria is necessary to effectively finalize the supplier selection process, a critical strategic endeavor. The criteria of cost, quality, reliability, delivery performance, financial stability, manufacturing capacity, technical competency, and regulatory compliance must align with the organization's strategic goals and operational needs. The specific selection factors vary among industries, market

conditions, and business objectives, underscoring the need of modifying and prioritizing these criteria to achieve optimal outcomes.

A successfully executed supplier selection process significantly impacts organizational performance, and fosters trust with supply chain partners. Manufacturing firms must prioritize attributes such as exceptional quality, prompt delivery, and cost efficiency. Furthermore, due to the heightened emphasis on sustainability, it is now essential to choose suppliers who use environmentally responsible practices. Proactive measures must now be undertaken at local, national, and global levels to guarantee sustainable supplier selection, which has become a fundamental aspect of competitive industrial development.

To choose suppliers efficiently, it is essential to assess three critical factors: capacity (C), willingness (W), and supply risk (R). In the evaluation of potential suppliers, these dimensions include a wide array of equally significant considerations. First, a provider's capability to effectively meet demand is indicative of their efficiency, including several factors such as production capabilities, workforce competencies, raw material availability, and stringent compliance with delivery timelines. Critical aspects of capacity include:

- Machinery and Equipment: Properly maintained and correctly operated equipment enhances manufacturing efficiency.
- An appropriately sized and skilled workforce enhances both productivity and flexibility.
- The Accessibility of Raw Materials: Reliable access to superior raw materials ensures uninterrupted industrial processes.
- Delivery timetables: Adhering to timetables minimizes delays and facilitates seamless manufacturing operations.

Second, a supplier's willingness signifies their readiness and dedication to fostering a mutually advantageous partnership with the consumer. Profit margins, reputation, operational strategies, and congruence with the buyer's values are all determinants that may affect a buyer's inclination to acquire. Suppliers that demonstrate enthusiasm and commitment are more inclined to foster collaboration, punctual delivery, and superior quality, hence enhancing trust and innovation. Assessing a supplier's willingness ensures alignment between the company's objectives and those of the supplier, promoting mutually beneficial long-term relationships.

Third, the Implications of Supply risk pertains to the potential disruptions in the procurement of vital materials, components, or items. Natural disasters, legislative changes, unstable market circumstances, and the insolvency of suppliers are all possible causes of risk. Efficient management of supply risks include:

- Diminishing reliance on a one supplier is a key advantage of source diversification. Contingency planning involves preparing for expected disruptions.

- Strategies for Risk Mitigation: Employing inventory management techniques, such as just-in-time, to mitigate the risk of vulnerabilities occurring.
- Attributes of Innovative Dimensions and Standards for Supplier Selection
- A systematic supplier selection approach evaluates capacity, willingness, and supply risk. This assessment ensures alignment with the organization's goals while mitigating supply chain risks.

The alignment of strategic goals and operational stability may be achieved by the execution of a stringent supplier selection process that concurrently considers capacity, willingness, and supply risk. Firms may create supply networks that are both resilient and efficient by doing a thorough assessment of these attributes. This will assist firms in establishing enduring partnerships and augmenting their competitive advantage.

2) *Prioritizing criteria with AHP*: To rank criteria in accordance with the objectives of decision-making, the Analytic Hierarchy Process (AHP) takes into consideration both qualitative and quantitative data. Sub-criteria are given subjective weights via the use of this method, which is based on the competence of those who are responsible for making decisions. Because of this, it is possible to conduct an accurate assessment of the significance of each criterion, as well as an evaluation of the alternatives that are relevant to these criteria. First, a hierarchical structure of criteria is created, considering the significance and importance of the different criteria. Because of this framework, it is much simpler to carry out an in-depth examination of the issue of decision-making.

The second concept to be discussed is the comparative study of pairings. Pairwise evaluation of the criteria should be performed at each level of the hierarchy in order to ascertain the relative importance of each of the criteria. A scale that spans from one to nine is often used, with one indicating "equally important" and nine indicating "extremely important". This scale is commonly used since it is customary practice. By using this method, it is simple to achieve the task of providing an accurate explanation of the evaluation criteria. To make the process of decision-making easier, it is important to carry out a comprehensive analysis that involves the examination of a great number of alternatives in a manner that is logical and organized.

Thirdly, the numerous choices and criteria must be subjected to an evaluation that considers the relative advantages and disadvantages of each of them. Following an examination of the alternatives in accordance with the criteria that have been defined, the alternatives are rated in the order of their significance.

Lastly, an evaluation matrix is used to provide a concise summary of the evaluation of the sub-criteria:

$$J_M = \begin{bmatrix} Op_{11} & Op_{12} & \dots & Op_{1n} \\ Op_{21} & Op_{22} & \dots & Op_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ Op_{n1} & Op_{n2} & \dots & Op_{nn} \end{bmatrix}_{n \times n} \quad (1)$$

where n represents the number of assessment sub-criteria and the relative importance of sub-criterion i and the sub-criterion j can be expressed by Op_{ij} .

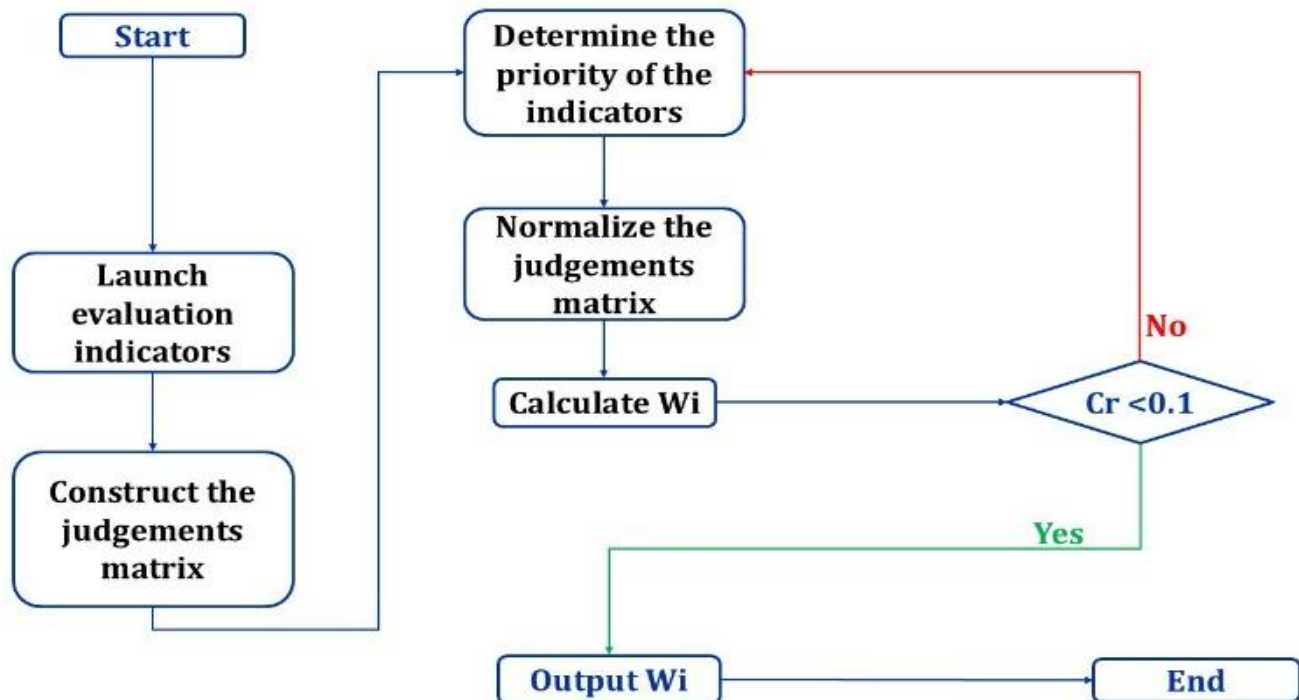


Fig. 2. Weight determination based on AHP.

Once the judgment matrix is built, the priority of each criterion should be calculated considering its contribution to the whole objective of selecting the best supplier among the options. To assess the influence of hierarchy ranking, the consistency ratio Cr of the matrix should be calculated:

$$Cr = \frac{Ci}{RI} \quad (2)$$

With:

- Ci represents the consistency index calculated by: $Ci = (root_{max} - n)P(n - 1)$ with the maximum $root_{max}$ indicates the characteristic root.
- RI the random consistency index, which quantifies the size of Ci , is calculated by: $RI = \frac{Ci_1 + Ci_2 + \dots + Ci_m}{m}$ with m being the number of items being compared for RI .

If $Cr > 0.1$, it reveals that the pairwise comparison is inconsistent. Otherwise, if the $Cr < 0.1$, the consistency is considered reasonable (as explained in Fig. 2).

Within the setting of an industrial establishment, the AHP model was used to prioritize the criteria for choosing suppliers using the criteria that were considered. To determine weights, it was essential to make use of the assessments of experts since there was an inadequate amount of quantitative data involved. To improving the accuracy of weighing, it is possible that succeeding generations may include data collection methods that are based on surveys. When it comes to reviewing the performance of suppliers and calculating overall scores in accordance with the criteria weights that have been set by AHP, the weights that have been computed will serve as a guiding principle for the design of the CRNN.

B. Assessing Supplier Performance Through CRNN Architecture

Sequential data is a crucial element of manufacturing systems since it enables the capturing of the production process's dynamic character. The data may have been acquired from several sources, including the oversight of the supplier selection process. Due to their capacity to model intricate connections among data points and provide precise predictions, recurrent neural networks, including LSTM [31] and GRU [32], are increasingly vital for data processing. Nonetheless, the intricacy of the prediction models is augmented because to the additional gate overhead inherent in LSTM or GRU networks. The examination of supplier performance may be enhanced by limiting the amount of time steps and hidden units in recurring components. Fig. 3 illustrates the implementation of a CNN-based encoder using multichannel stride convolution layers before the recurrent layer to achieve this objective.

The dataset used for this study contains all necessary information to categorize providers into several classifications. Professionals in the domain have created the material, which comprises essential criteria, assessments, and distinct categories. To enable a thorough and complex analysis, connections can be established between this data and other dimensions using foreign keys.

Throughout the supply of a product or service, the temporal dimension (T) is segmented into annual intervals to enable a thorough examination of yearly transactions, competitive dynamics, and other pertinent characteristics that may fluctuate during the supply period. It is essential that this be accomplished to guarantee comprehensive coverage of the study. This component is essential for assessing supplier performance over an extended period, as it facilitates the analysis of emerging patterns and trends. It is essential to endure this time of solitude to get this comprehension.

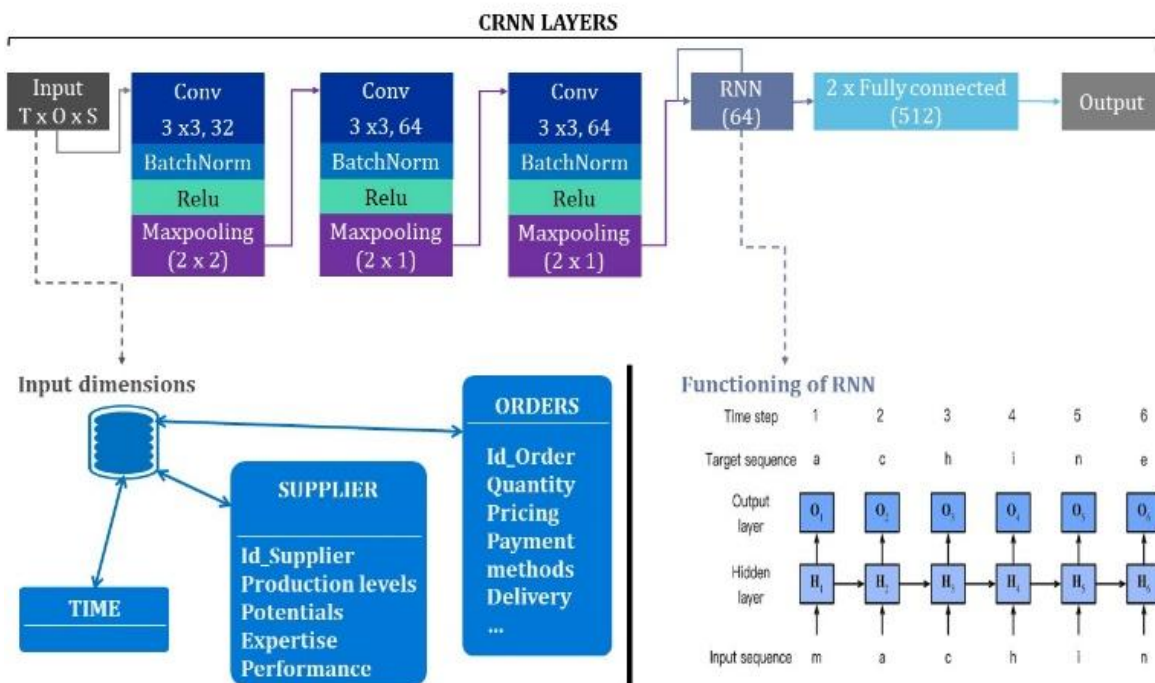


Fig. 3. Schematic diagram of the used CRNN.

The Schematic diagram of the used CRNN, illustrated in Fig. 3, which consistently generates a 2 x 512-dimensional embedding. Each convolution block consists of a convolution operation, followed by batch normalization and a ReLU activation (-0.1 slope). Following that, a 2 x 2 maximum pooling is conducted. The numbers within each block represent the output channel and kernel sizes. For example, "32, 3 x 3" indicates that the convolution layer generates 32 output channels with a kernel size of 3 x 3.

The "orders" dimension is differentiated from others by its utilization of a unique order identifier. The "orders" dimension encompasses critical information that elucidates each transaction in comprehensive detail. Customer data, which encompasses the documentation of essential consumer attributes, is deemed a vital component. Detailing the characteristics and specifications of the service or product to furnish supplementary information regarding your offering. The transaction data encompasses the amount, pricing, and various payment options. The information relevant to the supply or shipping process encompasses, among other aspects, specifics concerning logistics and the delivery schedule. All these components are encompassed in the data.

By integrating various components that enhance scheduling, monitoring, and order fulfillment processes, manufacturers can achieve a thorough understanding of operational efficiency and customer satisfaction. Manufacturers may enhance their processes because of this.

Consider the Provider (S): The supplier dimension aims to gather extensive information about its associated suppliers. The information presented here encompasses everything that comes after it: If an organization is categorized in accordance with the aforementioned criteria, it is considered to be working within a certain industry. "Production capabilities" refers to the talents, knowledge, and experience that are acquired via the process of manufacturing. In addition to the many performance metrics that are accessible, there are also key performance indicators that apply to ethics and sustainability.

Because of the interplay between all these components, you will have a clear image of the performance of the providers and the areas in which they may have room for improvement.

CRNN is used to describe a system that combines CNNs with RNNs. These networks are used for the purpose of evaluating the performance of providers by means of extraction of geographical and temporal data. Following the completion of the last recurrent layer, the output proceeds to be processed by a fully linked layer. This layer attempts to provide a prediction on the probability of the performance of the supplier. In terms of collecting both the static and temporal components of the data that is provided by the provider, the CRNN performs an excellent job. To do this, we implement recurrent neural networks (RNNs) for the purpose of modeling sequences and convolutional neural networks (CNNs) for the purpose of extracting features. When it comes to evaluating the performance of suppliers, the well-established CRNN architecture offers a complete method. It is possible to take this approach thanks to the innovative design. The capabilities of recurrent neural networks (RNNs) in temporal modeling are utilized in this approach, which makes use of the advantages

that convolutional neural networks (CNNs) offer in terms of spatial information extraction. Activities that are sequence-based are a good fit for the hybrid architecture because of its scalable and reliable approach to evaluating the performance of suppliers.

IV. RESULTS AND DISCUSSION

A. Data Description

As we propose a supplier selection approach combining criteria analysis and performance prediction, the evaluation phase requires the use of a dataset containing supplier information and supply operation history. To this end, we have chosen to use publicly available data to facilitate the evaluation of this approach, and to provide researchers with a basis for comparison using the Medicare & Medicaid Services (CMS) [33]. The website gives direct access to different data released by CMS. The datasets used for this study included information concerning durable medical Equipment and supplies with the supplier's information (payments, usage, submitted charges, beneficiary demographic...). This dataset is built on information gathered from CMS administrative during the period 2015-2020, whose dataset of each year exceeds 786040 elements.

B. Training Setups and Evaluation Metrics for CRNN

For the training of the CRNN, the Adam optimizer [34] is used, with a preliminary learning rate of 0.001. Every two epochs, this rate was reduced by a factor of 0.95, and the batch size was set to 32. The model was trained for 60 epochs in the whole experiment. We evaluated the performance of the CRNN model using the main evaluation metrics: the Mean absolute error (MAE), the mean absolute percentage error (MAPE), the Mean Squared Error (MSE), and the root mean square error (RMSE) [35]. For the CRNN modeling, we organized the training into a period sequence and feedback the sequence into the CRNN network constituted of various connected units, as explained above, to accomplish the current training model. Then, for the model optimization, the CRNN was trained to compute the values of the predicted variables at the set time.

The collected dataset has a total of 74,588 instances. These examples were randomly separated into three sets: a training set with 70% of the instances, a validation set with 20% of the instances, and a test set with 10% of the instances.

C. AHP Weightage

As stated previously, we applied AHP to calculate the weights and ranks of the various selection criteria. In the case of supplier selection, each level requires to be weighted to rate this large matrix. In this study, firstly, the weights of level 1 of each criterion are established and reviewed to determine the importance of each criterion. After that, the weight calculation steps of AHP are followed.

1) *Calculate the Weight of the selection criteria:* As supplier selection is paramount in manufacturing, this study presented a framework for analyzing its data, regardless of the size of the company, small, large, or medium. Manufacturing companies generate a large scale of diversified business processes. This is more convenient because they have a history

of transactions in addition to more recent data, with a strong experience of experts in the sector, which can ease the implementation of this approach. The chosen list of criteria in conjunction with the sub-criteria for each dimension was identified from the literature analysis, concerning the opinions of industry experts to ensure compatibility between the theoretical study and the practical aspects of supplier selection. The used sources offer a huge amount of data to study the previous records of the suppliers, which helped to confirm the list of criteria and sub-criteria.

At the preliminary stage, the criteria used were analyzed to recognize the most applicable criteria for the supplier selection process. Initially, there were ten criteria and 30 sub-criteria. Then preliminary discussions conducted with industrial experts were intended to gain a professional opinion about the criteria list. Then all these data were arranged and examined systematically.

The ranking results demonstrate that the most significant criteria that should be well studied while selecting suppliers for a specific product or service are quality and delivery of the suppliers followed by technological advances, performance improvement, and long-term relationship, which gained priority weightage of 0,462, 0,434, 0,359, 0,281 and 0,272 respectively the ranking weights. Based on the judgments given by the expert decision-makers, these criteria remain the most significant aspects that should be respected within a supplier selection process. According to these findings, it can be concluded that information sharing, subjective risks, intangible, cost-effective, and objective risks of the supplier gained relatively low priority weightings. When analyzing the priority weights for sub-criteria price appropriateness of the supplier is the most important criterion for them.

2) *Suppliers ranking*: The use of AHP to prioritize vital factors in manufacturing organizations may produce different importance values given to the specific requirements of each company. Moreover, these priorities may adjust regarding internal and external aspects, which can impact manufacturing operations.

TABLE I. PERFORMANCE RESULTS OF THE COMPONENTS OF PROPOSED METHOD

Method	AHP	CRNN	Proposed AHP_CRNN
Accuracy	90, 36 %	92,07%	95,96%
MAE	0.00554	0,05048	0,0771
MAPE	0,725782	1,004297	1,386251
MSE	0,00000602	0,00293	0,00262899
RMSE	0,00245	0,01711	0,04127

Table I demonstrates the results of supplier selection. Characteristically, the selection process ends once a supplier is chosen. However, other difficulties can occur regarding its performance and dedication, so it is quite important to analyze these aspects to avoid any potential risk that could affect the smooth running of manufacturing operations. Consequently, our study offers the possibility of having an optimized list of

suppliers to select the most efficient one that will meet the needs effectively and continuously, while ensuring the best gains and stability of manufacturing activities. Here best highest final values reveal that (Supplier_5), (Supplier_4), and (Supplier_2) are the most suitable suppliers for this supplier selection case, with the final values (2,031), (1,964), and (1,855) respectively.

D. Prediction of Supplier Performance

To quantify and assess the performance of the proposed method, the evaluation results of the AHP, CRNN, and the presented method. It can be noticed from the statistics in the table that the results of the three methods are all good with MAE < 0.1, MAPE < 1.5, MSE < 0.005, and RMSE rate < 0.5. The effects of using AHP and CRNN helped the proposed method to gain better results compared with traditional AHP and CRNN networks. The proposed strategic method proved higher prediction accuracy (95, 96%) with stronger generalization capability, and better operability, which shows that the AHP-CRNN proposed in this study is more appropriate for the supplier selection process in manufacturing systems.

While the first step, AHP, provided a methodological selection of the suppliers, the records of the best suppliers were captured and analyzed to reveal their performance. With all the completed preparations using AHP, the CRNN model computed iteratively the data, which contains the transaction history of the period 2015-2020 of the three suppliers, and provided the analytical results, as shown in Table II.

The anticipated values of providers exhibit significant consistency; however the projected values of some categories diverge considerably from prior assessments.

TABLE II. PREDICTION RESULTS OF THE PERFORMANCE OF SUPPLIERS REGARDING THE BEST-RANKED CRITERIA

Quality satisfaction								
Observed							Predicted	
Year	2015	2016	2017	2018	2019	2020	2025	2030
Supplier_2	29,5 0 %	44,3 2 %	51,2 5 %	57,5 0 %	63,5 2 %	69,3 2 %	71,3 9 %	76,4 1 %
Supplier_4	41,3 4 %	56,1 6 %	63,0 9 %	69,3 4 %	75,3 6 %	59,4 8 %	65,5 5 %	73,5 7 %
Supplier_5	53,1 8 %	68,0 0 %	74,9 3 %	81,1 8 %	87,2 0 %	71,3 2 %	77,3 9 %	85,4 1 %
Delivery transactions								
Observed							Predicted	
Year	2015	2016	2017	2018	2019	2020	2025	2030
Supplier_2	1230 0	1595 0	2507 1	3897 8	4532 8	5321 8	6730 9	9368 4
Supplier_4	1319 1	1684 1	2596 2	3986 9	4621 9	5410 9	6820 0	9457 5
Supplier_5	1408 2	1773 2	2685 3	4076 0	4711 0	5500 0	6909 1	9546 6
Technological advances								
Observed							Predicted	
Year	2015	2016	2017	2018	2019	2020	2025	2030

Supplier_2	33,5 6 %	55,2 3 %	61,9 8 %	68,7 8 %	70,6 2 %	76,5 5 %	80,3 2 %	85,6 9 %
Supplier_4	34,5 1 %	56,1 8 %	62,9 3 %	69,7 3 %	71,5 7 %	77,5 %	81,2 7 %	86,6 4 %
Supplier_5	34,2 8 %	55,9 5 %	62,7 %	69,5 %	71,3 4 %	77,2 7 %	81,0 4 %	86,4 1 %
Performance improvement								
Observed						Predicted		
Year	2015	2016	2017	2018	2019 %	2020	2025	2030
Supplier_2	34,9 2 %	56,5 9 %	63,3 4 %	70,1 4 %	71,9 8 %	77,9 1 %	81,6 8 %	87,0 5 %
Supplier_4	42,8 4 %	64,5 1 %	71,2 6 %	78,0 6 %	79,9 %	85,8 3 %	89,6 %	94,9 7 %
Supplier_5	40,1 7 %	61,8 4 %	68,5 9 %	75,3 9 %	77,2 3 %	83,1 6 %	86,9 3 %	92,3 %
Long-term relationship								
Observed						Predicted		
Year	2015	2016	2017	2018	2019 %	2020	2025	2030
Supplier_2	52,1 7 %	63,8 4 %	70,5 9 %	77,3 9 %	75,2 3 %	81,1 6 %	82,9 3 %	88,3 %
Supplier_4	60,0 9 %	71,7 6 %	78,5 1 %	85,3 1 %	83,1 5 %	89,0 8 %	90,8 5 %	96,2 2 %
Supplier_5	61,4 2 %	73,0 9 %	79,8 4 %	78,6 4 %	80,4 8 %	88,4 1 %	89,1 8 %	94,5 5 %

The analyzed data on long-term relationships clearly indicates that although Supplier_5 exhibited the best results from 2015 to 2017, there was a decline in performance in this criterion from 2018 to 2022, resulting in Supplier_4 outperforming Supplier_5, even in projected values. Supplier_4 marginally exceeded Supplier_5 in technology advancements and performance enhancement. Furthermore, the calculated performance metrics of the lower-ranked criterion exhibited varying levels of supplier performance. Particularly after 2018, when global economic problems emerged, affecting inflation rates and fluctuations in the international market following the coronavirus health crisis in 2020. The discrepancies in supplier performance underscore the significance of all selection criteria, not alone those identified by the AHP technique, to mitigate unanticipated factors that may disrupt production processes.

TABLE III. PREDICTION RESULTS OF THE PERFORMANCE OF SUPPLIERS CONSIDERING THE SUB-CRITERIA

Dimensions	Criteria	Detailed sub-criteria	Supplier_2	Supplier_4	Supplier_5
Capacity (C)	Cost-effective (C1)	Reduced cost/price of a product (C11)	44,67%	52,07%	49,55%
		Financial competence (C12)	57,64%	73,45%	58,52%
	Delivery (C2)	Available production (C21)	49,16%	52,23%	63,79%
		Delivery satisfaction (C22)	58,89%	63,96%	83,42%

Intangible (C3)	Performance history (C31)	43,29%	51,44%	73,01%	
	Responsiveness and situation in the industry (C32)	38,10%	77,65%	65,83%	
	Technological advances (C4)	Design (C41)	42,06%	50,48%	46,12%
		Quantity of patents applying (C42)	28,06%	56,20%	44,32%
Relative shares (C43)		24,04%	52,19%	48,35%	
Quality (C5)	R&D expenses input intensity (C44)	21,04%	67,79%	62,57%	
	Reliability of product (C51)	59,74%	76,89%	82,97%	
	Specific characteristics of remaining products (C52)	49,78%	68,92%	72,86%	
	Quality of products (C53)	43,06%	61,21%	59,35%	
Willingness (W)	Information sharing (W1)	Honest and regular communications (W11)	-	-	-
		Relationship proximity (W12)	-	-	-
	Long-term relationship (W2)	Dedication to quality (W21)	-	-	-
		Long-term commitment (W22)	21,26%	59,14%	54,99%
		Mutual honesty and respect (W23)	69,08%	77,06%	68,99%
	Performance improvement (W3)	Commitment to permanent development in products and processes (W31)	31,15%	56,65%	61,80%
Effort in supporting "just-in-time" standards (W32)		54,16%	57,16%	84,69%	
Risk of supply (R)	Objective risks (R1)	Geographical closeness (R11)	48,09%	73,19%	76,87%
		Bankruptcy (R12)	57,78%	86,07%	83,40%
		Strikes, natural disasters, pandemics (R13)	41,19%	81,30%	67,91%

Subjective risks (R2)	Transportation disruptions (R14)	40,96%	62,26%	53,54%
	Fluctuations in the market price of raw materials (R15)	44,24%	50,94%	70,73%
	Reputation (R21)	35,99%	71,10%	70,64%
	Organizational management (R22)	70,95%	80,94%	41,98%
	Social responsibility (R23)	59,89%	84,68%	56,99%
	Political and regulatory environment (R24)	53,92%	71,27%	64,12%
	Market conditions (R25)	40,97%	56,57%	56,09%
Global performance		46,84%	66,93%	66,03%

The list derived using the AHP approach identifies the top three suppliers: Supplier_5, Supplier_4, and Supplier_2. Nevertheless, the performance analysis of each supplier over the years indicates that Supplier_4 is a viable contender to Supplier_5. We used the AHP phase as input for the CRNN rather than the whole list of vendors. The use of AHP enabled us to generate a concise list, facilitating the CRNN's emphasis on the specifics of each supplier's performance according to the selection criteria. In the prior assessments, we only used the selection criteria from Table III, without considering the influence of the sub-criteria on supplier selection. To provide a fair comparison among the suppliers, the subsequent test involves evaluating the performance of each supplier based on the selection sub-criteria outlined in Table III. The performance

prediction findings, based on the selection sub-criteria, indicated that Supplier_4 outperforms Supplier_5.

This study's findings and existing literature demonstrate that using AHP enables manufacturing businesses to make supplier selection decisions based in methodical and objective assessments of available alternatives. This may mitigate the risk of bad decision-making and assure the selection of the appropriate supplier to fulfill the organization's objectives and specifications. Generally, selecting the appropriate provider to guarantee prompt delivery of superior quality.

Choosing appropriate items or services is crucial, since picking the incorrect option may result in many complications, such delivery delays, substandard quality, or even legal ramifications. To mitigate these possible issues, it is essential to adopt a comprehensive methodology for supplier selection that encompasses all relevant criteria and sub-criteria. By evaluating the suggested selection criteria with other pertinent organizational characteristics, decision-makers may mitigate the risk of supplier-related issues and assure the selection of an appropriate partner for their requirements. We have used deep learning to analyze and forecast the performance of the selected providers in order to mitigate risks. The AHP-CRNN methodology facilitates enhanced analysis to get increased revenues via the selection of the appropriate provider.

E. Comparison with Former Methods

The AHP-CRNN model was reviewed from multiple angles in the preceding sections. To properly demonstrate the operational effectiveness of AHP-CRNN, we chose four extensively proposed and used techniques (RNN, RDNN, CRNN, and LSTM). Based on the relevant published works, we retrieved the corresponding architectures and parameters of the abovementioned methodologies for this comparative analysis. As stated in the preceding sections, the experimental setting for the comparison maintained a consistent unified strategy. This included using the same database and keeping the percentages for the training, validation, and test datasets the same.

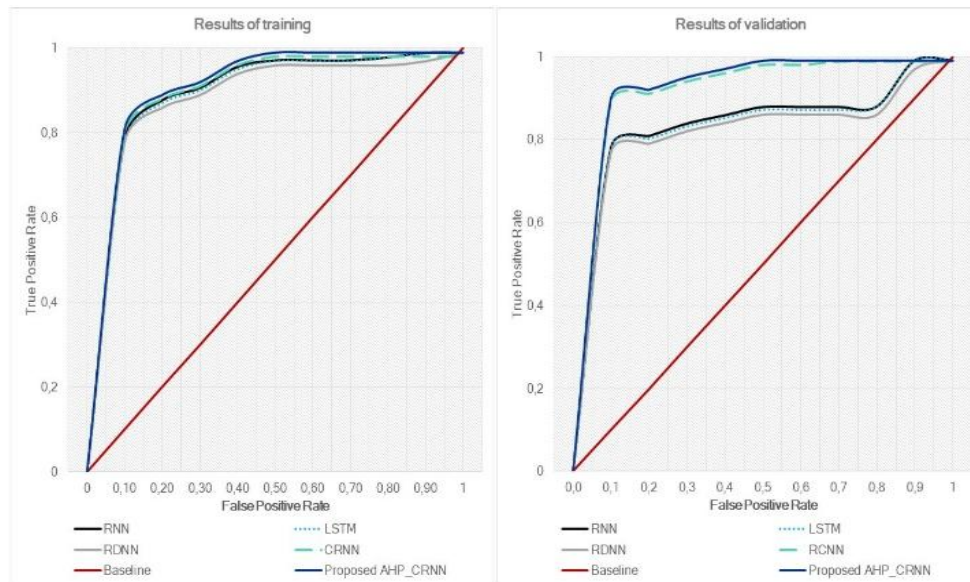


Fig. 4. The overall supplier selection assessment compared to other models in training and validation processes.

Because of the large amount of experimentation data, it is not possible to offer detailed convergence accuracy metrics for all tested approaches. As a result, in this section, we will instead provide statistical rankings. Fig. 4 depicts an overview of true positive (TP) and false positive (FP) rates for the training and validation sets, allowing for a thorough evaluation of the proposed supplier selection technique.

It allowed us to understand the model's ability to correctly identify positive examples and avoid false positives. As comparing the TP and FP rates of the training and validation sets is an important aspect of deep learning evaluation and tuning, the results show a massive improvement while using the AHP-CRNN model.

The proposed AHP-CRNN-based method for supplier selection employs a strategic process. We used the AHP phase to select a list of potential suppliers, which is used as the input of the CRNN model. The strategic AHP-CRNN-based approach strengthens the multi-objective analysis in the process of supplier selection. The CRNN employs CNN layers for feature extractions and RNN layers for the temporal dependencies assessment.

The proposed method was proven performant compared to traditional RNN [36], LSTM [31], RDNN [28], and CRNN [37]. To further compare and quantify the performance of the proposed AHP-CRNN strategy, the evaluation results of the literature models are displayed in Table IV.

TABLE IV. PERFORMANCE COMPARISON OF DIFFERENT LITERATURE MODELS

Method	Accuracy	MAE	MAPE	MSE	RMSE
RNN	89,99 %	0,009 28	0,9630 97	0,000500 7	0,01588 07
LSTM	92,56 %	0,018 38	0,9721 97	0,00072	0,0161
CRNN	92,07 %	0,050 48	1,0042 97	0,00293	0,01711
RDNN	91,73 %	0,059 98	1,0137 97	0,001505 99	0,05004 7
Proposed AHP_CRNN	95,96%	0,077 1	1,3862 51	0,002628 99	0,05127

The findings show that the LSTM surpasses the traditional RNN. That can be explained by the fact that LSTM networks can store long-term dependencies better than traditional RNNs. In traditional RNNs, the information from long-term dependencies can easily be forgotten or lost as it moves through the network. However, LSTMs have an internal memory cell that can store information for a longer period, allowing them to better capture long-term dependencies. However, traditional RNNs are often computationally simpler and more efficient than LSTMs, which can be more complex and computationally demanding. Applications combining RNNs with other types of neural networks, such as CNN or DNN, showed improved model performance. As the use of RNN with other networks makes it possible to address multiple tasks simultaneously or tackle more complex manufacturing data, the proposed method is based on a CRNN to have to ability to handle structured as well as unstructured data, ensure better generalization to new data, and reduce the overfitting.

F. Discussion

Comprehending these distinctions is crucial for efficient supplier selection. By analyzing historical data, decision-makers may identify suppliers who consistently perform effectively, even under challenging circumstances. This mitigates risks and ensures supply chain continuity. Moreover, analyzing supplier performance longitudinally allows for the identification of suppliers who consistently improve their performance. This information is crucial when considering long-term partnerships and developmental potential.

The proposed method facilitates the benchmarking of suppliers, highlighting top performance and identifying those requiring development. This data-driven approach enables decision-makers to make objective and informed choices, leading to cost reductions and enhanced efficiency. Moreover, evaluating supplier performance over time enables the optimization of your supply base. Organizations may establish robust relationships with reliable suppliers by acknowledging consistent performance, leading to enhanced negotiating leverage and favorable conditions. Evaluating supplier performance over time is essential for supplier management, since it offers insights into the consistency and dependability of providers.

The study outlined in the article examines the difficulties encountered by industrial units in a competitive and worldwide market, emphasizing the need of optimizing supply chains and choosing appropriate suppliers for success. The research underscores the necessity for enterprises to provide superior products/services at competitive pricing more swiftly than their rivals. This requires a rigorous supplier selection procedure to ensure supply chain integrity, preserve profit margins, and meet customer satisfaction.

The AHP-CRNN strategy in supplier selection offers several practical advantages, such as cost reduction, risk alleviation, quality enhancement, ethical procurement, and strengthened supplier relationships. Through the methodical assessment and selection of suppliers using both quantitative and qualitative criteria, firms may enhance their supply chains, secure a competitive advantage, and establish a robust and sustainable business environment.

The AHP-CRNN technique significantly influences supplier selection decisions. A primary advantage is the capacity to make well-informed supplier selection judgments. By methodically assessing prospective suppliers against several factors, including cost, quality, and delivery time, firms get an extensive understanding of their alternatives. This allows them to choose providers who fulfill urgent requirements while also aligning with long-term strategic objectives. In a more competitive business landscape, educated decision-making may profoundly influence a company's performance and competitiveness.

- Financial Implications and Efficiency: The AHP-CRNN methodology yields significant cost savings. By systematically evaluating suppliers, firms may discern those providing the most advantageous terms and pricing. This may result in substantial cost reductions,

an essential element in sustaining profitability. Furthermore, choosing suppliers that can adhere to stringent delivery timelines and adjust to fluctuating circumstances improves supply chain efficiency. Minimized supply chain interruptions and enhanced delivery times may result in decreased operating expenses and heightened customer satisfaction.

Quality assurance and risk management are essential in the selection of suppliers. The AHP-CRNN methodology facilitates the identification of suppliers with a demonstrated history of providing high-quality goods or services. Consistently choosing such suppliers guarantees the preservation of high-quality standards and mitigates the likelihood of product recalls or quality-related problems. Moreover, effective supplier selection is essential for risk minimization. Companies may choose suppliers recognized for their resilience and adaptability to unanticipated obstacles, thereby mitigating supply chain risks.

The AHP-CRNN methodology fosters a culture of ongoing improvement and data-driven decision-making. Organizations may evaluate previous supplier performance data and trends to perpetually enhance their selection criteria. This iterative procedure results in improved supplier selection over time. Furthermore, the methodology cultivates a data-driven culture throughout the firm, transcending supplier selection. It advocates for decision-makers to use data and analytics for informed decision-making across diverse business functions.

V. CONCLUSION

The suggested method of supplier selection using the Analytic Hierarchy Process (AHP) and Convolutional Recurrent Neural Network (CRNN) has significant consequences. Initially, using AHP provides a structured framework for systematically and openly evaluating and contrasting various criteria. The suggested technique enhances the long-term sustainability of manufacturing operations by ensuring efficient supplier selection. Fostering robust connections with suppliers and achieving favorable outcomes are essential for the seamless operation of production processes. The systematic strategies obtained from the AHP-CRNN approach facilitate the enhancement of production systems, guaranteeing the sustained availability of reliable and high-performing suppliers. The applicability of the suggested technique has been shown via several experimental experiments, highlighting its efficacy in supplier selection for sustainable manufacturing systems. This illustrates its potential for practical use. The study indicates that the intelligent judgment methodology may be improved to better the selection criteria for complex manufacturing systems, suggesting that the suggested method may be expanded and modified in future research.

REFERENCES

- [1] Schlemitz A, Mezhuyev V. 2024. Approaches for data collection and process standardization in smart manufacturing: Systematic literature review, *Journal of Industrial Information Integration*, Volume 38,2024,100578, <https://doi.org/10.1016/j.jii.2024.100578>.
- [2] J. Huang et al., "Deep Reinforcement Learning-Based Dynamic Reconfiguration Planning for Digital Twin-Driven Smart Manufacturing Systems With Reconfigurable Machine Tools," in *IEEE Transactions on Industrial Informatics*, vol. 20, no. 11, pp. 13135-13146, Nov. 2024, doi: 10.1109/TII.2024.3431095.
- [3] Sun, L., He, H., Yue, C. et al. Unleashing Competitive Edge in the Digital Era: Exploring Information Interaction Capabilities of Emerging Smart Manufacturing Enterprises. *J Knowl Econ* 15, 10853–10897 (2024). <https://doi.org/10.1007/s13132-023-01545-w>
- [4] Sahoo, S. K., Goswami, S. S., & Halder, R. (2024). Supplier Selection in the Age of Industry 4.0: A Review on MCDM Applications and Trends. *Decision Making Advances*, 2(1), 32–47. <https://doi.org/10.31181/dma21202420>
- [5] Sheykhizadeh, M., Ghasemi, R., Vandchali, H.R. et al. A hybrid decision-making framework for a supplier selection problem based on lean, agile, resilience, and green criteria: a case study of a pharmaceutical industry. *Environ Dev Sustain* 26, 30969–30996 (2024). <https://doi.org/10.1007/s10668-023-04135-7>
- [6] Bai, C., Zhu, Q. and Sarkis, J. (2022) 'Circular economy and circularity supplier selection: a fuzzy group decision approach', *International Journal of Production Research*, 62(7), pp. 2307–2330. doi: 10.1080/00207543.2022.2037779.
- [7] Pamucar, D., Ulutaş, A., Topal, A., Karamaşa, Ç., & Ecer, F. (2024). Fermatean fuzzy framework based on preference selection index and combined compromise solution methods for green supplier selection in textile industry. *International Journal of Systems Science: Operations & Logistics*, 11(1). <https://doi.org/10.1080/23302674.2024.2319786>
- [8] Acerbi, F., & Taisch, M. (2020). Information flows supporting circular economy adoption in the manufacturing sector. In Ed., L. B, Department of Management, Economics and Industrial Engineering pp. 703–710. Springer. Available at. https://doi.org/10.1007/978-3-030-57997-5_81
- [9] Emrouznejad, A. and Marra, M. (2017), "The state of the art development of AHP (1979–2017): a literature review with a social network analysis", *International Journal of Production Research*, Vol. 55 No. 22, pp. 6653-6675.
- [10] Liu, Y., Eckert, C.M. and Earl, C. (2020), "A review of fuzzy AHP methods for decision-making with subjective judgments", *Expert Systems with Applications*, Vol. 113738
- [11] Belotti Pedroso, C., Tate, W. L., Lago da Silva, A., & Ribeiro Carpinetti, L. C. (2021). Supplier development adoption: A conceptual model for triple bottom line (TBL) outcomes. *Journal of Cleaner Production*, 314(June), 127886. Available at. <https://doi.org/10.1016/j.jclepro.2021.127886>
- [12] Wang, C.-N., Pan, C.-F., Tinh Nguyen, V., & Tam Husain, S. (2022). Sustainable supplier selection model in supply chains during the COVID-19 pandemic. *Computers, Materials and Continua*, 70(2), 3005–3019. Available at. <https://doi.org/10.32604/cmc.2022.020206>
- [13] Mei, Y., Ye, J., & Zeng, Z. (2016). Entropy-weighted ANP fuzzy comprehensive evaluation of interim product production schemes in one-of-a-kind production. *Computers & Industrial Engineering*, 100, 144–152. Available at. <https://doi.org/10.1016/J.CIE.2016.08.016>
- [14] Kaur, H., Singh, S. P., Garza-Reyes, J. A., & Mishra, N. (2020). Sustainable stochastic production and procurement problem for resilient supply chain. *Computers & Industrial Engineering*, 139,105560. Available at. <https://doi.org/10.1016/j.cie.2018.12.007>
- [15] Zeydan, M., Çolpan, C., & Çobanoğlu, C. (2011). A combined methodology for supplier selection and performance evaluation. *Expert Systems with Applications*, 38(3), 2741–2751. Available at. <https://doi.org/10.1016/j.eswa.2010.08.064>
- [16] Shiri, I., AmirMozafari Sabet, K., Arabi, H. et al. (2021) Standard SPECT myocardial perfusion estimation from half-time acquisitions using deep convolutional residual neural networks. *J. Nucl. Cardiol.* 28, 2761–2779. <https://doi.org/10.1007/s12350-020-02119-y>
- [17] Yuan, Y.; Shao, C.; Cao, Z.; He, Z.; Zhu, C.; Wang, Y.; Jang, V. (2020) Bus Dynamic Travel Time Prediction: Using a Deep Feature Extraction Framework Based on RNN and DNN. *Electronics* 2020, 9, 1876. <https://doi.org/10.3390/electronics9111876>
- [18] Chien, CF. et al. (2020) Deep reinforcement learning for selecting demand forecast models to empower Industry 3.5 and an empirical study for a semiconductor component distributor, *International Journal of Production Research*, 58:9, 2784-2804, DOI: 10.1080/00207543.2020.1733125

- [19] Bera, S., Shrivastava, VK(2020) Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification, *International Journal of Remote Sensing*, 41:7, 2664-2683, DOI: 10.1080/01431161.2019.1694725
- [20] Acerbi, F., Rocca, R., et al. (2023) Enhancing the cosmetics industry sustainability through a renewed sustainable supplier selection model, *Production & Manufacturing Research*, 11:1, DOI: 10.1080/21693277.2022.2161021
- [21] Chai, J., Ngai, EWT. (2020) Decision-making techniques in supplier selection: Recent accomplishments and what lies ahead. *Expert Systems with Applications*. Volume 140, 2020, 112903, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2019.112903>
- [22] Stević, Z.(2020) Sustainable supplier selection in healthcare industries using a new MCDM method: Measurement of alternatives and ranking according to COmpromise solution (MARCOS), *Computers & Industrial Engineering*, Volume 140, 2020, 106231, <https://doi.org/10.1016/j.cie.2019.106231>.
- [23] Memari, A., Dargi, A., Akbari Jokar, M. R., Ahmad, R., & Abdul Rahim, A. R. (2019). Sustainable supplier selection: A multi-criteria intuitionistic fuzzy TOPSIS method. *Journal of Manufacturing Systems*, 50, 9–24. Available at. <https://doi.org/10.1016/j.jmsy.2018.11.002>
- [24] Zhang, J.; Yang, D.; Li, Q.; Lev, B.; Ma, Y. (2021) Research on Sustainable Supplier Selection Based on the Rough DEMATEL and FVIKOR Methods. *Sustainability* 2021, 13, 88. <https://doi.org/10.3390/su13010088>
- [25] Nair, R.K., et al. "The impact of formal supplier selection process on supplier performance and supplier relationship," *Journal of Supply Chain Management*, vol. 56, no. 4, 2020.
- [26] Mani, V., Agrawal, R., Sharma, V. (2014) Supplier selection using social sustainability: AHP based approach in India. *International Strategic Management Review*, Volume 2, Issue 2, Pages 98-112
- [27] Jessin, T.A., Rajeev, A. and Rajesh, R. (2023), "Supplier selection framework to evade pseudo-resilience and to achieve sustainability in supply chains", *International Journal of Emerging Markets*, Vol. 18 No. 6, pp. 1425-1452. <https://doi.org/10.1108/IJOEM-11-2021-1704>
- [28] Yuan, Y.; Shao, C.; Cao, Z.; He, Z.; Zhu, C.; Wang, Y.; Jang, V. (2020) Bus Dynamic Travel Time Prediction: Using a Deep Feature Extraction Framework Based on RNN and DNN. *Electronics* 2020, 9, 1876. <https://doi.org/10.3390/electronics9111876>
- [29] Abdulla, A.; Baryannis, G.; Badi, I. Weighting the Key Features Affecting Supplier Selection using Machine Learning Techniques. *Preprints* 2019, 2019120154 (doi: 10.20944/preprints201912.0154.v1).
- [30] Fernandez-Vazquez, S., Rosillo, R., de la Fuente, D. and Puente, J. (2022), Blockchain in sustainable supply chain management: an application of the analytical hierarchical process (AHP) methodology, *Business Process Management Journal*, Vol. 28 No. 5/6, pp. 1277-1300. <https://doi.org/10.1108/BPMJ-11-2021-0750>
- [31] Cahuantzi, R., Chen, X., Güttel, S. (2021) A comparison of LSTM and GRU networks for learning symbolic sequences. <https://doi.org/10.48550/arXiv.2107.02248>
- [32] Medicare & Medicaid Services (CMS). (2022) <https://data.cms.gov/provider-data/>
- [33] Feng, W., Zhu, Q., Zhuang, J., et al. (2019) An expert recommendation algorithm based on Pearson correlation coefficient and FP-growth. *Cluster Comput* 22 (Suppl 3), 7401–7412 (2019). <https://doi.org/10.1007/s10586-017-1576-y>
- [34] Bera, S., Shrivastava, VK(2020) Analysis of various optimizers on deep convolutional neural network model in the application of hyperspectral remote sensing image classification, *International Journal of Remote Sensing*, 41:7, 2664-2683, DOI: 10.1080/01431161.2019.1694725
- [35] Chicco D, Warrens MJ, Jurman G. (2021) The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE, and RMSE in regression analysis evaluation. *PeerJ Computer Science* 7:e623 <https://doi.org/10.7717/peerj-cs.623>
- [36] Guo, K., et al., Optimized Graph Convolution Recurrent Neural Network for Traffic Prediction, in *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 1138-1149, Feb. 2021, doi: 10.1109/TITS.2019.2963722.
- [37] Sheng, Z., Wang, H., Chen, G. et al. Convolutional residual network to short-term load forecasting. *Appl Intell* 51, 2485–2499 (2021). <https://doi.org/10.1007/s10489-020-01932-9>

Understanding Art Deeply: Sentiment Analysis of Facial Expressions of Graphic Arts Using Deep Learning

Fei Wang¹

Hubei University of Technology, Wuhan City, Hubei Province, 430068, China¹

Abstract—Art serves as a profound medium for humans to express and present their thoughts, emotions, and experiences in aesthetically and captivating means. It is like a universal language transcending the limitations of language enabling communication of complex ideas and feelings. Artificial Intelligence (AI) based data analytics are being applied for research domains such as sentiment analysis in which usually text data is analyzed for opinion mining. In this research study, we take art work and apply deep learning (DL) algorithms to classify seven diverse facial expressions in graphics art. For empirical analysis, state of the art deep learning algorithms of Inceptionv3 and pre-trained model of ResNet have been applied on large dataset. Both models are considered revolutionary deep learning architecture allowing for the training of much deeper networks and thus enhancing model performance in various computer vision tasks such as image recognition and classification tasks. The comprehensive results analysis reveals that the proposed methods of ResNet and Inceptionv3 have achieved accuracy as high as 98% and 99% respectively as compared to existing approaches in the relevant field. This research contributes to the fields of sentiment analysis, computational visual art, and human-computer interaction by addressing the detection of seven diverse facial expressions in graphic art. Our approach enables enhanced understanding of user sentiments, offering significant implications for improving user engagement, emotional intelligence in AI-driven systems, and personalized experiences in digital platforms. This study bridges the gap between visual aesthetics and sentiment detection, providing novel insights into how graphic art influences and reflects human emotions by highlighting the efficacy of DL frameworks for real-time emotion detection applications in diverse fields such as human psychological assessment and behavior analysis.

Keywords—Artificial intelligence; deep learning; sentiment analysis; art detection; image processing; convolutional network

I. INTRODUCTION

Art is a broad term that can be described as the work or process undertaken by man in creating physical skills, objects, or musicals involving painting, sculpture and dancing etc. They are not only appraisals of culture and individual encounters but also as a channel or medium of communication which enforces feeling and thinking. Interdisciplinary methods are used in the analysis of art that many disciplines incorporate into their analysis the impact and role of artistic creations on and from the social contexts of world. For example, in expressions analysis of art, the concern is on the feelings invoked by art and how such feelings differ among the subgroups and cultures

[1]. In more practical terms, researchers use semi structured interviews and questionnaires with the audience together with quantitative tools like sentiment analysis to elicit responses that contribute to a nuanced understanding of the use of art as a means of creating human experiences and engagement with various social processes. There are various types of sentiment analysis including the binary classification of subjectivity analysis [2], the tertiary classification containing the sentiment valance finding [3], the multi-classification containing the emotion detection [4], and the aspect-oriented sentiment analysis [5] that targets feature level deep understanding. a means of communication that evokes emotions and provokes thought. The analysis of art spans multiple research areas, including psychology, sociology, and cultural studies, where scholars examine how artistic expressions influence and are influenced by societal contexts. For instance, expressions analysis in art focuses on understanding the emotional responses elicited by artworks and how these responses vary across different demographics and cultural backgrounds [6]. Researchers employ qualitative methods such as interviews and surveys alongside quantitative techniques like sentiment analysis to gauge audience reactions, ultimately contributing to a deeper understanding of the role of art in shaping human experience and social discourse [7].

Moreover, sentiment analysis is combined with other technologies like computer vision and IoT devices so that a business can monitor information about customers' emotions and actions in the physical spaces in real-time mode [8]. Sentiment analysis is an important area as it continues to develop, the complexity of the models for such analysis will increase making it easier to determine the right strategies when they are required in different fields. Analysis of sentiment in images has been a popular trend in recent years mainly in the context of affective content in images. This field discusses how certain images create certain feelings and this is very essential because in areas like social media analysis or advertising. The research in this direction started around 2010, where the first attempts were made to place pictures into positive or negative sets according to their characteristics. To interpret affect, there has been the use of methods like Convolutional Neural Networks (CNNs) to perform context analysis of images, in relation to emotional content through organizations like Flickr and Twitter. From the research done, texture and color co-occurrence histogram are critical in identifying the sentiment of an image [9] [10]. In addition, proposing a method for combining both text and image features should help improve

the effectiveness of sentiment classifiers and provide additional knowledge of the users' emotions conveyed through images posted on social media [11].

As for social art, sentiment analysis becomes a crucial method on how the public perceives the art pieces and other materials that are a part of culture. By observing images of artworks or social art initiatives posted on social media, emotions and reception of a given subject in the community can be quantified [12]. This approach can help in measuring the level of audience participation but can also enlighten artists and curators regarding prevailing mood trends within their viewers. When applied in this case, the deep learning models enable the analysis of subtle differences in sentiment beyond basic positive-negative quality assessments [13]. In addition, the differentiation of emotions that are related to concrete artworks will enable targeted addressing and, thus, improve the effectiveness of social art activities.

A. Research Contributions

In this study, our main contributions include:

- For graphic art and identification of seven emotions, data preprocessing and diverse deep learning algorithms have been applied.
- Highlighted the limitations of existing studies by filling research gaps in emotion detection using digital image processing and deep learning.
- Developed a robust emotion classification framework using a deep learning pre-trained models including ResNet-50 and Inception v3 model by modifying the fully connected layer to adapt to the specific emotion classification task.
- Achieved highest classification performance of 99% with inception v3 model as compared to baseline models such as VGG16 and DCNN, demonstrating state-of-the-art results.

For the rest of the paper, Section II reviews the existing studies in relevant literature, then Section III shares the proposed research methodology along with experimental set-up discussing datasets which are prepared and used for empirical analysis and performance metrics used for results comparison. Section IV presents the results and discussion sharing findings from this study. Before concluding the manuscript, Section V discusses the results in detail sharing comparative analysis of the proposed model with the existing approaches.

II. BACKGROUND

The sentiment analysis of images by employing deep learning techniques has received increased attention in the last few years due mainly to the large availability of computer powers and the growing availability of image data in the social media platforms. In this approach, deep learning techniques, including Convolutional Neural Networks CNNs, are used to learn useful feature representations of images from social media which are indicative of emotional sentiments. Table I defines the summary of existing studies for deeper analysis. Studies show that CNNs are capable to capture spatial hierarchies of images for the task of categorizing sentiments

into positive, negative or neutral [14]. Recent works have shown that using transfer learning methods including Inception-V3 it is possible to librarian the accurate classification of the sentiment analysis tasks without requiring large, labeled datasets [15]. This capability is particularly valuable where good quality labeled data is hard to come by or in short supply.

The combination of two modalities, i.e., using image analysis in conjunction with textual sentiment analysis, has enriched the field even more. Analyzing the pictures together with the related texts helps researchers get a deeper insight into the people's attitudes [16]. For example, the integration with captions or hashtags used in Big Five Personality traits analysis helps models consider extra context that enhances the sentiment prediction accuracy up to multiple factors [17]. Moreover, improvements in the deeper architecture of the Capsule Networks besides the convolutions with RNN and hybrid Deep Learning also demonstrate great performance in sentiment analysis [18]. These models prove enhanced performances in comprehending intricate nonverbal emotions described through graphics. There is still a problem in image sentiment analysis, especially in terms of the stability of human emotions and different perceptions of images across cultures. Due to its subjectivity, there are variations in modeling sentiments which need to be dealt with by having rich training set with various emotions and occurrence [19]. A revolution in the recent few years in deep learning has revolutionized the field of image sentiment analysis where the general expressions of emotions in the image are processed and understood [20].

Among the more significant trends, it is possible to distinguish the combination of CNNs and RNNs as these networks have been used to extract spatial and temporal data in images and their textual descriptions. CNNs are good at detailing the local features protruding on architectural diagrams that make them significant in accentuated sentiment classifying assignments where visualization features are dominant [21]. Current research has also shown that combining CNNs with LSTMs results in finer outcomes in the sentiment classification owing to combining pros of both structures [22].

The use of people's emotions in multimodal textual and visual platforms has been considered as an active research area in this context. Combining information retrieved from both images and associated captions will give more light to researchers to get to the core point of this subject, which in this case is the sentiments. For example, it is established that the interaction of CNNs for image processing with individualized LSTMs or transformers for sports sentiment analysis can help improve sentiment outlook than individual modality alone [23]. This approach is especially useful in settings such as social media where messages are occasionally accompanied by images which indicate the authors' emotional state. Thus, the current state of and future trends for image sentiment analysis based on deep learning methods are steadily developing. We find a vast scope towards improving the existing sentiment analysis performance and its utility in different domains by integrating them with the latest architectural models like CNNs, RNNs, and transformers with a combination of multi-modal data for the prediction of gender violence based on

sentiment analysis [24]. As the issues concerning data quality and ethical implications are solved as potential issues, it will be possible that both quantity and quality aspects of deep learning-based sentiment analysis will expand.

A. Limitations of Existing Studies

Many of the prior works in the field of emotion detection, especially in digital image processing and deep learning algorithm, often encounter several limitations. Most use simple models such as the VGG16 or DCNN, which while serving basic image categorization lack the ability to discern the patterns for capturing emotion features required for categorization. These models often face challenges with overfitting, especially when dealing with limited or imbalanced datasets, resulting in suboptimal generalization to unseen data.

Additionally, previous studies frequently lack comprehensive evaluations across diverse emotion categories or robust datasets, which limit their applicability to real-world scenarios. Computational inefficiency is another significant drawback, as some models require extensive computational resources but fail to deliver proportionally high performance. In addition, few advanced data augmentation techniques are used and the absence of an extensive focus on specific domains leads to failures to reach better accuracy and reliability. Such limitations justify the need for more sophisticated and flexible strategies, as addressed in this research study. The novelty of our work lies in achieving the highest results with an increased number of classes, enabling more comprehensive emotion and sentiment analysis from diverse facial expressions in graphic art, surpassing the limitations of previous studies with fewer emotion categories.

TABLE I. SUMMARY OF EXISTING STUDIES

Ref	year	Model	Dataset	Classes	Results
[14]	2018	CNN	Twitter images	4	90%
[15]	2023	CNN	famous CK+, FER2013, and JAFFE	3	95%
[16]	2022	BERT,CNN	MVSA-Multiple and T4SA	2	93%
[17]	2024	LSTM	MBTI	5	92%
[18]	2020	ConvNet-SVMBovW model, SVM	IMDb	5	91%
[19]	2024	Capsule with Deep CNN and Bi structured RNN	Twitter data	4	95%
[20]	2024	LSTM-BiLSTM,BCNN	MVSA	3	92%
[21]	2023	CNN	CK+, FER2013, and JAFFE	4	94%
[22]	2019	CNN,LSTM	IMDB,Google news dataset	4	92%

[23]	2021	CNN,LSTM,KNN	2018 world tweets	FIFA cup	2	92%
[24]	2022	LSTM-CNN+GloVe	Tweets dataset		3	93%

III. METHODS AND ARCHITECTURES

The two areas of artificial intelligence and deep learning have prompted notable improvement in understanding and discriminating demanding patterns in digital image. This study proposed the models that are embedded in recognizing colorless images depicting diverse artistic facial expressions. It encompasses high complication preprocessing, architecture, and large dataset to provide robust and accurate results. The architectures for the employed models are respectively shown in Fig. 1, which shows the basic workflow of any typical DL models with multi-tier structures of the architectures. Firstly, known as Dataset Collection, the image data pertinent to the process is accumulated. Then, Data Preprocessing is done to remove all irrelevant or duplicate data and adjust the data format for the model. The next step is Training with Models such as Inception v3 and ResNet50, which are deep convolutional neural network, which we extract features from images and learn some patterns from them. Fully connected layers follow the training process to combine the features which the model has learned with for the purpose of classification. It is then classified under different categories of different models, having considered the learned patterns. Moreover, the figure shows that Subtractive operations like Activation Function, Pooling, Flattening, Reduction of Overfitting and Optimizer are required to enhance the efficiency of the model and to avoid overfitting and thus more generalized results.

A. Dataset Preparation and Preprocessing

Dataset consists of 32,298 grayscale images illustrating facial expressions and depicted as sketches with lead pencils. It includes seven emotion classes: This means the emotions, which are depicted in the images can be categorized into seven classifications, which are being angry, disgust, fear, happy, sad, surprise, and a neutral category. The challenges of the artistic and grayscale pictures as well as size and variability of the dataset are countered well, thus allowing the use of deep learning models.

The input data also handles through some manipulations to improve its compatibility with the selected deep learning models and more importantly to ensure that the models undergo stable training by applying two major steps of resizing and normalization. By changing the dimensions of the images to 299 by 299 pixels because that is the expected input size for model. There will be occasional variability in light conditions therefore the image's data are normalized at a mean of [0.5, 0.5, 0.5] and standard deviation [0.5, 0.5, 0.5].

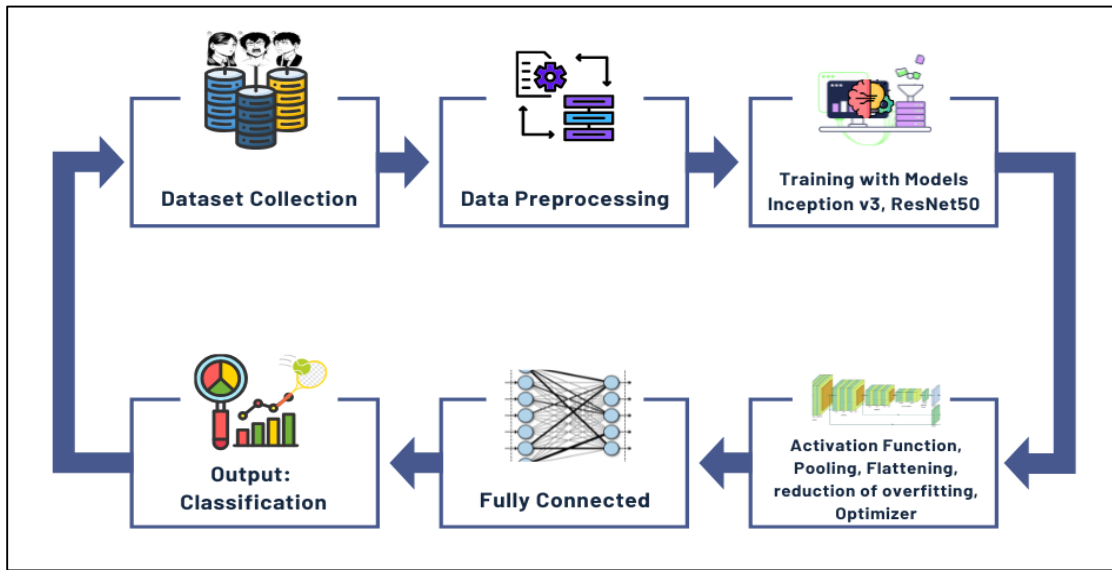


Fig. 1. Basic framework of any typical deep learning model.

These preprocessing steps assist in controlling fluctuations in the training curve using pixel values for normalization by following Eq. (1), thereby having an optimized training process with good features. Table II contains comprehensive details of all symbols that are used in equations for better understanding.

$$P_{normalized} = \frac{P-\mu}{\sigma} \quad (1)$$

B. Applied Deep Learning Models

In sentiment analysis method containing the feature extraction and the classification, both components are significant to the interpretation of the emotional context visually transferred. For example, in feature extraction, model captures the features such as facial expressions, patterns, or

textures as a representation of happy, sad, angry, and the like. Such features are embedded into the feature space of higher dimensions thus capturing relevant visual details. During this stage, these features extracted are then subjected to fully connected layers that try to map the image to sentiment categories. It makes it possible to capture slight changes in the expressions or other artistic features which are valuable to make robust sentiment categories even in highly diverse image sets.

1) *Inception v3 architecture*: Inception v3 is a deep learning architecture which optimizes computational speed and uses high effectiveness, as architecture shown in Fig. 2 for this study.

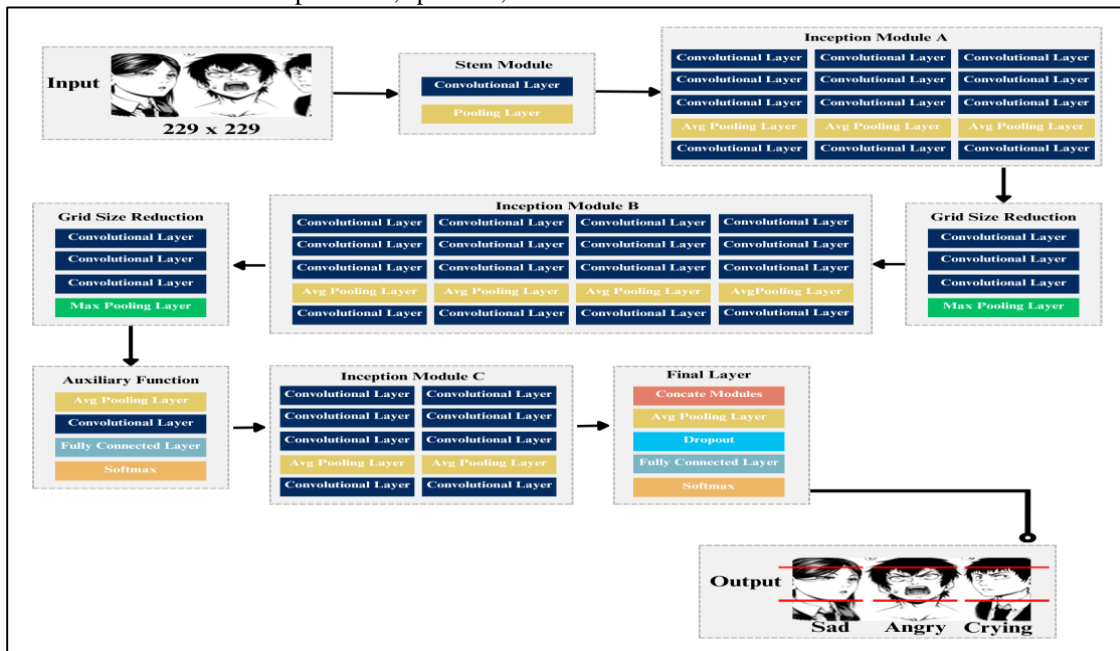


Fig. 2. Architecture of Inception v3 model.

This is done with the help of the inception module, which takes the input data and splits it through a series of channels every of which processes the data differently while trying to capture as many attributes of the input as $X \in \mathbb{R}^{H*W*D}$ where H is the height, W is the width and D is the depth showing number of channels. These paths are then added to produce a single output that can be represented as $O_{inception} \in \mathbb{R}^{H*W*D'}$; thus, the network can capture a diverse set of features at multiscale level. To minimize computational complexity, certain design strategies have been applied, such as the factorized convolution, where a complex convolution is replaced by a series of simpler steps, as well as the dimensionality reduction using 1x1 convolutions, computed as in Eq. (2).

$$O_{inception} = \text{concat}(\text{conv}_{1*1}(X), \text{conv}_{3*3}(X), \text{conv}_{5*5}(X), \text{conv}_{1*1(3*3)}(X)) \quad (2)$$

Where:

$$\begin{aligned} \text{conv}_{1*1}(X) &= W_1 * X + b_1, \text{ with } W_1 \in \mathbb{R}^{1*1*D*D_1} \text{ and } b_1 \in \mathbb{R}^{D_1} \\ \text{conv}_{3*3}(X) &= W_3 * X + b_3, \text{ with } W_3 \in \mathbb{R}^{3*3*D*D_3} \\ \text{conv}_{5*5}(X) &= W_5 * X + b_5, \text{ with } W_5 \in \mathbb{R}^{1*1*D*D_5} \end{aligned}$$

The auxiliary head is a technique of regularization in which gradients are added in the actual backpropagation processes during the training. The auxiliary classifier applies function over the featured space $\hat{y}_i^{auxiliary}$ produced by a certain layer of the network, computed as in Eq. (3).

$$L_{auxiliary} = -\sum_{i=1}^C y_i \log(\hat{y}_i^{auxiliary}) \quad (3)$$

Also, Inception v3 requires Global Average Pooling (GAP) to decrease the size of feature maps $F \in \mathbb{R}^{H*W*D}$ along with depth dimensions D for training model and enhancing generalization to each feature map f_a , computed as in Eq. (4).

$$GAP(F) = \frac{1}{H*W} \sum_{i=1}^H \sum_{j=1}^W F_{ij}^d \quad (4)$$

Finally, the output is classified using activation function layer combines both main classification and the auxiliary loss, computed as in Eq. (5), which indeed makes the model very effective for large scale image recognition.

$$L_{total} = L_{main} + \delta L_{auxiliary} \quad (5)$$

2) *ResNet 50 architecture*: ResNet-50 is a categorized deep convolutional network model resolving vanishing gradients issue in very deep learning networks, as working defined in Fig. 3.

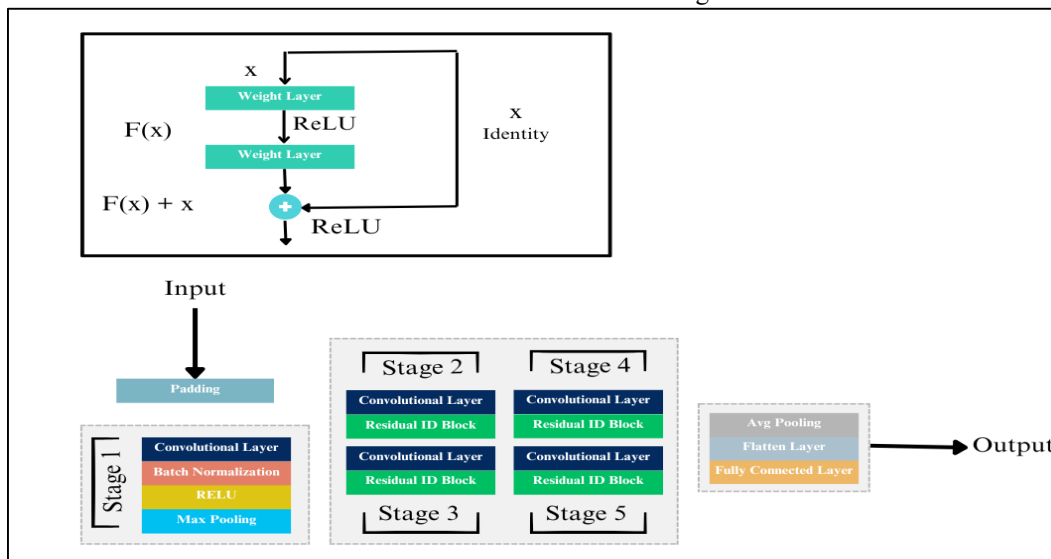


Fig. 3. Architecture of ResNet50 model.

With 50 layers and each layer of residual blocks. ResNet design is based on the concept of Residual block or skip connections that can bypass one or more layers of computations to provide the network with a method of training from scratch deeper networks that causes less degradation, at least theoretically. These skip connections are useful in reducing the vanishing gradient problem because they enable the gradients to flow directly through them using learning function \mathcal{F}_k which is calculated as in Eq. (6).

$$y_k = \mathcal{F}_k(x_k, \{W_{k,i}\}) + x_k \quad (6)$$

The ResNet-50 model has been designed mostly for relatively small image classification problems and uses batch

normalization and ReLU activation for enhanced results, computed as in Eq. (7).

$$z_k = \sigma(BN(\mathcal{F}_k(x_k, \{W_{k,i}\}) + x_k)) \quad (7)$$

As for the architecture, it is made up of the first convolutional layer, that is several stages of residual blocks where each block is made up of a 1x1 convolutional layer, a 3x3 convolutional layer and a final second 1x1 convolutional layer, defining the gradient flow using loss function \mathcal{L} , computed as in Eq. (8).

$$\frac{\partial \mathcal{L}}{\partial x_k} = \frac{\partial \mathcal{L}}{\partial y_k} + \frac{\partial \mathcal{L}}{\partial z_k} \cdot \frac{\partial z_k}{\partial x_k} \quad (8)$$

The architecture makes it easy to learn both low- and high-level features and that is the reason why this model is widely used in many computer vision tasks such as object detection and segmentation.

TABLE II. SYMBOLS DESCRIPTION OF APPLIED EQUATIONS

Symbols	Description
$P - \mu$	Mean values
ϑ	Standard deviation
P	Original pixel values
*	Convolution operation that are concate with final output
$y_i \in \mathbb{R}^C$	One hot-encoded true label vector for class i
$\hat{y}_i^{auxiliary} \in \mathbb{R}^C$	Predicted probability distribution from the auxiliary classifier
C	Number of output classes.
F_{ij}^d	Feature value at position (i, j) in the $d - th$ channel of the feature map
δ	Weight factor that balances the auxiliary loss relative to main loss
y_k	Output of the $k - th$ residual block
x_k	Input to the $k - th$ residual block
$\mathcal{F}_k(x_k, \{W_{k,i}\})$	Learned residual function with weight $W_{k,i}$
z_k	Output of the Batch Normalization (BN) and activation
σ	RELU activation function
$\frac{\partial \mathcal{L}}{\partial x_k}$	Gradient of loss with respect to input x_k
$\frac{\partial \mathcal{L}}{\partial y_k}$	Gradient with respect to output y_k
$\frac{\partial \mathcal{L}}{\partial z_k}$	Gradient of loss with respect to BN output z_k
$\frac{\partial z_k}{\partial x_k}$	BN output z_k depends on the input x_k .
TP and FP	True Positive and False Positive
TN and FN	False Positive and False Positive

C. Performance Measures

To fully assess the performance of models, standard assessment metrics are utilized including accuracy, precision, recall and F1-score. They include information on correct classified instances, and are useful in cases where classes are imbalanced, or misclassifications to different classes cost differently. Accuracy is the simplest of all the performance measurement metrics that give the percentage of correct prediction of instances to the entire instances. But it might not be that effective in handling those datasets with imbalanced classes. Precision becomes important due to this, together with recall. Precision is the measure of the accuracy of the positive predictions calculated as the proportion of true positives to the total positive predictions and the false ones, it is valuable in those application domains where false positives carry serious implications (e.g., medical diagnoses). Recall or Sensitivity is equal to the relation of true positive findings to the sum of true positives and false negative results, which highlights the ability of the model not to miss any relevant cases. F1-score defined

as the harmonic mean of precision and recall, is a valuable supplement to these two values, but most important when both measures are significant.

The training and validation accuracy and loss are important parameters to decide about the learning progress of a model. The accuracy of training measures the capability of the model on the training dataset, while the validation accuracy tests the model on how well it can perform on new dataset. The same about training loss that estimates the error during the learning process of the model on the training set, and validation loss that estimates the error on the validation dataset. Training loss should be decreasing while validation loss should be a plateau or on an increasing trend if there is not much data for training or training data is limited rather than showing a decreasing trend having a low value is best for a well-generalized model. Thus, Table III indicates the metrics employed to adjust the models depending on the context of evaluation, which will be highly beneficial for practitioners.

TABLE III. EVALUATION PERFORMANCE MEASURES

Sr. No	Metrics	Equation	Purpose
1	Accuracy	$\frac{TP+TN}{TF+FN+FP+TP}$	Measures overall correctness
2	Precision	$\frac{TP}{TP+FP}$	Focus on avoiding false alarms
3	Recall	$\frac{TP}{TP+FN}$	Emphasizes capturing all actual positives
4	F1-score	$\frac{2(Precision*Recall)}{Precision+Recall}$	Indicates overall performance balance

IV. RESULTS AND DISCUSSION

The findings on how facial emotions are classified using the deep learning pre-trained models, Inception v3 and ResNet50 model can be a useful guide on the efficiency of deep learning for facial emotions classification, as shown in table IV. The main aim of the study is to distinguish between emotions like ‘angry,’ ‘crying’, ‘embarrassed’, ‘happy’, ‘pleased,’ ‘sad,’ and ‘shock’ through facial expression, and the results of the study can be given multiple interpretations.

TABLE IV. RESULTS OF MODEL PERFORMANCE ACROSS ALL MEASURES

Model s	Training				Validation			
	Accur acy	Precis ion	Rec all	F1- Sco re	Accur acy	Precis ion	Rec all	F1- Sco re
ResNet50	98	98	98	98	67	66	67	65
Inception v3	99	99	99	99	76	78	76	76

A. Hyperparameter Settings

The configurations of the model are adjusted to improve its performance for the emotion classification task, as shown in Table V. An optimizer learning rate is used to adjust generalization of the pre-trained model while practicing on the dataset without causing much alteration of the learnt parameterization. The number of batches has been defined in the manner to optimize the computational resources and to maintain steady gradients during the training phase. This algorithm is used due to its adaptive learning rate for each

parameter what makes possible to obtain faster convergence with better generalization. Like the previous optimizers, this optimizer does not require specific parameter settings since it utilizes standard settings for its operation for efficient weight updates during the optimization process.

TABLE V. VALUES OF HYPERPARAMETERS

Parameters	Values
Learning Rate	0.0001
Batch Size	32
Optimizer	Adam
Adam Settings	Lr = 0.0001, beta1 = 0.9, beta2 = 0.999
Loss Function	Cross Entropy Loss
Epochs	30
Model Architecture	Inception v3, ResNet50

For the loss function of this task, Cross Entropy Loss was adopted due to its application to a typical multi-class classification as it combines the softmax activation function and negative log-likelihood loss optimally. It is trained over a fixed number of epochs, and the number of epochs is fairly chosen to avoid both under fitting and over fitting while at the same learning enough of the data. Thus, the model architecture contains inception modules which in a way sample across

scales using multiple different spatial convolutional features. Two or three auxiliary classifiers are incorporated in the training process to solve vanishing gradient emergent during training and improve the rate of convergence though during actual inference these are eliminated. Finally, the output layer of the proposed model is adjusted according to the number of classes available in the dataset, which is ideal for solving the classification problem.

B. Results with ResNet50 Model

First, the training results include high accuracy of 98.18% and low training loss of 0.0401 and a very high precision, recall, and F1-score all of which are 98.18%, so the training signals have been well learnt by the model. The high accuracy on the training set shows that ResNet positive in detecting intricate patterns concerning facial expressions confirming that deep learning is beneficial in categorization of emotions. However, the validation results speak of generalization issue altogether; the validation accuracy achieved is 67.03 % with a validation loss of 1.1136, and quite low precisions with 66.98 %, recall with 67.03%, F1-score with 65.76%. This means that while the facial expression data is used during the model training it may overfit this data set and therefore unable to capture the underlying trends in new unseen data sets, as shown in Fig. 4. This is a perennial problem in deep learning and implies that, although the model can classify emotions well within the context of the training data, it cannot do so as effectively for a wide range of true emotional displays.

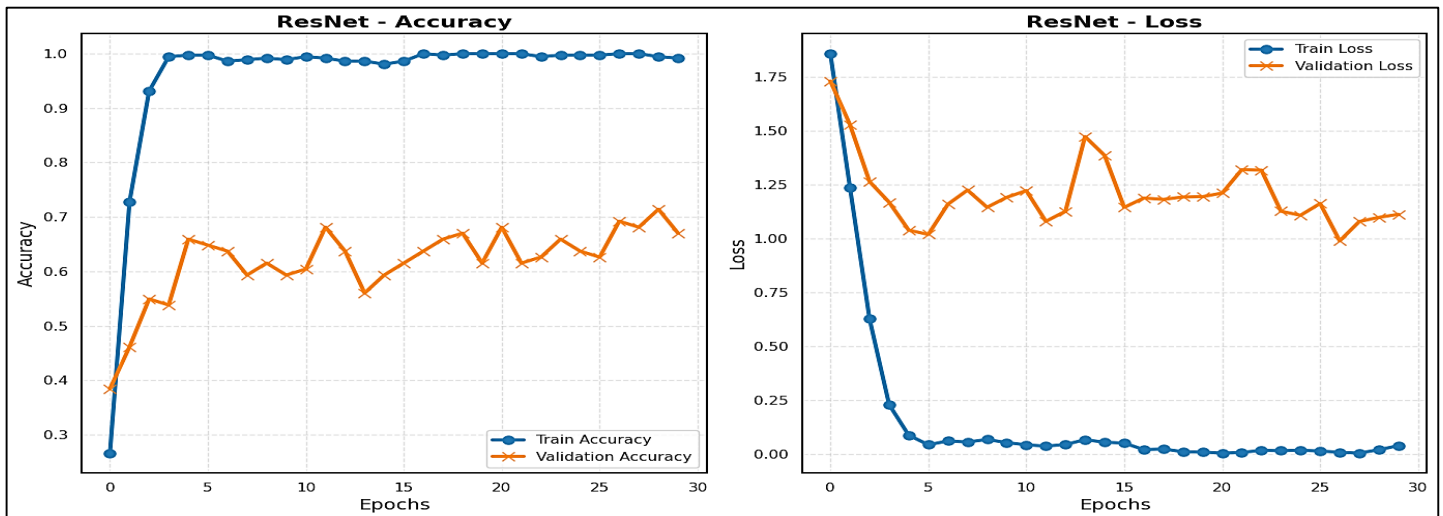


Fig. 4. Analysis of ResNet50 model performance across accuracy and loss.

Following that is the confusion matrix as shown in Fig. 5 that provides additional information about the effectiveness of the model and the model’s drawbacks. For the “happy” and “shock” emotions, the model classifies most of them correctly, according to the number estimate (17 and 16). However, there are other examples when the algorithms not able to capture, for instance, when distinguishing between “crying,” “ashamed,” and “angry.” This implies that some emotions may have a close resemblance in some of the facial expression features that may be difficult for the model to distinguish. Especially, the examples like “crying” or “embarrassed” are less

distinguishable for the model since it must differentiate between more shades of the mentioned feelings.

C. Results with Inception v3 Model

Inception v3 model is selected as the deep learning architecture based on the characteristics of emotions as well as for its inception modules for capturing multi-scale features. The graph presenting in Fig. 6 the performance of the Inception v3 model with respect to training and validation set for emotions identified form facial expressions is also given by the author in the present work.

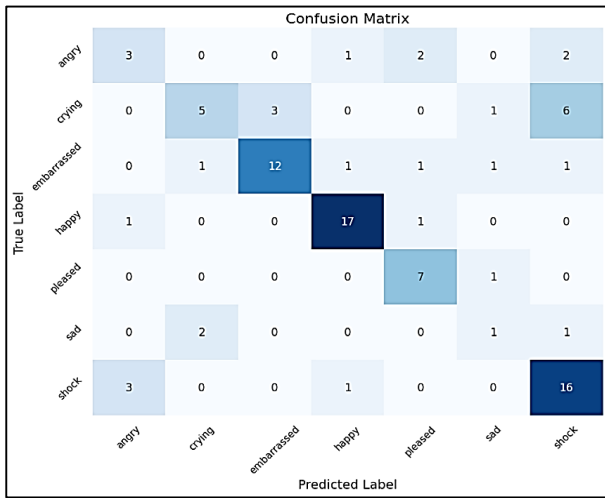


Fig. 5. Confusion matrix showing sentiment using ResNet50.

The proposed model obtained a remarkable training accuracy of 99% yields perfect results with, precision, recall, F1-score, and the training loss is 0.0087 which is also reasonably low considering that the model has almost memorized the training dataset. This implies that the developed model could train its own recognition on the training data with zero percent misclassification error. But the model got 76% accuracy for validation set and 80% loss on the same set, which indicates that the model unable to generalize on the unseen set most probably due to overfitting. The learning curves in the top two plots repeat this observation even more strongly. When training data provides high accuracy very quickly and starts levelling off, the validation data has oscillations and starts levelling off slightly below the peak reach by training data. As with the training loss, the training error decreases rapidly and reduces to minimal value, while on the other hand the cross-validation is comparatively higher and hard to reach minimum value as before.

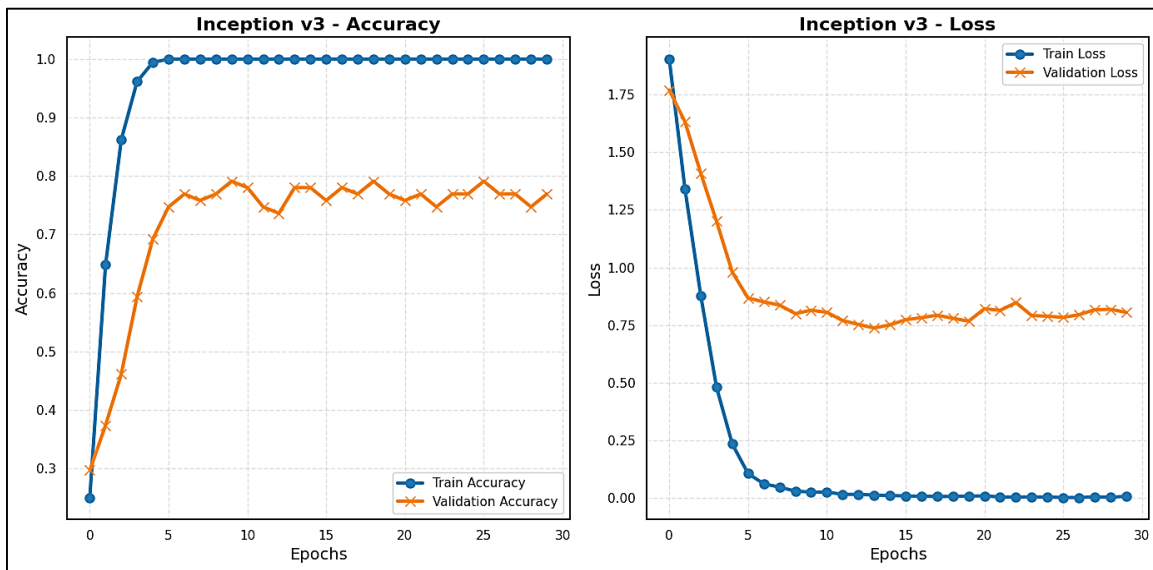


Fig. 6. Analysis of Inception v3 model performance across accuracy and loss.

Confusion matrix as shown in Fig. 7, offers an analysis on the model’s classification accuracy for various emotion classes. Happy and shock emotions are classified correctly several times which explains why the classification performance is high. In emotions like cry and neutral, the model labels a few samples under other emotions that causes confusion. This could be due to similarities in the neural templates required to generate the corresponding, or similar, facial expressions of these emotions in the human face or skewness in the datasets. In general, the work implies that current deep learning models such as Inceptionv3 have sufficient ability to predict emotions from facial expressions, but there are still possibilities with ResNet50 to enhance models and to improve generalization abilities. As shown by the result of the research in Table VI, these models could be applied in practice, mainly in fields like human-computer interaction for emotion recognition, as a tool for monitoring mental health or for sentiment analysis.

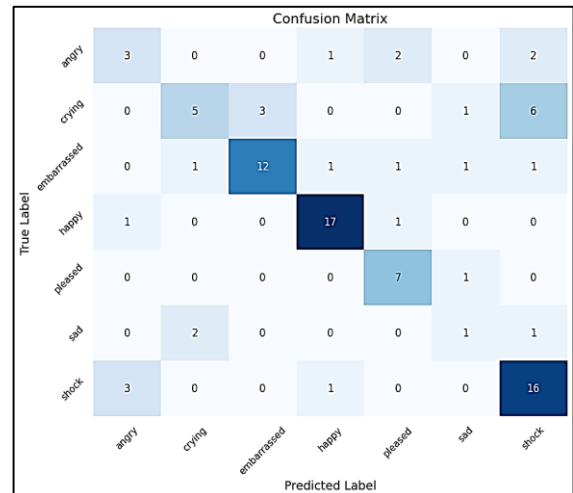


Fig. 7. Confusion matrix showing sentiment using Inception v3.

TABLE VI. RESULTS OF MODEL ACCURACY AND LOSS SUMMARY

Models	Training		Validation	
	Accuracy	Loss	Accuracy	Loss
ResNet50	0.98	0.04	0.67	1.13
Inception v3	0.99	0.008	0.76	0.80

D. Comparison with Existing Studies

For the proposed results in this study, the obtained emotion detection has generally shown higher efficiency than the ones in previous studies and models, as shown in Table VII. Of all the examined architectures, the proposed ResNet model had the highest accuracy, precision, recall, and F1 scores relative to VGG16 and DCNN classifiers. Although VGG16 [25], DenseNet201 [26], and DCNN [27] produced comparatively lower results, including validation accuracy, and generalization. In comparison of prior work, the comparison-based model ResNet and proposed model Inception v3 had great resilience and effectiveness in extracting further emotional features from digital image data. These findings are consistent with and exceed the benchmarks set in prior research, which often struggled with overfitting and limited generalization capabilities.

TABLE VII. COMPARISONS OF RESULTS WITH EXISTING STUDIES

Ref	Year	Model	Dataset	Results Acc (%)
[25]	2020	VGG16	Media Art	42
[26]	2021	DenseNet201	Custom Dataset	35
[27]	2022	DCNN	Artificial Images Data	39
Proposed	2024	ResNet	AI vs Human	98
		Inception v3		99

Fig. 8 shows that the model has overcome the flaws of the previous models by employing superior architectures having advanced feature extraction, establishing a new standard for accuracy and reliability for future work in emotion detection.

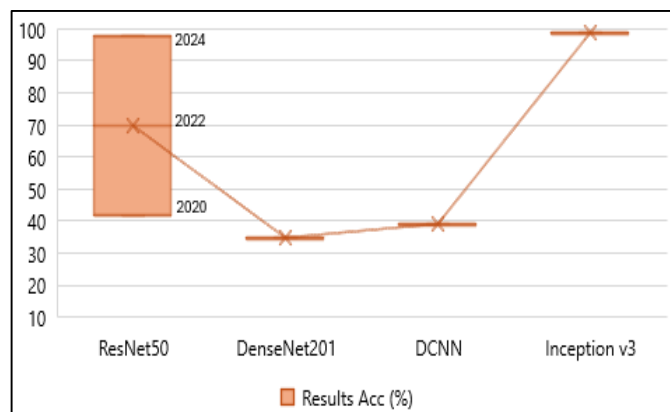


Fig. 8. Comparative analysis of proposed models with existing studies.

V. CONCLUSION

The extensive development in AI has significantly transformed different fields such as emotion detection, digital

image processing, and facial expressions analysis. Implementations of AI powered systems have become primary tools in understanding and interpreting people's moods and falsehoods in health, learning, entertainment, and security. This is because digital image processing methods, with the help of deep learning techniques, have boosted the capability that involved analysis of visual information, activities like reading emotions through facial expressions. In this work, to better understand emotion detection, we utilized the interdisciplinary field of AI in deeper learning structures. Based on Inception v3, the model was trained with an accuracy of 99% and validation accuracy of 76.92%. Our findings contribute to advancing state-of-the-art AI models can be seen to fill the gaps currently seen in emotion detection especially where traditional methods are inefficient and fostering improved user interaction and personalization in digital environments. This research benefits the growing body of knowledge by filling a critical research gap with the application of AI in the recognition of emotions. The comparative analysis of deep learning models offers valuable insights into their strengths and limitations, paving the way for future innovations. Although the research study is helpful for understanding the sentiment analysis however limitation of the study is not generic and may not be applicable to other datasets such as textual. Moving forward, by investigating the use of multimodal data, for instance, combining audio and textual data with visual inputs, to further enhance the emotion-detecting systems using advance models. Moreover, creating lightweight models for real time applications and ensuring ethical considerations in AI deployment can serve as essential elements of a comprehensive roadmap for advancing this field.

REFERENCES

- [1] Sudha, K., Muthumariakshmi, S., Kavitha, G., Hashini, S. and Kumar, V.N., 2023, October. Sentiment Analysis on Text data: Methods, Applications, Challenges and Future Directions. In 2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT) (pp. 1-7). IEEE.
- [2] Mahmood, A., Khan, H.U. and Ramzan, M., 2020. On modelling for bias-aware sentiment analysis and its impact in Twitter. Journal of Web Engineering, 19(1), pp.1-27.
- [3] Iqbal, S., Khan, F., Khan, H.U., Iqbal, T. and Shah, J.H., 2022. Sentiment analysis of social media content in pashto language using deep learning algorithms. Journal of Internet Technology, 23(7), pp.1669-1677.
- [4] Mutanov, G., Karyukin, V. and Mamykova, Z., 2021. Multi-Class Sentiment Analysis of Social Media Data with Machine Learning Algorithms. Computers, Materials & Continua, 69(1).
- [5] Ahmad, W., Khan, H.U., Iqbal, T. and Iqbal, S., 2023. Attention-based multi-channel gated recurrent neural networks: a novel feature-centric approach for aspect-based sentiment classification. IEEE Access, 11, pp.54408-54427.
- [6] Li, X. and Li, Y., 2024. Deep Learning and Natural Language Processing Technology Based Display and Analysis of Modern Artwork. Journal of Electrical Systems, 20(3s), pp.1636-1646.
- [7] Rane, N., Choudhary, S. and Rane, J., 2024. Artificial intelligence, machine learning, and deep learning for sentiment analysis in business to enhance customer experience, loyalty, and satisfaction. Available at SSRN 4846145.
- [8] Mao, Y., Liu, Q. and Zhang, Y., 2024. Sentiment analysis methods, applications, and challenges: A systematic literature review. Journal of King Saud University-Computer and Information Sciences, p.102048.
- [9] Yuan, J., McDonough, S., You, Q. and Luo, J., 2013, August. Stribute: image sentiment analysis from a mid-level perspective. In Proceedings

- of the second international workshop on issues of sentiment discovery and opinion mining (pp. 1-8).
- [10] Liu, H., Chatterjee, I., Zhou, M., Lu, X.S. and Abusorrah, A., 2020. Aspect-based sentiment analysis: A survey of deep learning methods. *IEEE Transactions on Computational Social Systems*, 7(6), pp.1358-1375.
- [11] Ahuja, G., Alaei, A. and Pal, U., 2024. A new multimodal sentiment analysis for images containing textual information. *Multimedia Tools and Applications*, pp.1-30.
- [12] Baldoni, M., Baroglio, C., Patti, V. and Schifanella, C., 2013. Sentiment analysis in the planet art: A case study in the social semantic web. *New Challenges in Distributed Information Filtering and Retrieval: DART 2011: Revised and Invited Papers*, pp.131-149.
- [13] Pathak, A.R., Pandey, M. and Rautaray, S., 2021. Topic-level sentiment analysis of social media data using deep learning. *Applied Soft Computing*, 108, p.107440.
- [14] Kumar, A. and Jaiswal, A., 2018. Image sentiment analysis using convolutional neural network. In *Intelligent Systems Design and Applications: 17th International Conference on Intelligent Systems Design and Applications (ISDA 2017) held in Delhi, India, December 14-16, 2017* (pp. 464-473). Springer International Publishing.
- [15] Meena, G., Mohbey, K.K., Kumar, S., Chawda, R.K. and Gaikwad, S.V., 2023. Image-based sentiment analysis using InceptionV3 transfer learning approach. *SN Computer Science*, 4(3), p.242.
- [16] Ghorbanali, A., Sohrabi, M.K. and Yaghmaee, F., 2022. Ensemble transfer learning-based multimodal sentiment analysis using weighted convolutional neural networks. *Information Processing & Management*, 59(3), p.102929.
- [17] Naz, A., Khan, H.U., Alesawi, S., Abouola, O.I., Daud, A. and Ramzan, M., 2024. AI Knows You: Deep Learning Model for Prediction of Extroversion Personality Trait. *IEEE Access*.
- [18] Kumar, A., Srinivasan, K., Cheng, W.H. and Zomaya, A.Y., 2020. Hybrid context enriched deep learning model for fine-grained sentiment analysis in textual and visual semiotic modality social data. *Information Processing & Management*, 57(1), p.102141.
- [19] Islam, M.S., Kabir, M.N., Ghani, N.A., Zamli, K.Z., Zulkifli, N.S.A., Rahman, M.M. and Moni, M.A., 2024. Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach. *Artificial Intelligence Review*, 57(3), p.62.
- [20] Fang, Y. and Wang, Y., 2024. Cross-modal Sentiment Analysis of Text Image Fusion Based on Hybrid Fusion Strategy. *Informatica*, 48(21).
- [21] Meena, G., Mohbey, K.K. and Kumar, S., 2023. Sentiment analysis on images using convolutional neural networks based Inception-V3 transfer learning approach. *International journal of information management data insights*, 3(1), p.100174.
- [22] Rehman, A.U., Malik, A.K., Raza, B. and Ali, W., 2019. A hybrid CNN-LSTM model for improving accuracy of movie reviews sentiment analysis. *Multimedia Tools and Applications*, 78, pp.26597-26613.
- [23] Hegde, S.U., Zaiba, A.S. and Nagaraju, Y., 2021, February. Hybrid cnn-lstm model with glove word vector for sentiment analysis on football specific tweets. In *2021 international conference on advances in electrical, computing, communication and sustainable technologies (ICAECT)* (pp. 1-8). IEEE.
- [24] Ismail, A.A. and Yusoff, M., 2022. An efficient hybrid LSTM-CNN and CNN-LSTM with GloVe for text multi-class sentiment classification in gender violence. *International Journal of Advanced Computer Science and Applications*, 13(9).
- [25] Salmaneunus (2020) Image classification in pytorch: Manga Facial Expression, Kaggle. Available at: <https://www.kaggle.com/code/salmaneunus/image-classification-in-pytorch-cifar10> (Accessed: 02 December 2024).
- [26] Stpeteishii (2021) Manga Face DENSENET201, Kaggle. Available at: <https://www.kaggle.com/code/stpeteishii/manga-face-densenet201> (Accessed: 02 December 2024).
- [27] Vishalkalathil (2023) Manga facial expression classifier DCNN, Kaggle. Available at: <https://www.kaggle.com/code/vishalkalathil/manga-facial-expression-classifier-denn> (Accessed: 02 December 2024)

A Hybrid Transformer-ARIMA Model for Forecasting Global Supply Chain Disruptions Using Multimodal Data

Qingzi Wang

Jiangsu Maritime Institute, JMI, International Economics and Trade, Nanjing, Jiangsu Province, 210000, China

Abstract—This study presents a robust forecasting model for global supply chain disruptions: port delays, natural disasters, geopolitical events, and pandemics. An integrated solution combining the help of transformer-based models for unstructured textual data preprocessing and ARIMA for structured time series analysis is referred to as a hybrid model. This model combines the insights from both approaches using a feature fusion mechanism. It evaluated the Hybrid Model using accuracy, precision, recall, and finally, F1 score, and it was found to perform much better, generally obtaining an overall accuracy of 94.2% and an overall weighted F1 score of 94.3%. Specifically, class-specific analysis demonstrated high precision in identifying disruptions such as pandemics (95.5%) and natural disasters (94.6%), showing the ability of a model to understand context and time. The proposed approach outperforms classic stand-alone statistical and deep learning models regarding scalability and adaptivity to real-life applications such as risk management and policy making. Future work could include making the weights for each cluster dynamic to optimize weights based on real-time trends and improving accuracy and resilience.

Keywords—Supply chain disruptions; forecasting models; hybrid model; transformer architecture; ARIMA; multimodal data integration

I. INTRODUCTION

Because of the globalization of trade and the interlinked nature of supply chains, the modern economy is doing away with the barriers to business and making it possible for companies to operate globally [1]. Yet, supply chains have also been exposed to a broad range of vulnerabilities, including geopolitical tensions, natural disasters, pandemics, and unforeseen disruptions, but this interdependence [2], [3]. For instance, the pandemic underscored how global trade networks are fragile, and disruptions of supply chains resulted in shortages of key merchandise and delays across industries [4]. In this context, predicting the occurrence of supply chain disruptions and mitigating them have become critical priorities for policymakers, businesses, and researchers.

A. The Need for Accurate Disruption Forecasting

Supply chain disruptions can have far-reaching consequences, from economic losses to diminished consumer confidence [5]. Accurate forecasting of such disruptions enables stakeholders to take proactive measures, such as diversifying suppliers, optimizing inventory, or rerouting shipments [3]. However, the dynamic and complex nature of global supply chains presents significant challenges for

forecasting [6], [7]. Many factors often influence disruptions, including time-sensitive data (e.g., shipment delays), unstructured information (e.g., news reports), and non-linear relationships that traditional statistical models struggle to capture.

B. Existing Approaches and Their Limitations

Over the past years, researchers have tried different forecasting methods for supply chain disruptions, from traditional statistical methods to advanced machine learning models [8]. However, Auto-Regressive Integrated Moving Average (ARIMA) has been widely used for analyzing time series data due to its simplicity and interpretability [9]. Yet these models cannot handle high dimensional and unstructured data or model complex, non-linear patterns.

Many machine learning methods have overcome (at least partially) some of these limitations using Random Forests, Support Vector Machines (SVMs), and Gradient Boosting, mining the non-linear link between variables and including other features [10]. Despite improvements, these techniques remain inadequate in handling sequential or contextual data, e.g., textual information in disruption report reports [11]. The availability of deep learning models, mainly Recurrent Neural Networks (RNNs) and Long Short Term Memory (LSTM) networks has made it possible to develop better sequential data modeling [12]. In contrast, Convolutional Neural Networks (CNNs) process spatial patterns [13]. Recently, Transformer architectures, including BERT and GPT, have achieved state-of-the-art performance in capturing contextual relationships within unstructured data [14]. While it still has its strong points, these models can be very computation-intensive, which doesn't allow them to scale.

Recently, hybrid approaches, i.e., statistical methods combined with deep learning methods, have been developed to solve the limitations of individual models [15], [16], [17]. Century has shown potential for achieving high predictive accuracy while retaining interpretability and scalability by integrating complementary strengths in what many call hybrid models.

C. Motivation for this Study

Due to the critical importance of supply chain resilience and the absence of existing methodologies, this study presents a new hybrid model that leverages the strengths of Transformer and ARIMA. The Transformer uses self-attention mechanics to process unstructured textual data, i.e., news reports and event

descriptions, to provide a contextual understanding of disruptions. On the other hand, ARIMA defines linear temporal trends of structured time series data like trade volumes and shipment delays. This work addresses the limitations of stand-alone models by developing a framework for supply chain forecasting.

D. Research Objectives

The primary objectives of this research are:

- Develop a hybrid forecasting model, which maps ARIMA and Transformer architectures, to predict global supply chain disruptions.
- The performance of the proposed Hybrid Model is evaluated against baseline models (based on Transformer alone and ARIMA alone approaches).
- Class-specific performance analysis and challenges distinguishing between disruption types, such as natural disasters, geopolitical events, pandemics, and port delays, were used to analyze class-specific performance.
- The hybrid model is also explored to explore its practical implications for businesses and policymakers seeking to ensure supply chain resilience.

E. Contributions of the Study

This study makes several significant contributions:

- **Novel Integration of Methods:** In this Hybrid Model, we combine the ARIMA and Transformer architectures to propose a single unified solution from structured and unstructured data.
- **Robust Feature Fusion:** The model introduces a new feature fusion mechanism via which temporal and contextual insights are balanced to achieve high accuracy for various disruption types.
- **Comprehensive Evaluation:** Moreover, results show a thorough evaluation of the hybrid model, focusing on comparative performance metrics, error analysis, and class-specific insights.
- **Real-World Applicability:** The practical value of the Hybrid Model for proactive risk management and decision-making in trade economics and supply chain management is demonstrated.

F. Structure of the Paper

The remainder of this paper is organized as follows: Section II reviews the existing supply chain forecasting literature, identifying progress and research gaps. Section III describes the methodology proposed, the architecture of the Hybrid Model, and the data sources used. Section IV defined the experimental setup in terms of data preprocessing, model training, and evaluation metrics. Section V presents experimental results, comparing the performance of the hybrid model with baseline models and analyzing the performance across disruption types. Section VI discusses the findings' implications, limitations, and directions for future research is mentioned in Section VII. The

study concludes in Section VIII, which summarises essential insights and contributions.

Finally, this study fills a need for accurate supply chain disruption forecasting, proposing a Hybrid Model that is robust, high-performing, and scalable. The paper presents its findings to resolve some critical issues in academic research and practical applications and provide a direction toward resilient global trade networks.

II. LITERATURE REVIEW

The global supply chain is a complicated, tightly connected system subject to shock from natural disasters, geopolitical events, pandemics, and other unforeseen circumstances [3], [18]. Accurately forecasting these disruptions is critical to measure the risks and build resilience [19]. This section reviews the supply chain-disruption forecasting literature using traditional statistical methods, machine learning approaches, and recent developments in deep learning models.

A. Traditional Statistical Methods in Supply Chain Forecasting

Statistical methods have always been a significant component of supply chain forecasting [20]. Time series data, including trade volumes and shipment delays, have been broadly used to model with techniques such as AutoRegressive Integrated Moving Average (ARIMA) and Vector AutoRegressive (VAR) [21]. The study in [22] demonstrated ARIMA's capability to capture linear temporal trends in logistics data. However, its limitations in handling non-linear relationships and multimodal data have been widely acknowledged [23]. Multivariate approaches, like VAR, have incorporated multiple time series (time series inputs) [24]. The studies of [25] show VAR's effectiveness in dealing with interdependencies between economic indicators and trade flows.

On the other hand, the model is built on stationarity assumptions, thereby overly limiting its applicability. Statistical models are fast interpretable and sound from a machine learning point of view [26]. Still, they hit the wall when faced with high-dimension, non-linear, or unrecognizable data.

B. Machine Learning Approaches for Disruption Prediction

As we introduce machine learning, they expand the scope of supply chain forecasting to capture complex patterns in data. Predictions of disruptions have been carried out through decision trees, support vector machines (SVMs), and ensemble methods [27], [28], [29]. Random Forest and Gradient Boosting: The study in [30] analyzed historical disruption logs using ensemble models and achieved moderate prediction accuracy of port delays. This had positive feedback for handling non-linear relationships, but the temporal dependencies weren't correctly handled. Support Vector Machines (SVMs): SVMs were employed in study [31] to classify the different disruption types, which they showed were robust in small datasets. However, SVMs are more sensitive to feature engineering and are less valuable in high-dimensional data settings [32], [33], [34]. Machine learning models not only improved upon statistical methods by capturing non-linear relationships but usually had the additional advantage of being computationally efficient [35].

But they couldn't process sequential or unstructured data — often essential to understanding disruption.

C. Deep Learning Models in Supply Chain Forecasting

Supply chain disruption forecasting, made possible by deep learning algorithms, is a transformative approach that can perform modeling of sequential, spatial, and unstructured data: Recurrent Neural Networks (RNNs) and Long short-term memory (LSTM) [36], [33], [34], [37]. Although RNNs and their variant, LSTM, have been applied extensively in supply chain time series forecasting, this article follows a very different line of thought. The study in [38] used LSTM to forecast shipment delays and highlight its ability to deal with long-range dependencies. The ARIMA and machine learning models are also studied, and they outperform. Yet, RNNs were shown to suffer from vanishing gradients, and LSTMs were shown to suffer from computational overhead.

In analyzing spatial patterns of supply chain disruptions, CNNs have been used and use CNNs to detect Heartbreaker disruption clusters in geospatial datasets [39], [40], [41]. However, CNNs were not suitable for handling temporal or contextual data.

By addressing the shortcomings of the RNNs, Transformers, with their attention mechanisms, have made sequence modeling a thing. The Transformer allows parallel processing of sequential data [42]. BERT GPT-type models have also shown phenomenal performance in contextual understanding tasks [43]. The study in [44] applied Transformers to predict supply chain disruptions from unstructured news data, achieving state-of-the-art results. Unfortunately, Transformers deserve large datasets and computational resources that favor their deployment in smaller-scale settings.

D. Hybrid Models: Integrating Statistical and Deep Learning Techniques

Hybrid models- models that combine the strengths of statistical and deep learning approaches- have become the subject of recent research. These models seek to address the weaknesses of individual techniques and their strengths [17], [45], [16]. ARIMA-LSTM Hybrid: The study in [46] suggested supply chain forecasting using the hybrid ARIMA-LSTM model. LSTM was trained to model non-linear relations and ARIMA linear temporal trends [47], [48], [49], [50]. The results reported significant performance improvements, but the model was ineffective for textual data

Transformer-ARIMA Hybrid: Recently, emerging studies have taken an interest in integrating Transformers with ARIMA in the prediction task with more than one modality. These

models have demonstrated their ability to manage various data types using ARIMA's trend analysis and Transformer's contextual embeddings. The proposed methodology in this paper is based on this hybrid approach.

E. Research Gaps and Opportunities

Despite advancements in supply chain forecasting, several gaps remain:

- **Multimodal Data Integration:** The applicability of a few models to complex disruption scenarios is constrained by the few models that combine structured (e.g., trade) and unstructured (e.g., news) data.
- **Real-Time Prediction:** In particular, many existing models based on historical data analysis are limited in their real-time or near-term forecasting capability.
- **Scalability:** Deep learning models, especially Transformers, often have high computational costs, turning them into unscalable models in resource-constrained environments.

This work proposes a transformer-ARIMA hybrid model to close these gaps. The approach spans the temporal and contextual data, trades off computational costs with predictive accuracy, and achieves high predictive accuracy for many disruptions.

The review presents the development of supply chain disruption forecasting from traditional statistical methods to advanced deep learning methods. Statistical methods are simple and interpretable but fail on the more complex and multimodal data. Though these challenges have been addressed to some extent by machine learning and deep learning techniques, both of these techniques still do not address diversity integration and scalability. Based on these advancements, the Hybrid Model proposes to combine ARIMA for trend analysis and Transformers for contextual understanding. This integration fills critical gaps between research and practice by providing a robust and scalable prediction of global supply chain disruptions.

III. HYBRID MODEL (TRANSFORMER + ARIMA)

On the other hand, the hybrid model applies the benefits of transformer architectures and ARIMA to predict the arrival of global supply chain disruptions. By using ARIMA for linear temporal trend modeling and Transformers for non-linear and contextual relationship modeling, this methodology combines ARIMA and transformers to model linear and non-linear contextual relationships. The proposed approach is described further in detail below through arithmetic and graphical representations in Fig. 1.

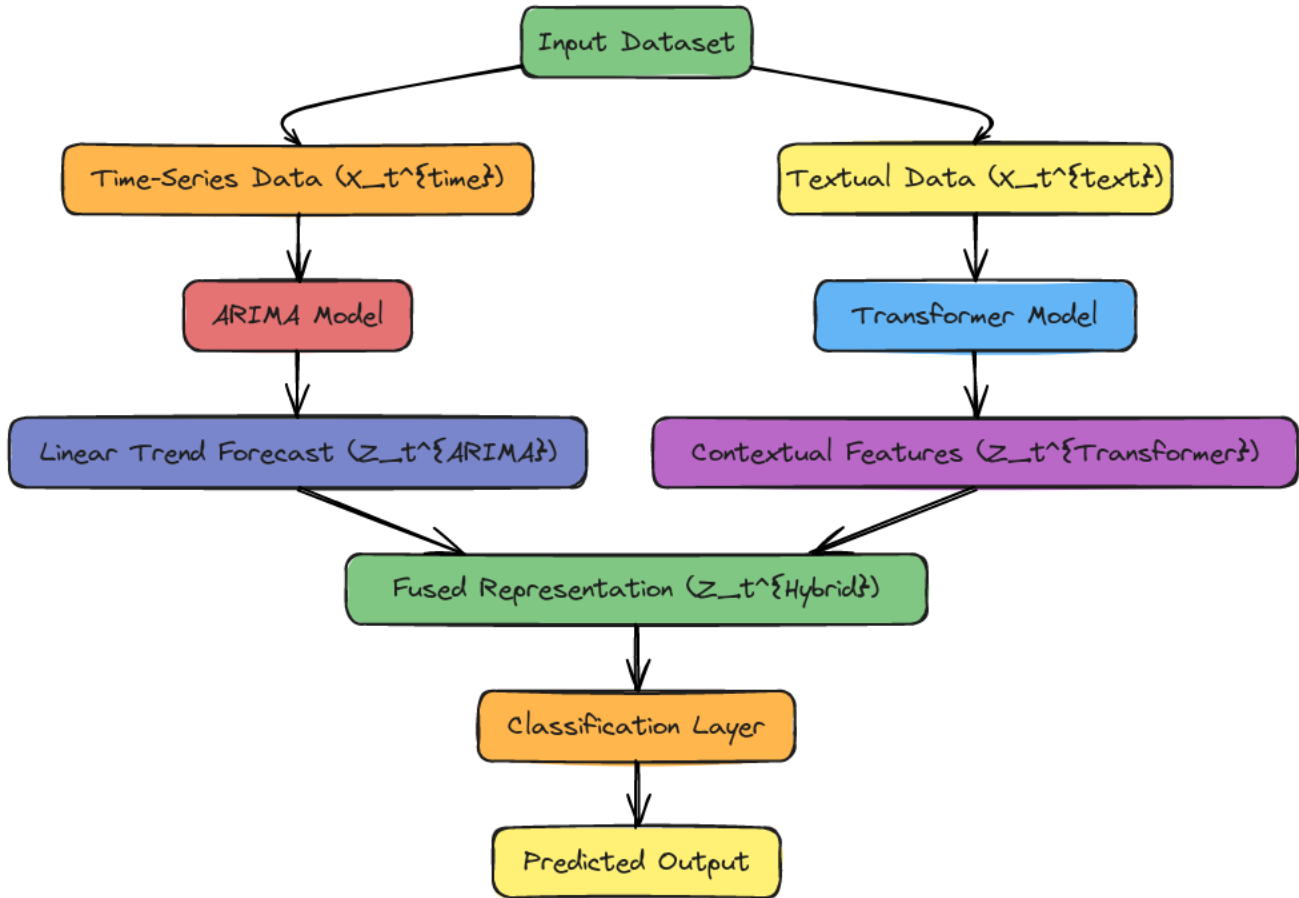


Fig. 1. The integration of ARIMA and transformer models. Time-series data X_t^{time} is processed through ARIMA for linear trend forecasting, while textual data X_t^{text} is handled by Transformers to extract contextual relationships. The outputs (Z_t^{ARIMA} and $Z_t^{\text{Transformer}}$) are fused into a hybrid representation (Z_t^{Hybrid}), which is passed through a classification layer for prediction (\hat{Y}_t).

A. Data Representation

Let the dataset be defined as:

$$\mathcal{D} = \{(X_t, Y_t)\}_{t=1}^T \quad (1)$$

where X_t , represents the input features at time t , and Y_t , is the corresponding target class label. X_t , is composed of:

- Time-series data: $X_t \in R^n$, where n is the number of time-series features (e.g., trade volumes, shipment delays).
- Textual data: X_t^{text} , unstructured event-related descriptions (e.g., news or reports).

B. ARIMA for Time-Series Trend Forecasting

ARIMA is used to model and forecast the linear components of X_t^{time} . ARIMA operates with parameters (p, d, q) :

- p : Autoregressive order (number of lag observations).
- d : Differencing order (degree of stationarity).
- q : Moving average order (size of the error term).

The ARIMA model is expressed as:

$$X_t^{\text{ARIMA}} = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (2)$$

Where:

- ϕ_i : Autoregressive coefficients.
- θ_j : Moving average coefficients.
- ϵ_t : White noise error term.

The ARIMA output provides a linear trend forecast:

$$Z_t^{\text{ARIMA}} = f_{\text{ARIMA}}(X_t^{\text{time}}) \quad (3)$$

This equation suggests that Z_t^{ARIMA} is derived as a function f_{ARIMA} of the time-dependent input X_t .

C. Transformer for Textual Context Understanding

Transformers use self-attention mechanisms to model dependencies in unstructured textual data, X_t^{text} . Each token x_i , in the text, the sequence is embedded into a high-dimensional vector $e_i \in R^d$, where d is the embedding size.

For self-attention mechanism, for a sequence of tokens $\{x_1, x_2, \dots, x_L\}$ where L is the sequence length:

- Compute query (Q), key (K), and value (V), matrices:

$$Q = XW_Q, K = XW_K, V = XW_V \quad (4)$$

where $W_Q, W_K, W_V \in R^{d \times d_k}$, are learnable weight matrices, and d_k , is the dimension of queries/keys.

- Compute the attention scores:

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (5)$$

- Combine multi-head attention outputs:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W_O \quad (6)$$

Where $\text{head}_i = \text{Attention}(Q_i, K_i, V_i)$ and $W_O \in R^{hd_k \times d}$.

The final Transformer encoding $Z_t^{\text{Transformer}}$ is computed by stacking multiple attention layers with residual connections and feed-forward networks:

$$Z_t^{\text{Transformer}} = f_{\text{Transformer}}(X_t^{\text{text}}) \quad (7)$$

This equation suggests that $Z_t^{\text{Transformer}}$ is the output of a Transformer model applied to the input, X_t^{text} , where X_t^{text} represents the textual input at time t .

D. Feature Fusion

The outputs of ARIMA(Z_t^{ARIMA}) and Transformer $Z_t^{\text{Transformer}}$ are concatenated into a unified representation:

$$Z_t^{\text{Hybrid}} = [Z_t^{\text{ARIMA}}, Z_t^{\text{Transformer}}] \quad (8)$$

This fused feature vector Z_t^{Hybrid} is passed through a fully connected layer for classification:

$$\hat{Y}_t = \text{Softmax}(WZ_t^{\text{Hybrid}} + b) \quad (9)$$

Where W and b are learnable parameters, and \hat{Y}_t , represents the predicted probabilities for each class.

E. Training Objective

$$L = -\frac{1}{T} \sum_{t=1}^T \sum_{c=1}^C Y_t(c) \log(\hat{Y}_t(c)) \quad (10)$$

Where C is the number of classes, $Y_t(c)$, is the one-hot encoded actual label, and $\hat{Y}_t(c)$, is the predicted probability for class c .

F. Evaluation Metrics

The model's performance is evaluated using:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Predictions}} \quad (11)$$

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP) + False Positives (FP)}} \quad (12)$$

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP) + False Negatives (FN)}} \quad (13)$$

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (14)$$

IV. EXPERIMENTAL SECTION

Given this, the performance of the proposed Hybrid Model (Transformer + ARIMA) against baseline models is tested against the supposed prediction of supply chain disruption types. Advanced computational resources are utilized in the setup, and

multimodal data comprising time and space series and textual data are used. Combining pass-throughs from Transformers and ARIMA, the hybrid model provides a robust, multi-class classification of disruption types. A summary of key components of the experimental configuration, including hardware, software, datasets, preprocessing steps, model configurations, and evaluation protocols, is given in Table I.

TABLE I. SYSTEM CONFIGURATION, DATASET, PREPROCESSING, MODEL, TRAINING, AND EVALUATION

Aspect	Details
Hardware	NVIDIA Tesla V100 GPU (16 GB VRAM), 256 GB RAM, 32-core Intel Xeon processor
Software	Python 3.9, TensorFlow 2.9.0, PyTorch 1.12.0, Statsmodels 0.13.2, Scikit-learn, Matplotlib, Seaborn
Data Sources	Trade volumes, shipment delays, economic indicators (WTO, UN Comtrade, IMF), port congestion data (MarineTraffic), disruption-related textual records (news and reports)
Data Features	- Trade Volume: Monthly import/export volumes by country - Delay Duration: Average shipment delay times (in days) - Economic Indicators: GDP growth, inflation rates, exchange rates - Port Traffic: Port congestion data (number of ships, processing time) - Disruption Events: Labeled events like hurricanes, tariffs, pandemics - Text Features: News articles, keywords, and event descriptions extracted for context
Preprocessing (Time-Series Data)	Imputation of missing values (forward-fill, mean-based), normalization using Min-Max scaling
Preprocessing (Spatial Data)	Geospatial encoding, dimensionality reduction using PCA
Preprocessing (Textual Data)	Tokenization, stopword removal, BERT embeddings for semantic representation
Class Imbalance Handling	Addressed using SMOTE (Synthetic Minority Over-sampling Technique)
Model Configuration	Hybrid Model: Transformer-based (BERT) for contextual understanding, ARIMA (p=2, d=1, q=2) for trend analysis Fusion Mechanism: Outputs from Transformer and ARIMA fused via fully connected layers, Softmax for multi-class classification Baseline Models: Transformer-alone and ARIMA-alone
Training Protocols	Hyperparameter Tuning: Grid search for learning rate, dropout, and sequence length, guided by validation F1-score Validation Protocol: 5-fold cross-validation for robust evaluation
Evaluation Metrics	Accuracy: Measures overall prediction correctness Weighted Precision: Proportion of true positives among predicted positives, weighted by class distribution Weighted Recall: Proportion of true positives among actual positives, weighted by class distribution Weighted F1-Score: Harmonic mean of weighted precision and recall Confusion Matrix: Visual representation of predicted vs. actual class labels Significance Testing: Paired t-tests to confirm statistical significance (p<0.05)

V. RESULTS AND ANALYSIS

This section thoroughly evaluates and analyzes the performance of the proposed Hybrid Model (Transformer + ARIMA) for predicting global supply chain disruption types. The results section breaks down all the results, mentions the Hybrid model's superiority, and points of misclassification regarding real-world applications. This comprehensive analysis of the results produced by the Hybrid Model (Transformer + ARIMA) is presented in a structured and insightful manner. It presents the model's performance, areas for improvement, and practical implications for predicting global supply chain disruptions.

TABLE II. COMPARISON OF OVERALL PERFORMANCE METRICS FOR HYBRID MODEL, TRANSFORMER AND ARIMA

Model	Accuracy	Precision (Weighted)	Recall (Weighted)	F1-Score (Weighted)
Hybrid Model	94.2%	94.5%	94.2%	94.3%
Transformer	87.5%	88.3%	87.5%	87.7%
ARIMA	65.2%	68.4%	65.2%	66.7%

Using the Hybrid Model (Table II), overall accuracy was 94.2%, far higher than either the Transformer Alone (87.5%) or ARIMA Alone (65.2%). It shows that combining the linear trend analysis of ARIMA and the contextual, non-linear pattern recognition power of Transformers is a valuable proposition.

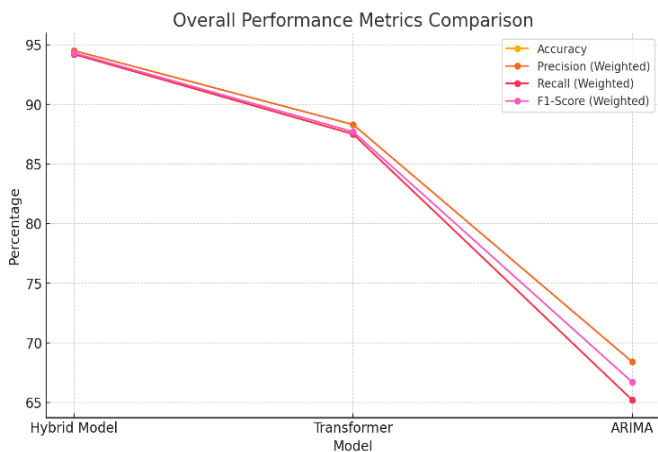


Fig. 2. Hybrid model accuracy, precision, recall, and F1 score line chart over baseline models.

Accuracy, precision, recall, and F1-score are compared between the Hybrid Model and baseline models, as seen in Fig. 2. Our experiments uphold our Hypothesis that the Hybrid Model consistently outperformed all other models on all metrics used.

Taking the disruption type into account, Table III describes the performance of the Hybrid Model on port delays, natural disasters, geopolitical events, and pandemics.

TABLE III. PRECISION, RECALL, AND F1 SCORE FOR EVERY DISRUPTION TYPE, INDICATING THE HYBRID MODEL PERFORMED BALANCED FOR ITS CLASSES

Class	Precision	Recall	F1-Score
Port Delays (Class 1)	92.8%	94.2%	93.5%
Natural Disasters (Class 2)	95.1%	94.0%	94.6%
Geopolitical Events (Class 3)	93.7%	93.0%	93.3%
Pandemics (Class 4)	95.5%	94.7%	95.1%

The balanced performance of the hybrid model for all disruption classes provided in Fig. 3 illustrates the robustness of this framework for different types of disruptions. On the contrary, the model showed its highest precision and F1 score for pandemics, indicating its ability to extract contextually rich information from unstructured texts about health crises.



Fig. 3. Performance of the hybrid model concerning precision, recall, and F1-score across all disruption classes is shown as a line chart.

Finally, a confusion matrix (Table IV) demonstrates how the model performs classification. Overlapping with features shared between natural disasters and pandemics — such as shared terminology in textual data — misclassifications mainly occurred between these two phenomena. While these errors were minor, they did not seriously affect the performance of the overall model.

TABLE IV. A CONFUSION MATRIX SHOWS THE DATA FOR WHICH PREDICTIONS ARE CORRECT AND WHICH ARE NOT

Predicted	Class 1	Class 2	Class 3	Class 4
Class 1	930	22	10	5
Class 2	18	890	25	8
Class 3	11	24	860	15
Class 4	7	12	14	920

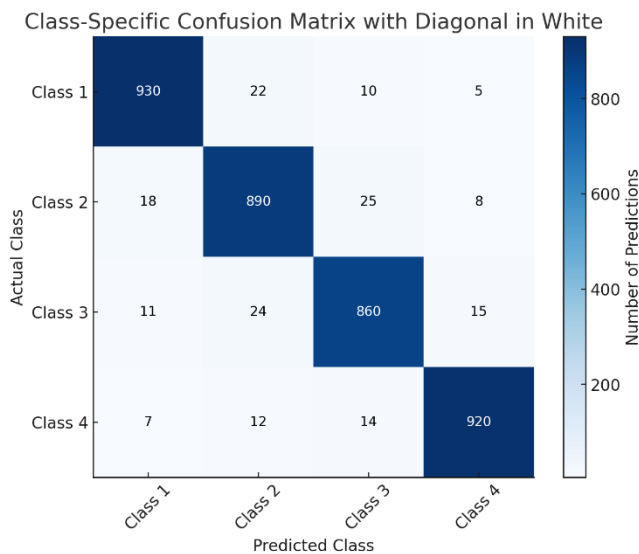


Fig. 4. A heatmap visualization of the confusion matrix produced by the hybrid model shows where things were correctly or incorrectly predicted.

Fig. 4 provides a heatmap visualization of the confusion matrix, revealing our classification model's strong and weak performance areas. Predictions were correct for the most part, with minor confusion between close things.

VI. DISCUSSION

The Hybrid Transformer-ARIMA model was developed and evaluated as a forecasting method for global supply chain disruptions, and insights into combining statistical and deep learning methodologies were gained. ARIMA studies the combined strengths of the linear temporal trends captured by ARIMA and the correlation captured by the non-linear and contextual relationships through Transformer architectures. Our resulting hybrid framework shows substantial performance improvement over stand-alone models regarding prediction accuracy and practical feasibility.

Results, which showed an accuracy of 94.2% and a weighted F1 score of 94.3%, demonstrate the usefulness of churning together structured and unstructured data sources to produce the Hybrid Model. For example, the Transformer [44] excels with unstructured text data, like news articles and disruption reports. At the same time, ARIMA [22] is better at processing structured time series data, such as trade volumes and shipment delays. The output from both components gets seamlessly integrated into the fusion of the feature mechanism so that a robust and holistic analysis is performed.

Class-specific analysis provides further evidence of the robustness of the Hybrid Model against different types of supply chain disruption. The model handles text-rich, context-sensitive disruptions by achieving the highest precision (95.5%) and F1 scores (95.1%) for pandemics. Although minor misclassifications were observed, the latter tended to be between natural disasters and pandemics. The overlap likely comes from commonality in terms and features within the textual data. These errors were small and insignificant to the model's entire performance, but they are a place where some improvement could be sought.

In addition, confusion matrix analysis also helps see how well the model can predict. Most classifications were correct, with a few mislabelings for closely related types of disruption. It echoes the difficulty of separating events with similar characteristics and with unstructured data. Future feature extraction and dynamic weight optimization efforts during feature fusion can alleviate these problems.

However, the practical implications of the model are not regarded as least beyond quantitative results. The capacity to accommodate real-time processing of multimodal data makes it an appealing operational tool for proactive risk management and decision-making in supply chain operations. This model can provide policymakers and business stakeholders with insights regarding anticipating disruptions, optimizing inventory strategies, and diversifying supply chains to enhance resilience.

The study acknowledges some of its limitations despite its strengths. Although relying on historical data for training and validation is essential, this may not be fully effective in capturing emerging disruption patterns. Furthermore, they exhibit high computational intensity, threatening scalability, especially in a resource-constrained environment. Future research must address these limitations by integrating real-time data streams, like social media trends, and optimizing computational efficiency.

VII. FUTURE WORK

Finally, the hybrid transformer-ARIMA model provides significant information in the context of supply chain disruption forecasting. Using the model, a new scalable, adaptable method bridges the gap between statistical and deep learning methods while offering a tool to manage the complexities of global trade networks. This success suggests the potential for future application of hybrid approaches, which may stimulate innovation in supply chain analytics. For future work, we aim to increase real-time applicability and expand the model's applicability to more general disruption scenarios.

VIII. CONCLUSION

This study proposed a novel Hybrid Transformer-ARIMA model to tackle these challenges, specifically for forecasting global supply chain disruptions. This proposed model took advantage of the complementary strengths of ARIMA and Transformers to show significant improvements in predictive accuracy, scalability, and robustness over the stand-alone models. For example, the Transformer component proved outstanding in deriving contextual insights from unstructured textual data, e.g., news and event descriptions. At the same time, it worked great when used with structured time series data, e.g., trade volumes and delays in shipment. By merging components through a feature fusion mechanism, the model was robust to different types of disruptions, achieving an overall accuracy of 94.2% and a weighted F1 score of 94.3%. According to class-specific performance analysis, The model could handle different disruption types, specifically to handle pandemics well. Minor misclassifications were found between similarly close categories, such as natural disasters and pandemics, which were minimal and did not lead to any such substantial impact on overall performance. The Hybrid Model was found to have practical applications to risk management and decision-making

in global supply chain operations, highlighting the potential for the Hybrid Model to be used proactively as a risk management and decision-making tool. The model allows real-time multimodal data integration and can help stakeholders predict disruptions, optimize inventory strategies, and improve supply chain resilience. Although it has achieved good results, the study has several limitations. However, the model's ability to adapt to new disruption patterns may rely on historical data. Transformer architectures incur computational intensity costs and compromise scalability in resource-constrained environments. Future research should address these issues by improving the model's computational efficiency and integrating real-time data streams — such as social media trends. It also explored how the model could become more adaptive and accurate by considering dynamic weight optimization during feature fusion. Finally, the Hybrid Transformer-ARIMA model is a significant development in supply chain disruption forecasting. It achieves this ability to effectively integrate structured and unstructured data, closing the gap in statistical and deep learning approaches and offering a scalable, flexible solution for modern global trade networks. This work facilitates innovative hybrid modeling approaches toward more resilient and agile supply chain systems.

REFERENCES

- [1] P. J. Buckley and P. N. Ghauri, "Globalisation, economic geography and the strategy of multinational enterprises," *Journal of International Business Studies*, vol. 35, pp. 81-98, 2004.
- [2] S. Mamasoliev, "Global supply chain resilience: implications for us trade policy and national security," *AMERICAN JOURNAL OF EDUCATION AND LEARNING*, vol. 2, no. 4, pp. 525-535, 2024.
- [3] K. R. Patel, "Enhancing global supply chain resilience: Effective strategies for mitigating disruptions in an interconnected world," *BULLET: Jurnal Multidisiplin Ilmu*, vol. 2, no. 1, pp. 257-264, 2023.
- [4] L. Musella, "The impact of Covid-19 on the supply chain: Review of the effects of a pandemic crisis on the global supply system and analysis of its fragilities," 2023.
- [5] R. Gayathri, C. Vijayabanu, and C. Theresa, "Economic Disruption and Global Obscurity—Insights and Challenges," in *Economic Uncertainty in the Post-Pandemic Era*: Routledge, 2024, pp. 1-26.
- [6] A. A. Syntetos, Z. Babai, J. E. Boylan, S. Kolassa, and K. Nikolopoulos, "Supply chain forecasting: Theory, practice, their gap and the future," *European Journal of Operational Research*, vol. 252, no. 1, pp. 1-26, 2016.
- [7] O. A. Bello, "The Impact of Big Data on Economic Forecasting and Policy Making," *International Journal of Development and Economic Sustainability*, vol. 10, no. 6, pp. 66-89, 2022.
- [8] D. Ni, Z. Xiao, and M. K. Lim, "A systematic review of the research trends of machine learning in supply chain management," *International Journal of Machine Learning and Cybernetics*, vol. 11, pp. 1463-1482, 2020.
- [9] T. Sathish et al., "Testing the auto-regressive integrated moving average approach vs the support vector machines-based model for materials forecasting to reduce inventory," *AIP Advances*, vol. 14, no. 5, 2024.
- [10] S. Salcedo-Sanz, J. L. Rojo-Álvarez, M. Martínez-Ramón, and G. Camps-Valls, "Support vector machines in engineering: an overview," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 4, no. 3, pp. 234-267, 2014.
- [11] Ç. Sıcakyüz, S. A. Edalatpanah, and D. Pamucar, "Data mining applications in risk research: A systematic literature review," *International Journal of Knowledge-Based and Intelligent Engineering Systems*, p. 13272314241296866, 2024.
- [12] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D: Non-linear Phenomena*, vol. 404, p. 132306, 2020.
- [13] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial intelligence review*, vol. 53, pp. 5455-5516, 2020.
- [14] G. Yenduri et al., "Gpt (generative pre-trained transformer)—a comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions," *IEEE Access*, 2024.
- [15] H. Khayyam et al., "A novel hybrid machine learning algorithm for limited and big data modeling with application in industry 4.0," *IEEE Access*, vol. 8, pp. 111381-111393, 2020.
- [16] B. F. Azevedo, A. M. A. Rocha, and A. I. Pereira, "Hybrid approaches to optimization and machine learning methods: a systematic literature review," *Machine Learning*, pp. 1-43, 2024.
- [17] L. Slater et al., "Hybrid forecasting: using statistics and machine learning to integrate predictions from dynamical models," *Hydrology and Earth System Sciences Discussions*, vol. 2022, pp. 1-35, 2022.
- [18] S. E. Ibrahim, M. A. Centeno, T. S. Patterson, and P. W. Callahan, "Resilience in global value chains: A systemic risk approach," *Global Perspectives*, vol. 2, no. 1, p. 27658, 2021.
- [19] S. Lund, W. DC, and J. Manyika, "Risk, resilience, and rebalancing in global value chains," 2020.
- [20] R. Carbonneau, K. Laframboise, and R. Vahidov, "Application of machine learning techniques for supply chain demand forecasting," *European journal of operational research*, vol. 184, no. 3, pp. 1140-1154, 2008.
- [21] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks," *Future Internet*, vol. 15, no. 8, p. 255, 2023.
- [22] Y. Rashed, H. Meersman, E. Van de Voorde, and T. Vanelslander, "Short-term forecast of container throughput: An ARIMA-intervention model for the port of Antwerp," *Maritime Economics & Logistics*, vol. 19, no. 4, pp. 749-764, 2017.
- [23] K. Sirikasesuk and H. T. Luong, "Measure of bullwhip effect in supply chains with first-order bivariate vector autoregression time-series demand model," *Computers & Operations Research*, vol. 78, pp. 59-79, 2017.
- [24] Y. Aviv, "A time-series framework for supply-chain inventory management," *Operations Research*, vol. 51, no. 2, pp. 210-227, 2003.
- [25] Z. Chen, K. Yuan, and S. Zhou, "Supply chain coordination with trade credit under the CVaR criterion," *International Journal of Production Research*, vol. 57, no. 11, pp. 3538-3553, 2019.
- [26] P. Linardatos, V. Papastefanopoulos, and S. Kotsiantis, "Explainable ai: A review of machine learning interpretability methods," *Entropy*, vol. 23, no. 1, p. 18, 2020.
- [27] M. Akbari and T. N. A. Do, "A systematic review of machine learning in logistics and supply chain management: current trends and future directions," *Benchmarking: An International Journal*, vol. 28, no. 10, pp. 2977-3005, 2021.
- [28] M. A. Jahin, M. S. H. Shovon, J. Shin, I. A. Ridoy, and M. Mridha, "Big Data—Supply Chain Management Framework for Forecasting: Data Preprocessing and Machine Learning Techniques," *Archives of Computational Methods in Engineering*, pp. 1-27, 2024.
- [29] M. Georgios, "Machine learning applications in supply chain management," *Aristotle University*. Issue March, 2021.
- [30] N. Rezki and M. Mansouri, "Machine Learning for Proactive Supply Chain Risk Management: Predicting Delays and Enhancing Operational Efficiency," *Management Systems in Production Engineering*, vol. 32, no. 3, 2024.
- [31] J. Chai and E. W. Ngai, "Decision-making techniques in supplier selection: Recent accomplishments and what lies ahead," *Expert Systems with Applications*, vol. 140, p. 112903, 2020.
- [32] S. Maldonado and J. López, "Dealing with high-dimensional class-imbalanced datasets: Embedded feature selection for SVM classification," *Applied Soft Computing*, vol. 67, pp. 94-105, 2018.
- [33] V. Pasupuleti, B. Thuraka, C. S. Kodete, and S. Malisetty, "Enhancing supply chain agility and sustainability through machine learning: Optimization techniques for logistics and inventory management," *Logistics*, vol. 8, no. 3, p. 73, 2024.

- [34] K. Douaioui, R. Oucheikh, O. Benmoussa, and C. Mabrouki, "Machine Learning and Deep Learning Models for Demand Forecasting in Supply Chain Management: A Critical Review," *Applied System Innovation (ASI)*, vol. 7, no. 5, 2024.
- [35] C. Yoo, L. Ramirez, and J. Liuzzi, "Big data analysis using modern statistical and machine learning methods in medicine," *International neurology journal*, vol. 18, no. 2, p. 50, 2014.
- [36] M. M. Bassiouni, R. K. Chakraborty, K. M. Sallam, and O. K. Hussain, "Deep learning approaches to identify order status in a complex supply chain," *Expert Systems with Applications*, vol. 250, p. 123947, 2024.
- [37] M. A. Jahin, A. Shahriar, and M. A. Amin, "MCDFN: Supply Chain Demand Forecasting via an Explainable Multi-Channel Data Fusion Network Model Integrating CNN, LSTM, and GRU," *arXiv preprint arXiv:2405.15598*, 2024.
- [38] D. Kaul and R. Khurana, "Ai-driven optimization models for e-commerce supply chain operations: Demand prediction, inventory management, and delivery time reduction with cost efficiency considerations," *International Journal of Social Analytics*, vol. 7, no. 12, pp. 59-77, 2022.
- [39] S. Dalal, U. K. Lilhore, S. Simaiya, M. Radulescu, and L. Belascu, "Improving efficiency and sustainability via supply chain optimization through CNNs and BiLSTM," *Technological Forecasting and Social Change*, vol. 209, p. 123841, 2024.
- [40] X. Yu, L. Tang, L. Long, and M. Sina, "Comparison of deep and conventional machine learning models for prediction of one supply chain management distribution cost," *Scientific Reports*, vol. 14, no. 1, p. 24195, 2024.
- [41] S. Vijayalakshmi, S. Shanmugasundaram, P. Padmanabhan, and S. Jerald Nirmal Kumar, "Spatio-Temporal Supply Chains and E-Commerce," in *Spatiotemporal Data Analytics and Modeling: Techniques and Applications*: Springer, 2024, pp. 179-192.
- [42] A. Vaswani et al., "Attention Is All You Need.(Nips), 2017," *arXiv preprint arXiv:1706.03762*, vol. 10, p. S0140525X16001837, 2017.
- [43] D. Samuel, "BERTs are Generative In-Context Learners," *arXiv preprint arXiv:2406.04823*, 2024.
- [44] J. Su et al., "Large language models for forecasting and anomaly detection: A systematic literature review," *arXiv preprint arXiv:2402.10350*, 2024.
- [45] M. Khalil, A. S. McGough, Z. Pourmirza, M. Pazhoohesh, and S. Walker, "Machine Learning, Deep Learning and Statistical Analysis for forecasting building energy consumption—A systematic review," *Engineering Applications of Artificial Intelligence*, vol. 115, p. 105287, 2022.
- [46] D. Xu, Q. Zhang, Y. Ding, and D. Zhang, "Application of a hybrid ARIMA-LSTM model based on the SPEI for drought forecasting," *Environmental Science and Pollution Research*, vol. 29, no. 3, pp. 4128-4144, 2022.
- [47] M. Elsaraiti and A. Merabet, "A comparative analysis of the arima and lstm predictive models and their effectiveness for predicting wind speed," *Energies*, vol. 14, no. 20, p. 6782, 2021.
- [48] K. Ullah et al., "Short-Term Load Forecasting: A Comprehensive Review and Simulation Study With CNN-LSTM Hybrids Approach," *IEEE Access*, 2024.
- [49] S. Khan, "Application of Deep Learning LSTM and ARIMA Models in Time Series Forecasting: A Methods Case Study analyzing Canadian and Swedish Indoor Air Pollution Data," *Austin J Med Oncol*, vol. 9, no. 1, p. 1073, 2022.
- [50] M. I. A. Efat et al., "Deep-learning model using hybrid adaptive trend estimated series for modelling and forecasting sales," *Annals of Operations Research*, vol. 339, no. 1, pp. 297-328, 2024.

Marine Predator Algorithm and Related Variants: A Systematic Review

Emmanuel Philibus¹, Azlan Mohd Zain², Didik Dwi Prasetya³, Mahadi Bahari⁴, Norfadzlan bin Yusup⁵,
Rozita Abdul Jalil⁶, Mazlina Abdul Majid⁷, Azurah A Samah⁸

Department of Computer Science, Universiti Teknologi Malaysia, Johor Bahru, Malaysia^{1, 2, 8}

Department of Computer Science, Kaduna State College of Education, Gidan Waya, Kafanchan, Nigeria¹

Department of Electrical Engineering and Informatics, State University of Malang, Malang, Indonesia³

Department of Information Systems, Universiti Teknologi Malaysia, Johor Bahru, Malaysia⁴

Department of Software Engineering, Universiti Malaysia Sarawak, Kota Samarahan, Malaysia⁵

Department of Software Engineering, Universiti Tun Hussein Onn Malaysia, Batu Pahat, Malaysia⁶

Centre for Artificial Intelligence & Data Science, Universiti Malaysia Pahang Al-Sultan Abdullah, Kuantan, Malaysia⁷

Abstract—The Marine Predators Algorithm (MPA) is classified under swarm intelligence methods based on its type of inspiration. It is a population-based metaheuristic optimization algorithm inspired by the general foraging behavior exhibited in the form of Levy and Brownian motion in ocean predators supported by the policy of optimum success rate found in the biological relationship between prey and predators. The algorithm is easy to implement and robust in searching, yielding better solutions to many real-world problems. It is attracting huge and growing interest. This paper provides a systematic review of the research progress and applications of the MPA by analyzing more than 100 articles sourced from Scopus and Web of Science databases using the PRISMA approach. The study expounded the classical MPA's workflow. It also unveiled a steady upward trend in the use of the algorithm. The research presented different improvements and variants of MPA including parameter-tuning, enhancement of the balance between exploration and exploitation, hybridization of MPA with other techniques to harness the strengths of each of the algorithms towards complementing the weaknesses of the other, and more recently proposed advances. It further underscores the application of MPA in various areas such as Engineering, Computer Science, Mathematics, and Energy. Findings reveal several search strategies implemented to improve the algorithm's performance. In conclusion, although MPA has been widely accepted, other areas remain yet to be applied, and some improvements are yet to be covered. These have been presented as recommendations for future research direction.

Keywords—Exploitation-exploration; marine predator algorithm; metaheuristic algorithms; metaheuristic-hybridization; meta-heuristics; optimization; predator prey systems

I. INTRODUCTION

There is a proliferation of optimization methods for finding optimum solutions to engineering, scientific, real-world, and social problems [1, 2]. This is necessitated by the corresponding increase in complex optimization problems that require

solutions. These methods can broadly be classified into deterministic and stochastic methods (Fig. 1). The deterministic methods can be further classified into gradient-based and non-gradient-based methods. For instance, mathematical linear and non-linear programming methods are all gradient-based since they rely on gradient computation to locate global solutions. Conversely, non-gradient-based deterministic methods use direct algorithms, conditions, and static, and dynamic data structures instead of gradients to compute the global optimum solution [3–6].

One prevailing limitation of mathematical programming methods includes greater chances of local optima stagnation while searching in non-linear space. As such, researchers have used different initial designs, hybridization, and modifications to overcome the drawbacks. This, however, makes the solution problem specific. Non-gradient deterministic methods possess weaknesses including difficult implementation and require a deep knowledge of mathematics before application.

One of the ways by which researchers address the drawbacks of the deterministic methods is by exploring alternatives from the stochastic approaches. The popular stochastic method in use is metaheuristics [1, 7] which uses random variables and operators to perform a global search while trying to avoid being trapped in local optima. Metaheuristic algorithms are now being applied in several research fields such as business management, medical imaging, environmental studies, engineering design, mathematics, robotics, image segmentation, etc., changing the trends and the look and feel of the research world. These methods are simple and easy to understand and implement. However, they do not guarantee a global solution despite possessing outstanding qualities such as being gradient-free, problem-independent, adaptable, and near-global solutions over other optimization methods.

This work was supported in part by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme (FRGS) [FRGS/1/2022/ICT02/UTM/01/1]. Thank you to the Research Management Center and Faculty of Computing, Universiti Teknologi Malaysia (UTM) for thorough research support

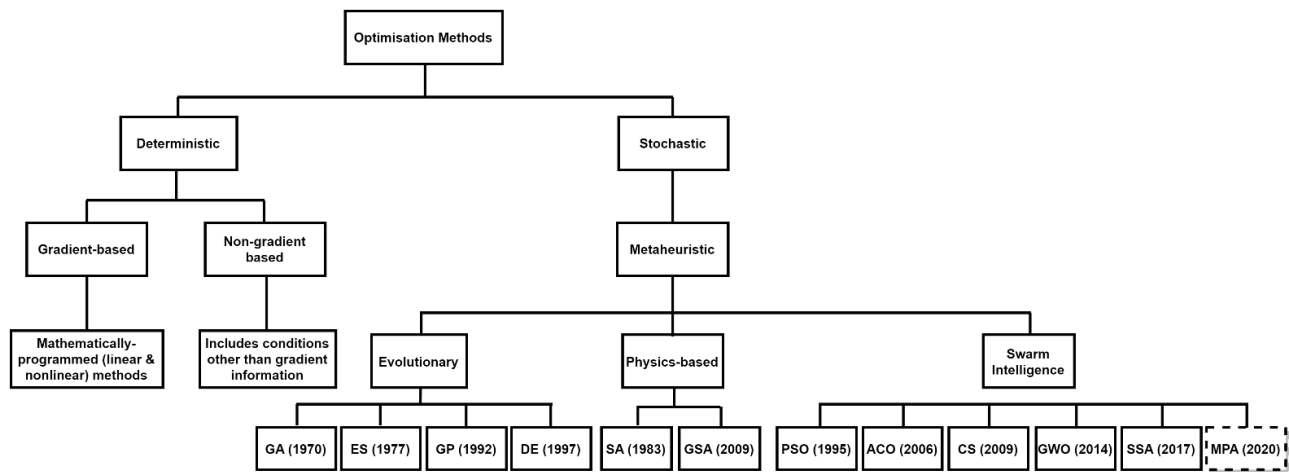


Fig. 1. Category of optimization algorithms featuring metaheuristics.

Metaheuristic methods can be grouped into three groups based on their type of inspiration (Fig. 1). These are evolutionary algorithms, physics-based, and swarm intelligence methods.

Evolutionary algorithms are the oldest form of metaheuristics, grouped based on biological interaction within the space of nature. In this group, the earliest method proposed in the 1970s was the Genetic Algorithm (GA) [8]. GA is hinged on two biological concepts: mutation and cross-over, used in domain search and improvement of initialized random populations. Other popular algorithms proposed by this group about the same time include Evolution Strategy (ES) in 1977 [9], Genetic Programming (GP) in 1992 [10], Differential Evolution (DE) in 1997 [11], etc.

The second group of metaheuristic methods classified in this study is physics-based. In this group, inspiration is drawn from various laws of physical nature. The search for optimal solutions is strictly based on the laws of physics. Inspired by the laws of thermodynamics, the oldest popular method first proposed under this group is Simulated Annealing (SA) in 1983 [12]. A Gravitational Search Algorithm (GSA) was later proposed in 2009 [13] which is based on Newton's law of masses gravity and interaction as a way of position update to search for the optimum solution. Swarm intelligence is the third group of these metaheuristic approaches in this study. In this group, the algorithms imitate a set of behaviors found in flocks, swarms, schools, and herds of several natural creatures. The first method proposed in this group was Particle Swarm Optimization (PSO) in 1995 [14]. PSO is an optimization algorithm inspired by the behavior of schools of birds or fish. Subsequent algorithms proposed in this group include Ant Colony Optimization (ACO) in 2006 [15], Cuckoo Search (CS) in 2009 [16], Grey Wolf Optimizer in 2014 [17, 18], Salp Swarm Algorithm (SSA) in 2017 [19], and Marine Predator Algorithm (MPA) in 2020 [1] to mention a few.

The Marine Predators Algorithm (MPA) is a population-based metaheuristic optimization algorithm inspired by the general foraging behavior exhibited in the form of Levy and Brownian motion in ocean predators supported by the policy of optimum success rate found in the biological relationship between prey and predators [1]. MPA is characterized by being simple in implementation and robust in solution search yielding

better solutions to many real-world problems [20]. It is swarm-based, a relatively new algorithm introduced in 2020 by Faramarzi and his team, and it is attracting huge and growing interest. The algorithm was originally proposed for use in engineering and mathematical problems. However, due to its high performance and search success, it has gained wide acceptance, and it has been applied in several domains. It uses two motions: Levy flight and Brownian motions to perform a search for local or global solutions. The strategies employed by MPA for use in different situations as originally proposed by [1] are:

- When the search encounters sparsely populated prey, MPA applies the Levy flight grazing strategy and later changes to Brownian motion when a crowded population of prey is detected.
- In addition to the swift fluctuation of the hunting strategy, the predators transform their actions towards finding locations with more crowded prey.
- The predators are too smart in retention of visited locations, keeping the memory to provide information that could help other predators when needed.
- Being easy to implement, possessing fewer parameters, and yielding good results, MPA has taken over the metaheuristic space as seen in the literature.

The MPA, introduced in 2020 and utilized across various domains, faces challenges associated with exploration-exploitation imbalance common among intelligent algorithms [21–23]. In addition, weaknesses such as poor solution quality, easily trapped in local optima, and slow convergence speed have been noticed. Consequently, many researchers have proposed various improvements and variants of the algorithm through parameter tuning, hybridization, and enhancements (modifications). Among these include a hybridization of Improved MPA and PSO known as IMPAPSO [24], enhanced MPA (EMPA)[25], four new variants of MPA: (i) multi-objective MPA (MMPA) (ii) modified MMPA (M-MMPA) (iii) Gaussian-based mutation M-MMPA (M-MMPA-GM), and (iv) Nelder-Mead simplex technique into M-MMPA (M-MMPA-NMM)[2], Three-scale image decomposition (TSD), Kirsch

compass operator (FR-KCO), and MPA (TSD-FR-KCO-MPA) [26], Local Escaping Operator MPA (LEO-MPA) [20], opposition based learning MPA and grey wolf optimization (MPOBL-GWO) [27], Tuned-MPA [28], a hybrid method that combines MPA with Fuzzy Proportional-Integral-Derivative with Filter (FPIDF) (MPA-FPIDF) [29], Boost MPA (BMPA)[30], combining the MPA with CNN (IMPA-CNN)[31], a modified version of MPA known as MMPA[32], MPALS and HMPA [33], modified type of MPA (MMPA)[34], hybrid MPA-Support Vector Machine (MPA-SVM) [35], MPA to optimize a trained ANN (MPA-ANN) [36], an improved MPA and ResNet50 (IMPA-ResNet50) [37], MPA and Proportional-Integral-Derivative-Acceleration (PIDA) (MPA-PIDA)[38], advanced MPA (AMPA) [39], MPA and multi-verse optimization algorithm (MPA-MVO)[40], Learning-Automata (LA)-based Jellyfish search MPA (LA-JS-MPA) [41], fractional-order comprehensive learning MPA (FOCLMPA) [42], Fusion Multi-Strategy Marine Predator Algorithm (FMMPA) [43], reinforcement learning (RL) and MPA (Deep-MPA)[44], MPA and naked mole-rat algorithm (NMRA)(MpNMRA) [45], Dynamic Foraging Strategy MPA (DFSMPA) [46], MPA with mechanism for teaching and learning (MTLMPA) [47], diversity-aware MPA (DAMPA) [48], MPA, modified conformable fractional-order accumulation operation (MCFAO) [49], two variants: BBD-based MPA, and CCD-based MPA [50], an enhanced version of the MPA (EMPA)[51], an enhanced multi-strategy MPA - Variational Mode Decomposition (MPA-VMD) [52], Tuned-MPA proportional-integral-derivative proportional derivative (PID-PD) controller [53], Open Circuit Voltage MPA (OCV-MPA) [54], Marine Predator Algorithm and Hide Object Game Optimization (MPA-HOGO) [55], multi-stage improvement of the MPA (MSMPA)[56], etc. This study presents an extensive review of MPA and its variants based on improvements. It analyzes its strengths and improvements and provides future research directions. The major contributions of this study can be summarized as follows:

- A detailed and clear explanation of the workflow of the classical MPA including a flowchart and pseudocode is provided, see Section II.
- A steady upward trend in MPA has been revealed based on some qualitative statistics of the articles published over the years, see Section III.
- The review highlights MPA's uniqueness based on the predator's ability to execute various movements corresponding to the prey's behavior.
- Several variants of the MPA have been presented which are made up of various search improvement strategies, see Section VI.

The rest of this paper is organized as follows: Section II presents the standard MPA with its source of inspiration, major components, and flowchart steps. Section III presents the materials and method used in this research, where the PRISMA approach is highlighted. Section IV discusses proposed variants of MPA for performance improvement. Section V showcases the application of MPA in different areas. Furthermore, Section VI gives supporting discussions. Section VII presents future

research directions. Finally, Section VIII presents the conclusion of the entire research work.

II. STANDARD MPA

The MPA is a population-based metaheuristic optimization algorithm inspired by the general foraging behavior exhibited in the form of Levy and Brownian motion in ocean predators supported by the policy of optimum success rate found in the biological relationship between prey and predators [1]. The algorithm was objectively proposed for use in engineering and mathematical problems.

It is a popular fact that the entire search strength of every metaheuristic algorithm is measured in three characteristics: exploration, exploitation, and the ability to escape local minimum/optima [57]. Exploitation serves as the main ability of the algorithm to search for every nearby detail while exploration ensures that the algorithm completes its search of the entire search space. The MPA uses two motions, Levy flight, and Brownian motions to search for local or global solutions. Because Levy flight is associated with mostly short steps, it is well suited to local search or exploitation. However, the Brownian motion on the other hand is associated with larger step sizes and hence it is suitable for global search or exploration. Either of these two motions alone cannot be sufficient in performing a search, and therefore the two are combined to improve the searchability of MPA. The algorithm is unique and widely acceptable compared to other metaheuristic algorithms due to its search strategies and memory recall as proposed by [1].

Based on its similarities to other metaheuristic algorithms, the MPA begins by defining an initial uniform population distribution of solutions in the search space based on trial using Eq. (1).

$$X_0 = X_{\min} + \text{rand}(X_{\max} - X_{\min}) \quad (1)$$

Here, X_{\min} and X_{\max} are referred to as the lower and upper bound variables, respectively, while rand is the uniform random vector of a range 0 to 1.

Next, a matrix of top predators also called Elite is created based on the generated distribution in Eq. (1). Additionally, top predators according to the survival of the fittest theory are more gifted at foraging. Therefore, a matrix of top predators is constructed that serves as a tentative solution known as Elite. This matrix's array supervises the search for locating prey relative to its available information or address as given in Eq. (2).

$$\text{Elite} = \begin{bmatrix} X_{1,1}^1 & X_{1,2}^1 & \cdots & X_{1,d}^1 \\ X_{2,1}^1 & X_{2,2}^1 & \cdots & X_{2,d}^1 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ X_{n,1}^1 & X_{n,2}^1 & \cdots & X_{n,d}^1 \end{bmatrix}_{n \times d} \quad (2)$$

Here, \vec{X}^1 denotes a vector of the top predator that is duplicated 'n' times to form the Elite matrix, 'n' is known as the number of search agents, and 'd' represents the dimensions. Every predator is a search agent and a potential prey as they both search for

food. When each iteration is completed, the Elite is updated where better predators replace top predators.

Furthermore, a second matrix of the same size as the Elite matrix known as Prey is formed and its predators' addresses are updated according to the Elite's as depicted in Eq. (3), where $X_{i,j}$ denotes the j th dimension of i th prey. MPA depends on these two matrices throughout its iterations.

$$\text{Prey} = \begin{bmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,d} \\ X_{2,1} & X_{2,2} & \dots & X_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n,1} & X_{n,2} & \dots & X_{n,d} \end{bmatrix}_{n \times d} \quad (3)$$

A. The Core of the MPA Optimization Process and Modeling

Based on the proposed model by [1], the optimization workflow of the MPA goes through three conditions while imitating the entire life of predators and prey. These three phases are split across three levels of velocity scenarios experienced by these aquatic creatures:

Condition 1: "When the speed of the predator gets faster than that of the prey" (high-velocity ratio).

Condition 2: "When the speed of the predator becomes almost equal to that of the prey" (unit velocity ratio).

Condition 3: "When the speed of the predator becomes slower than that of the prey" (low velocity ratio).

When condition 1 holds, this implies that the velocity ratio is high ($V \geq 10$), and consequently, the algorithm applies the best strategy for the predator which is to stand still without any movement. This approach [1] is expressed and modeled mathematically by Eq. (4).

From Eq. (4), $\overrightarrow{Prey}_l = (\overrightarrow{Prey}_l + P \cdot \vec{R} \otimes \overrightarrow{stepsize}_l)$ where \vec{R}_B is a vector containing the Brownian motion's normal distribution of random numbers.

While $Iter < \frac{1}{3} \text{Max_Iter}$ then,

$$\overrightarrow{stepsize}_l = \vec{R}_B \otimes (\overrightarrow{Elite}_l - \vec{R}_B \otimes \overrightarrow{Prey}_l), l = 1, \dots, n \quad (4)$$

The operator \otimes is an element-wise product. Computing the product of \vec{R}_B by prey simulates the prey's movement. The symbol $P=0.5$ is a constant control parameter that minimizes/maximizes predator or prey's step sizes, while R denotes a vector of uniform random numbers in the range $[0, 1]$. The first condition's scenario occurs at one-third of the entire iterations where the step size is high due to the high velocity of movement toward achieving high exploration. The variable $Iter$ represents the current iteration while Max_Iter stands for the maximum iteration.

Next, when condition 2 occurs, that is, the predator and prey are moving at almost the same velocity (unit velocity ratio i.e., $V \approx 1$), depicting a scenario where both are searching for their

food, the algorithm tries to detect the type of motion used by each. At this point, if the prey is moving in Levy motion, the predator's best approach becomes switching to Brownian motion. The situation happens in the middle of the optimization process when exploration attempts to switch to exploitation. Thus, both behaviors would matter, and half of the population would be assigned to exploration while the other half would be assigned to exploitation. Assuming the prey and predator are moving in Levy and Brownian motions, respectively, this can be represented or modeled mathematically by Eq. (5 and 6):

While $\frac{1}{3} < \text{Max_Iter} < \frac{2}{3} \text{Max_Iter}$, then

Considering the first half of the population,

$$\begin{aligned} \overrightarrow{stepsize}_l &= \vec{R}_L \otimes (\overrightarrow{Elite}_l - \vec{R}_L \otimes \overrightarrow{Prey}_l), l=1, \dots, \frac{n}{2} \\ \overrightarrow{Prey}_l &= \overrightarrow{Prey}_l + P \cdot \vec{R} \otimes \overrightarrow{stepsize}_l \end{aligned} \quad (5)$$

From Eq. (5), \vec{R}_L is a vector containing the Levy motion's distribution of random numbers. The vector \vec{R}_L and $Prey$ are multiplied to simulate the Levy-wise movement of the $Prey$. The step size is added to the location of the $Prey$ to complete this simulation.

On the other hand, the assumption for the second half of the population is thus:

$$\begin{aligned} \overrightarrow{stepsize}_l &= \vec{R}_B \otimes (\vec{R}_B \otimes \overrightarrow{Elite}_l - \overrightarrow{Prey}_l), l=\frac{n}{2}, \dots, n \\ \overrightarrow{Prey}_l &= \overrightarrow{Elite}_l + P \cdot CF \otimes \overrightarrow{stepsize}_l \end{aligned} \quad (6)$$

Where $CF = (1 - \frac{Iter}{Max_Iter})^{(2 \frac{Iter}{Max_Iter})}$ represents an adaptive parameter that regulates the step size of a predator's movement. The vector \vec{R}_B is multiplied with the $Elite$ to mimic the movement of the predator Brownian-wise and update the prey's location based on the predator's Brownian-wise movement.

Condition 3 occurs when the velocity of the predator becomes slower than that of the prey (low-velocity ratio, usually $V = 0.1$). This scenario usually occurs at the final phase of the optimization workflow, and it commonly targets high exploitation performance. The best option for the predator at this point is Levy's motion. This scenario is modeled mathematically by Eq. (7):

$$\begin{aligned} \text{While } Iter > \frac{2}{3} \text{Max_Iter} \\ \overrightarrow{stepsize}_l &= \vec{R}_L \otimes (\vec{R}_L \otimes \overrightarrow{Elite}_l - \overrightarrow{Prey}_l) \quad l=1, \dots, n \\ \overrightarrow{Prey}_l &= \overrightarrow{Elite}_l + P \cdot CF \otimes \overrightarrow{stepsize}_l \end{aligned} \quad (7)$$

where the vector \vec{R}_L is multiplied with the $Elite$ to simulate the predator's movement in Levy form and adding the step-size to the location of the $Elite$ to mimic the predator's movement aids in updating the location of the prey. The algorithm is presented thus (Algorithm 1):

Algorithm 1: Standard MPA Pseudocode

Step 1: **Initializing Phase**
(1) Initialize the parameters of the algorithm (Population size, dimensions, maximum Iterations)
(2) Uniformly distribute the initial solution using *Equation 1*.

Step 2: **Evaluation Phase**
(3) **while** (the termination condition does not satisfy)
(4) **Evaluate** the fitness of the solutions

Step 3: **Construction Phase**
(5) **Construct** the **Elite matrix** using *Equation 2*
(6) **Construct** the **Prey matrix** using *Equation 3*

Step 4: **Optimisation Phase**
Stage 1: **High Velocity Ratio**
(7) if ($Iter < \frac{1}{3} Max_Iter$) then
(8) **Update Prey** using *Equation 4*
Stage 2: **Unit Velocity Ratio**
(9) Else if ($\frac{1}{3} Max_Iter < Iter < \frac{2}{3} Max_Iter$) then
(10) Considering the first half of the population ($l = 1, \dots, \frac{n}{2}$)
(11) Update Prey using *Equation 5*
(12) For the second half of the population ($l = \frac{n}{2}, \dots, n$)
(13) Update Prey using *Equation 6*
Stage 3: **Low Velocity Ratio**
(14) Else if ($Iter > \frac{2}{3} Max_Iter$) then,
(15) Update Prey using *Equation 7*
(16) **end if**
(17) **end if**
(18) **end if**

Step 5: **Update Phase**
(19) **Update** the **Elite matrix** and save it in memory.
(20) **Apply the FADs** effect, then update using *Equation 8*
(21) Further, **Update the Elite matrix** and update the memory.
(22) **end while**

Overall, the steps proposed by [1] imitate the movement of predators and prey when seeking food in aquatic habitats. Their work assumes that there is an equal percentage of Levy and Brownian motion over the lifetime of a predator.

Because Fish Aggregating Devices (FADs) influence the time taken by predators at a particular place and point in time,

$$\vec{Prey}_l = \begin{cases} \vec{Prey}_l + CF[\vec{X}_{min} + \vec{R} \otimes (\vec{X}_{max} - \vec{X}_{min})] \otimes \vec{U} & \text{if } r \leq FADs \\ \vec{Prey}_l + [FADs(1 - r) + r](\vec{Prey}_{r1} - \vec{Prey}_{r2}) & \text{if } r > FADs \end{cases} \quad (8)$$

From Eq. (8), FADs are assigned the value 0.2 (i.e., FADs = 0.2) which is defined as the probability of its effect in the optimization process, and \vec{U} denotes a binary vector that contains arrays inclusive of zero and one. The array is formed by first generating random numbers in [0, 1] and thereafter, transforming it such that the array becomes zero if it is less than 0.2 and one otherwise. The parameter r represents a uniform number in [0,1]. \vec{X}_{max} and \vec{X}_{min} are vectors that contain upper and lower limits respectively of the dimensions. The subscripts r1 and r2 represent the prey's matrix indexes [1].

e.g., sharks spend 80% of the time around them and 20% at other places, the attraction by FADs is creating a local optimum and their jump to search other places is seen as avoidance of being trapped. The effect of FADs is therefore modelled mathematically as follows:

The MPA as depicted in 'Algorithm 1' and Fig. 2, has good provision for memory tracking and recalling. This helps the predator remember foraging success from the places it has visited. The flow requires updating the prey, applying the FADs effect, and evaluating the matrix for possible fitness updates for the Elite. At each stage of the iteration, the fitness value is compared with that of the previous iteration, and the best overwrites the current solution. This continually refines the quality of the solution as each iteration elapses.

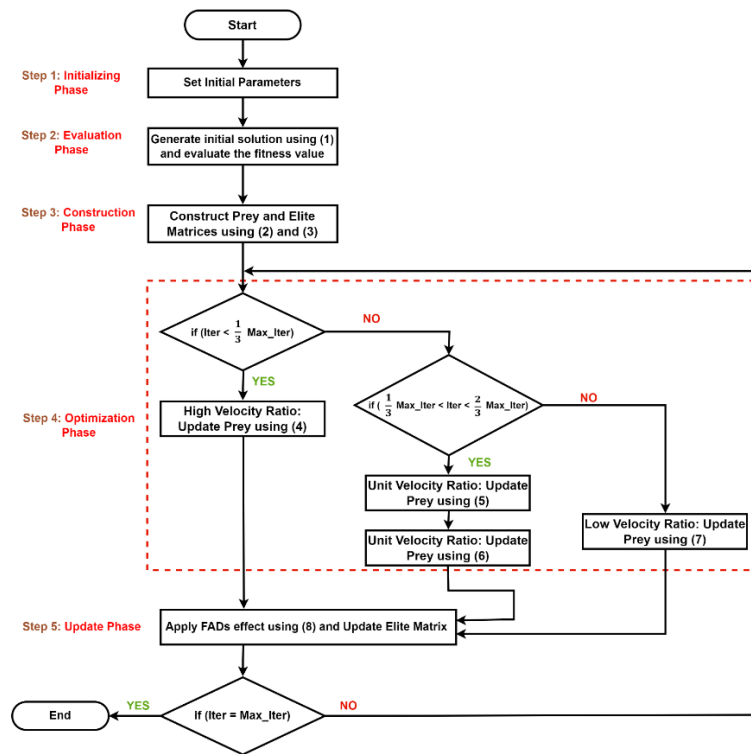


Fig. 2. MPA flowchart.

III. MATERIALS AND METHOD

This study applies the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) approach [58] in searching, collecting, synthesizing, and analyzing a systematic literature review (SLR) of the original MPA, proposing related modifications, and variants according to some selected articles. The study uses two databases: Scopus and Web of Science, and an additional database: Google Scholar (for verification purposes only).

First, to validate the proposed topic, a search for the terms “Marine Predator Algorithm and Related Variants: A Systematic Review” was carried out which gave no single result from the Scopus database. A similar search was also conducted with the same search string in the Web of Science database, and it also did not produce a result. Furthermore, a search for the exact match of the same title on Google Scholar yielded no results.

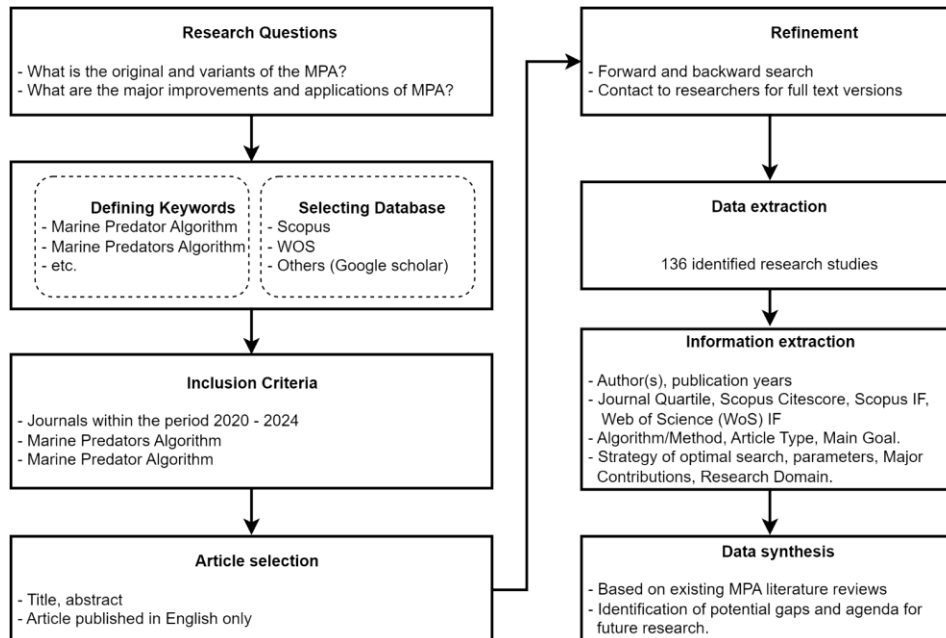


Fig. 3. The complete SLR process.

In Fig. 3, the complete SLR process is presented beginning with the formation of the research questions until the final data synthesis. It is important to note that the extraction of information from the articles was limited to the name of the author(s), publication years, journal quartile, Scopus CiteScore, Scopus IF according to Journal Indexed by Thomson Reuters (Clarivate Analytics), Web of Science IF, the algorithm or method used by the author(s), the article type (i.e., experimental result or review), main goal of the research, strategy for optimal search, parameters, major contributions, and the research domain.

Secondly, a careful search string was constructed to obtain relevant information and related articles (Fig. 4). An advanced search of the Scopus database using the constructed search string yielded 140 documents and a similar search conducted on the Web of Science database gave 143 papers as of 1st August 2024. The two search results were combined, and duplicate records were removed, reducing the document size to 170. A title-abstract screening was conducted where 11 articles were further excluded based on relevance and 1 other article was excluded, being written in Chinese language. Furthermore, 22 articles were excluded due to lack of full access. Overall, 136 articles were used in the entire synthesis of this review process.

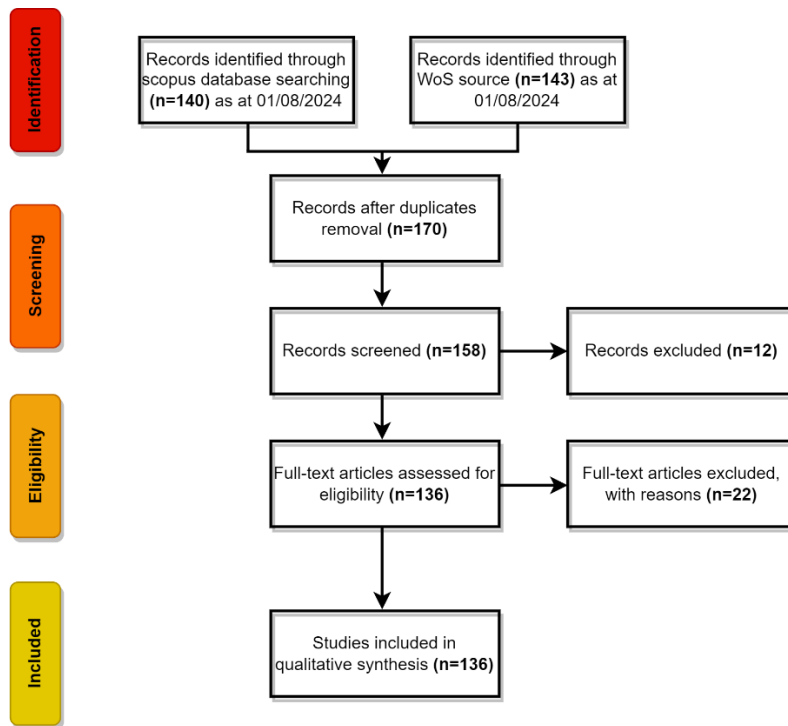


Fig. 4. Articles search and screening process.

The pie chart in Fig. 5 presents the diagrammatic distribution of retrieved MPA-related articles according to subject areas based on Scopus data. The top five subject areas are

Engineering, Computer Science, Mathematics, Energy, and Material Science, with 28.2%, 26.0%, 10.9%, 7.1%, and 5.8%, respectively.

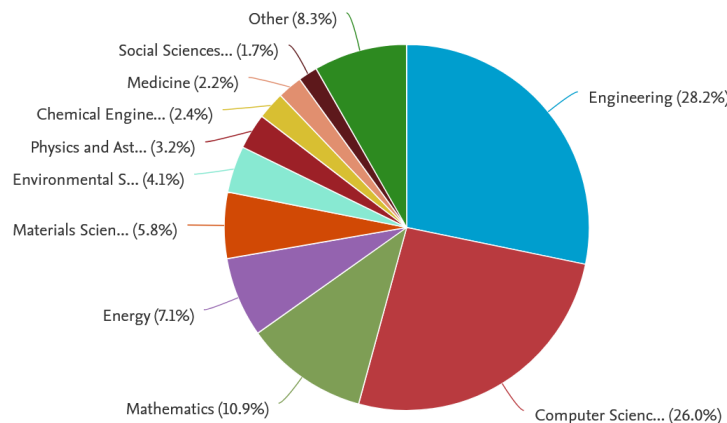


Fig. 5. MPA-related articles according to subject areas (Source: Scopus).

The research trend of the application of MPA for solving various problems is depicted based on the number of research articles that are published per year (Fig. 6.). The record shows a steady upward trend in the number of articles published over the years, reflecting a strong growth in the algorithm's usage. Beginning with 12 journal articles in 2020, the number progressively increased to 168 by 2023, showcasing a 93% rise

over the observed period. As of the search date, the record of published articles in 2024 was 148 while still counting. More publications are underway as the year progresses. This significant growth highlights the major rapid application of the algorithm, indicating positive acceptance and potential usage for further development and innovation.

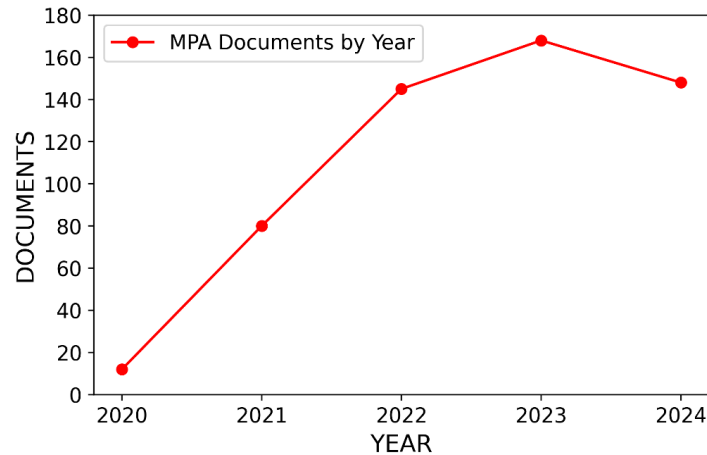


Fig. 6. MPA-related publications by year (Source: Scopus).

MPA stands out from other algorithms due to the predator's ability to execute various movements based on prey behavior. The predator can opt for Levy motion or Brownian motion based on the best encounter strategy, ensuring a dynamic connection between predator and prey. Specifically, the predator employs the Levy strategy when prey density is low and switches to Brownian motion when prey density is high.

As a metaheuristic algorithm, MPA is expected to meet some requirements of the major characteristics that measure its ability to solve optimization problems which include the ability to handle exploration, exploitation, local optimums, and convergence rate [41]. Each metaheuristic method differs in the way it does this based on the nature of the problem under consideration.

Three control parameters determine the sensitivity of MPA. The first is fish aggregating devices (FADs), which control the effect of FADs alongside their influence on the optimization flow. Secondly, P minimizes/maximizes predator's or prey's step sizes. Adjusting the step sizes in MPA helps to regulate its exploration and exploitation. The third parameter is the control factor (CF), which is an adaptive parameter that regulates the step size of a predator's movement. In study [1], it was found that the parameter ' P ' became more sensitive than FADs to optimizing some unimodal functions. However, in multimodal functions, FADs gave higher performance. In some instances, the parameters presented no sensitivity.

IV. PROPOSED VARIANTS OF MPA FOR PERFORMANCE IMPROVEMENT

A. Parameter-tuned MPA

One of the earliest approaches adopted by researchers towards improving the performance of MPA was the application of parameter tuning. It is a general approach used in

optimization and machine learning modeling to obtain optimum parameter values. The process requires tweaking some set of parameters used in controlling the behavior of the model/algorithm that are also adjustable to obtain an improved model with optimal performance.

In the original MPA, all population position updates are influenced by a constant value, denoted as P . As per the described position updating equations of MPA, there is a risk of premature convergence during the optimization iteration, limiting the exploration of the entire search space. Additionally, the alternation between Brownian and Lévy motions in the optimization process may lead to significant steps, causing the optimal solution to be crossed. Instead, dynamic updates of population positions could be achieved by incorporating other approaches such as the sine and cosine functions to improve the MPA's performance [32].

Some researchers such as [28] tried to improve the performance of MPA through a tuning process. Their study looked at the three most sensitive aspects of performance control of MPA which included the way the iterations are distributed across the iterations' phases, the size of the population in the second phase, and the effect of FADs. The experiment first tested different values of the iterations on each phase of the algorithm's optimization process e.g., allocating one-third of the iterations to each phase and later changing it produced little improvement in the results. The tests reveal that the least cost of optimal power flow (OPF) optimization in IEEE 48-Bus using MPA can be obtained in the first phase of the algorithm at three-fifths, phase two at one-fifth, and phase three at one-fifth of the iterations, respectively.

Concerning population size, the tasks of exploration and exploitation in MPA ideally require splitting the entire population size into two halves. However, it is important to note

that some optimizations' minimum can be achieved when the population is divided into two-thirds and one-third for the prey and predators, respectively [28]. While maintaining the classical fact that FAD or eddy current effect in MPA is meant to keep the iteration from being trapped at the local minimum, however, the mathematical representation shows that FADs are also local minima. Therefore, a search could be conducted to obtain the optimal value of FADs as in [28], starting with an initial value of 0.2 until better performance was obtained at FAD = 0.3. This shows that tuning the value of FAD in the IEEE 48-Bus system could yield better performance. After tuning and obtaining the optimal parameter values for the iterations' distribution, population size, and FADs for the MPA, an experiment was set up in two folds: holistic and inter-bounded OPF and ultimately comparing the performance of the tuned-MPA with GA. Overall findings showed that the tuned MPA outperformed GA in convergence, accuracy, and computational requirements. Furthermore, the holistic fold produces better solutions and requires higher computational power. While the inter-bounded

OPF generates faster results and is less computationally intensive.

Another similar MPA parameter-tuning case is found in [53], which in a bid to obtain the optimum load frequency control (LFC) settings to create a balance between power generation and demand, proposed a novel PD-P-PID cascade controller for LFC applications, utilizing the MPA for optimal parameter tuning. Tested on various power systems, including single and multi-area setups, the MPA-tuned PID-PD controllers exhibited superior performance compared to existing literature. The controller's robustness was evaluated on various power systems, and its parameters were optimized using MPA, demonstrating superior performance in terms of settling time and oscillations in frequency and tie-line power deviation compared to existing works. Their findings underscore the effectiveness of the MPA-tuned PD-P-PID controller in LFC applications. Table I provides more related works on MPA hyperparameter tuning where various degrees of success were achieved using parameter tuning.

TABLE I. RELATED WORKS ON MPA HYPERPARAMETER-TUNING

Ref.	J. Quart.	Scopus CiteScore (2022)	Scopus IF	WOS IF	Algorithm	Article Type	Strategies for optimal search	Major Contributions	Research Domain
[28]	Q1	9.0	4.342	3.9	Tuned-MPA	Experiment/Result	Parameter tuning	Tuning was done to obtain the optimal parameter values for the iteration distribution, population size, and FADs for the MPA.	Optimal power flow
[26]	Q1	8.2	5.861	3.88	TSD-FR-KCO-MPA	Experiment/Result	Levy and Brownian	MPA was used to obtain optimum parameters	Medical image fusion
[40]	Q4	2.0	1.771	0.5	MPA, MVO	Experiment/Result	Levy and Brownian	MPA is used to solve the final optimization problem	Electric vehicles
[59]	Q3	3.0	3.536	2.0	MPA	Experiment/Result	barrier parameters' influence	Incorporating barrier parameters influence, MPA is compared with GWO, and EO for effectiveness	Transformer oil breakdown
[60]	Q2	5.6	5.127	6.87	MPA	Experiment/Result	Levy and Brownian motion	MPA was used to obtain optimum parameters	Wind renewable energy
[61]	Q1	3.5	2.4	2.4	MPA	Experiment/Result	MPA is combined with the principle of key-term separation	MPA is combined with the principle of key-term separation	Mathematical computation
[62]	Q1	10.0	5.599	5.606	MPA	Experiment/Result	Levy and Brownian motion	compares several metaheuristic optimization algorithms that are used as frameworks for optimization	Selective harmonic elimination
[7]	Q1	14.1	9.177	9.7	MPA	Review			Microgrid, feature selection, etc.
[63]	Q2	4.7	3.271	2.9	MPA	Experiment/Result	Seven robust battery models were proposed for Lithium-ion batteries.	MPA is used as an optimizer of the objective function for the proposed seven models.	Li-ion batteries
[53]	Q2	7.7	4.203	4.1	Tuned-MPA PID-PD	Experiment/Result	Tuning	the performance of the Load Frequency Controller (LFC) was greatly improved by using tuned-MPA.	load frequency controller design

B. Improvements in MPA Exploitation-exploration Balance

The balance between exploration and exploitation is crucial in metaheuristic algorithms for effective optimization. The MPA addresses this balance by dynamically adjusting the exploration rate during optimization iterations. This adjustment facilitates a combination of exploration and exploitation, strategically applied at the start and end of the optimization process. MPA utilizes a control factor (CF) as an adaptive parameter to regulate step size for predator movement, contributing to the algorithm's

effectiveness in navigating the search space. Because of the spread of iterations that are partitioned into stages, the search agents in MPA do not have sufficient trials for the search and discovery of spaces and the exploitation of optimal solutions [42].

Some researchers criticized the exploitation and exploration searchability of the classical MPA proposed by study [1]. The study in [45] opined that the step sizes that are randomly

generated by the Levy distribution are large and best suited for exploration. This happens in some instances, probably occasioned by sudden jumps from smaller step sizes to larger ones during the search transition from exploitation to exploration [1]. They further stated that many modifications would be required to improve its exploitation ability. While possessing a convergence factor advantage, the larger steps generated by the Levy motion could jump the global minimum. As such, many of them focused on how to improve this aspect of the algorithm.

One possible solution found in the literature in this aspect is in the work done by [24], which proposed a hybridization of Improved MPA and PSO known as IMPAPSO algorithm for the optimization of the non-linear optimal reactive power dispatch (ORPD) problem. To improve MPA's exploration stage, they replaced the Brownian motion's random walk of the search agents with a high-tailed Weibull distribution. Secondly, the exploitation stage of classical MPA which is in phase 3 was also modified to use either PSO or MPA based on probability, to improve the convergence of the algorithm. The proposed IMPAPSO was evaluated using various test suites including IEEE 30, IEEE 57, and IEEE 118 bus systems. The strength of the proposed algorithm was examined in a rigorous comparison with other methods. Overall, the proposed IMPAPSO yielded an outstandingly high speed of convergence, outperforming its counterparts. The power loss was minimized to 96%, 10%, and 9% in IEEE 30, IEEE 57, and IEEE 118 bus systems, respectively.

Another example is found in study [2], which applies a strategy known as the dominance strategy based on exploration-exploitation (DSEE) to improve search exploitation-exploration. First, the classical MPA was modified to produce a multi-objective MPA (MMPA). Secondly, a strategic technique called dominance strategy based on exploration-exploitation (DSEE) was applied to count the returned dominant solutions in every returned solution, from which exploitation is carried out during the exploitation phase. This version was called M-MMPA. Thirdly, the Gaussian-based approach was incorporated into MPA to produce M-MMPA-GM which is a version that delves deeper into the present to discover better non-dominated solutions. This helps to discover better solutions by taking some distance from the present solution. The fourth version was incorporated with Nelder Mead simplex at the beginning of the optimization phase to build a front that helps MPA realize better solutions within the optimization flow.

Additionally, a multi-stage improvement of the MPA (MSMPA) was proposed by study [56]. MSMPA maintains the multi-stage search advantage and incorporates a linear flight

strategy in the middle stage to enhance predator interaction, especially for those further from the historical optimum, promoting exploration. In the middle and late stages, the search mechanism of PSO is integrated to boost exploitation capabilities, reducing stochasticity and effectively constraining predators from jumping out of the optimal region. Additionally, a self-adjusting weight was employed to regulate convergence speed, achieving a balanced exploration-exploitation capability. The algorithm was tested on various CEC2017 benchmark test functions and three multidimensional nonlinear structure design optimization problems, which demonstrated superior convergence speed and accuracy compared to other recent algorithms.

Furthermore, the study in [45] applied the exploitation ability of NMRA to MPA in a bid to address its poor search exploitation. The authors proposed the hybridization of MPA and a naked mole-rat algorithm (NMRA) named MpNMRA – a self-adaptive algorithm. While retaining all the main parameters of both approaches, the basic part of MPA was attached to the worker stage of NMRA to improve search exploitation and exploration. The MpNMRA converges faster than other algorithms in comparison. The study in [55] applied HOGO to modify the search transition in MPA to gradually shift from exploration to exploitation as iterations progress, utilizing the global most appropriate solution at each iteration.

Other studies that worked on improving the exploitation-exploration of MPA are highlighted in Table II. These include [20] which incorporated local escaping operator (LEO) into classical MPA to tackle poor exploitation and exploration; [27] applied opposition based learning (OBL) strategy with Grey Wolf Optimizer (GWO) into MPA to overcome weaknesses; [43] incorporated MPA with spiral complex path search strategy based on Archimedes' spiral curve for perturbation, expanding the global exploration range and strengthening the algorithm's overall search capabilities; [44] integrated reinforcement learning (RL) into MPA to improve its global searchability; [39] combined chaotic sequence parameter and adaptive mechanism for velocity update to better MPA's exploitation and exploration search; [34] adopted comprehensive learning (CL) approach that improves search and transitioning within exploration and exploitation on MPA; [51] applied ranking-based mutation operator to identify the best search agent, enhancing exploitation capabilities and preventing premature convergence; [46] used dynamic foraging strategy (DFS) to tackle sudden transition between the Levy Flight and Brownian Motion; and [47] incorporated teaching mechanism into MPA's first phase to promote its global search ability. Table II summarizes the major improvements in MPA exploitation and exploration with various major contributions.

TABLE II. IMPROVEMENTS IN MPA EXPLOITATION-EXPLORATION

Ref.	J. Quart.	Scopus CiteScore (2022)	Scopus IF	WOS IF	Algorithm	Article Type	Strategies for optimal search	Major Contributions	Research Domain
[24]	Q3	5.5	3.542	3.2	IMPAPSO	Experiment/Result	high-tailed Weibull distribution and PSO	Exploration is improved by replacing the Brownian motion's random walk of the search agents with a high-tailed Weibull distribution. The exploitation stage in phase 3 is also modified to use PSO or MPA based on probability.	Optimal reactive power dispatch
[64]	Q1	19.1	11.057	10.4	EMPA	Experiment/Result	Differential Evolution (DE) operator	DE operator is integrated into the exploration face of the standard MPA to escape local solution	PV Modelling
[2]	Q1	9.0	4.342	3.367	MMPA M-MMPA M-MMPA-GM	Experiment/Result	multi-objective, dominance strategy based on exploration-exploitation (DSEE), Gaussian-based approach, and Nelder Mead simplex	MMPA adopts classical MPA's search for MOPs, multi-objective modified MPA (M-MMPA) is a modification of the classical MPA to use DSEE strategy search phase for exploration and exploitation, Gaussian-based mutation (GM) was integrated into M-MMPA to have a new model M-MMPA-GM, and Nelder-Mead simple method (NMM) was integrated to M-MMPA-GM to create a front for it to get to a better solution while maintaining the minimum possible time (NMM-M-MMPA-GM).	Engineering design
[26]	Q1	8.2	5.861	3.88	TSD-FR-KCO-MPA	Experiment/Result	Levy and Brownian motion	MPA is combined with two other methods to address some drawbacks faced in medical image fusion that includes loss of edges due to ineffective high-frequency part of the fusion's rules, and low-contrast in fused images	Medical image fusion
[20]	Q1	12.3	8.664	8.038	LEO-MPA	Experiment/Result	Local Escaping Operator (LEO)	LEO is incorporated into classical MPA to tackle poor exploitation and exploration	Engineering design
[27]	Q1	12.6	9.602	8.5	MPAOBL-GWO	Experiment/Result	OBL and GWO	OBL strategy with Grey Wolf Optimizer (GWO) is integrated into MPA to overcome weaknesses.	PV System
[43]	Q2	4.5	3.143	2.7	FMMPA	Experiment/Result	Fusion multi-strategy	MPA is incorporated with a spiral complex path search strategy based on Archimedes' spiral curve for perturbation, expanding the global exploration range and strengthening the algorithm's overall search capabilities	Robot path planning
[44]	Q1	12.3	8.635	8.0	Deep-MPA	Experiment/Result	reinforcement learning (RL)	RL is integrated with MPA to improve its global searchability.	Renewable energy system design
[39]	Q2	6.8	4.352	3.9	AMPA	Experiment/Result	chaotic sequence parameter and adaptive mechanism for velocity update	AMPA combines chaotic sequence parameters and adaptive mechanisms for velocity update to better MPA's exploitation and exploration search	Antenna Signals
[34]	Q1	11.9	7.811	5.431	MMPA	Experiment/Result	Comprehensive Learning (CL) approach	CL approach that improves search and transitioning within exploration and exploitation is used on MPA	Economic emission dispatch.
[51]	Q2	4.7	3.308	3.5	EMPA	Experiment/Result	ranking-based mutation operator	the ranking-based mutation operator is used to identify the best search agent, enhancing exploitation capabilities and preventing premature convergence.	ANN classification
[45]	Q1	12.6	9.602	8.5	MpNMRA	Experiment/Result		the basic part of MPA is attached to the worker stage of NMRA	Engineering Design
[46]	Q1	12.3	8.635	8.0	DFSMPA	Experiment/Result	Dynamic Foraging Strategy (DFS)	DFS is used to tackle sudden transitions between the Levy Flight and Brownian Motion	Real-world engineering
[47]	Q3	3.9	2.393	2.6	MTLMPA	Experiment/Result	Mechanism for Teaching & Learning	teaching mechanism is incorporated into MPA's first phase to promote its global search ability	Engineering Design

C. Hybridization of MPA with other Techniques

Hybridization is the combination of two or more techniques to solve problems. The primary purpose of doing this is to harness the strengths of each approach and use them to complement the weaknesses of the other. Literature has shown that by combining MPA with other algorithms, there could be high-performance improvements. For instance, [45] combined MPA with NMRA with the sole aim of addressing the

limitations of MPA (i.e., poor exploitation) and NMRA (i.e., narrow exploration) while leveraging the strengths of the two. MPA suffers from poor exploitation while NMRA suffers from weak exploration, and they both get into local optimum stagnation due to early or untimely convergence. Therefore, the strengths of MPA (i.e., good exploration) and NMRA (i.e., good exploitation) were used to address their weaknesses and to improve the entire performance. Overall, the authors reported a

significant performance improvement. The proposed hybridization (MpNMRA) was found to be more suitable for lower dimensional problems, even though it also provides satisfactory performance in high dimensional cases.

Another example is found in study [29], which proposed a hybrid method known as MPA-FPIDF, a combination of MPA and Fuzzy Proportional-Integral-Derivative with Filter (FPIDF) to optimize Fuzzy PIDF-LFC to enhance the performance of a hybrid microgrid system, incorporating PV and wind energy sources along with real irradiance and wind speed data, as well as energy storage devices. The MPA was used to optimize the input scaling factors, output gains, and membership function boundaries of the proposed FPIDF controller. The performance of MPA-FPIDF controller is compared with the conventional MPA-PIDF controller and other controllers reported in the literature for the same case study, including PSO-PIDF, COR-PIDF, and COR-FPIDF controllers. In addition, various scenarios are implemented to assess the robustness and sensitivity of the proposed controller to step load perturbations, variations in system parameters, and uncertainties associated with renewable energy sources such as wind speed fluctuations and solar irradiance variations.

In study [44], a hybrid method that combines reinforcement learning (RL) and MPA known as Deep-MPA was proposed to minimize the cost of the microgrid power system. RL was integrated with MPA to improve global searchability. The proposed Deep-MPA design was validated against various algorithms, demonstrating a 6% reduction in energy costs.

Furthermore, the study in [52] proposed an enhanced multi-strategy MPA-Variational Mode Decomposition (MPA-VMD) method for pipeline leakage detection. This was meant to

address the limitations of MPA by focusing on improving convergence speed and avoiding local optima. The enhanced MPA was used to find critical parameters in variational mode decomposition (VMD), and dynamic entropy was employed to select effective modes. The algorithm incorporates strategies like a good point set at the initial population stage to enhance search accuracy. It introduces a nonlinear convergence factor and Cauchy distribution during the search process to optimize the predator step size for better global search capabilities. The method effectively escapes local optima, leading to improved convergence speed.

The studied literature reported diverse hybridizations of MPA with other techniques, yielding various performance improvements (Table III). These include [24] which combined high-tailed Weibull distribution's improved MPA and PSO; [27] which integrated MPA, OBL, and GWO; [29] where MPA was integrated with Proportional-Integral-Derivative-Acceleration (PIDA); [35] coupled MPA and SVM where MPA was used to optimize SVM classifier's hyper-parameters for FS and classification; [36] which hybridized MPA and ANN, where MPA was used to optimize a trained ANN along with its fitness function; [37] which proposed IMPA-ResNet50, an improved version of MPA (IMPA) that was improved using OBL and TL, and ResNet50; [31] combined IMPA and CNN – a modified MPA algorithm for CNN hyperparameter selection, enhancing output performance for classification; [52] combined MPA with variational mode decomposition (MPA-VMD); [54] hybridized Open Circuit Voltage (OCV) reconfiguration model and MPA; [55] integrates MPA with HOGO; [45] integrated MPA with NMRA called MpNMRA, etc. Table III summarizes the major hybridizations of MPA with other techniques.

TABLE III. HYBRIDIZATION OF MPA WITH OTHER TECHNIQUES

Ref.	J. Quart.	Scopus CiteScore (2022)	Scopus IF	WOS IF	Algorithm	Article Type	Strategies for optimal search	Major Contributions	Research Domain
[24]	Q3	5.5	3.542	3.2	IMPAPSO	Experiment/Result	high-tailed Weibull distribution and PSO	Exploration is improved by replacing the Brownian motion's random walk of the search agents with a high-tailed Weibull distribution. The exploitation stage in phase 3 is also modified to use PSO or MPA based on probability.	Optimal reactive power dispatch
[27]	Q1	12.6	9.602	8.5	MPAOBL-GWO	Experiment/Result	OBL and GWO	OBL strategy with Grey Wolf Optimizer (GWO) is integrated into MPA to overcome weaknesses.	PV System
[29]	Q2	9.0	4.342	3.9	MPA-FPIDF	Experiment/Result	Fuzzy Proportional-Integral-Derivative with Filter (FPIDF)	MPA is combined with FPIDF to optimize Fuzzy PIDF Load Frequency Controller (PIDF-LFC) to enhance the performance of a hybrid microgrid system	Microgrid system
[38]	Q1	9.1	6.765	6.8	MPA-PIDA	Experiment/Result	Levy and Brownian motion	MPA is used to optimize the gains of the PIDA controller.	Power modulation
[35]	Q1	11.9	7.415	5.772	MPA-SVM	Experiment/Result	Levy and Brownian motion	MPA was used to optimize the SVM classifier's hyper-parameters for FS and classification.	ligament deficiency detection
[36]	Q2	3.2	2.59	NA	MPA-ANN	Experiment/Result	Levy and Brownian motion	MPA is used to optimize a trained ANN along with its fitness function.	Transistor's design
[37]	Q1	10.0	5.599	5.606	IMPA-ResNet50	Experiment/Result	Transfer Learning and Opposition-Based Learning	OBL is used to improve MPA and TL is used to improve IMPA-ResNet50	Breast cancer diagnosis

Ref.	J. Quart.	Scopus CiteScore (2022)	Scopus IF	WOS IF	Algorithm	Article Type	Strategies for optimal search	Major Contributions	Research Domain
[31]	Q1	12.6	9.602	8.5	IMPA-CNN	Experiment/Result	automating the tuning of hyperparameters in CNN time-delay polynomials are applied to improve the model's performance in prediction	using a modified MPA algorithm for CNN hyperparameter selection, enhancing output performance for classification.	arrhythmia classification
[49]	Q1	11.9	7.811	5.431	MCFAO	Experiment/Result	Incorporates two Response Surface Methodologies (RSMs): Box Behnken Design (BBD) and Central Composite Design (CCD)	MPA is used to optimize the model's hyperparameters	Time series prediction
[50]	Q3	5.0	2.606	2.363	BBD-based MPA CCD-based MPA	Experiment/Result		MPA is used for biological decolorization process parameter optimization on BBD and CCD.	Biological Processes
[52]	Q2	4.8	2.795	4.1	MPA-VMD	Experiment/Result	Good point set in the initial population stage	improvements involves initializing a good point set and enhancing convergence factor (CF) and Cauchy distribution.	Pipeline leakage detection
[54]	Q2	5.4	5.784	4.0	OCV-MPA	Experiment/Result	Open Circuit Voltage (OCV) reconfiguration model and MPA	OCV model is used to measure the internal aging mechanism as influenced by the external factors of the lithium battery capacity decay, while MPA is used to detect the aging mode associate parameters.	Battery aging mechanism
[55]	Q3	2.0	1.347	0.6	MPA-HOGO	Experiment/Result	Hide Object Game Optimization (HOGO)	HOGO modifies the search transition in MPA to gradually shift from exploration to exploitation as iterations progress, utilizing the global best solution at each iteration.	Engineering design
[45]	Q1	12.6	9.602	8.5	MpNMRA	Experiment/Result		the basic part of MPA is attached to the worker stage of NMRA	Engineering Design

D. Proposed MPA Variants

Another way researchers address the limitations found in classical MPA is by modifying one or more aspects of the algorithm to create variants. An example is in study [46], which proposed a soft dynamic transformation to tackle the MPA's tendency to be trapped in local optima during transitioning from Levy Flight to Brownian motion when optimizing real-world problems. The proposed Dynamic Foraging Strategy MPA (DFSMPA) replicates the traditional MPA, imitating the step size taken to grab prey. It then applies the dynamic foraging strategy (DFS) to reach deeper search locations for a complete global, faster, and more efficient search. This could help prevent being trapped. Instead of the usual three phases that are used in classical MPA to mimic the behavior of predator and prey, the DFSMPA uses the continuous model to convert the various phases. In the two phases of exploration and extraction, the continuous model alternates between the search agents.

In addition, the study in [25] developed an enhanced MPA (EMPA) to identify hidden parameters in various PV and static PV models. In their work, the differential evolution (DE) operator was integrated into the exploration face of the standard MPA to escape local solutions for stability and performance reliability in handling nonlinear optimization cases of modeling PV. The strengths of the proposed enhancement are: (i) maintaining various new solutions in the search and optimizing unexpected convergence. (ii) avoiding being trapped by leaders and the population. (iii) using diverse search mechanisms that combine populations to create a balance between exploration and exploitation. (iv) dynamically changing the solutions by the

algorithm to ensure effectiveness and efficiency. (v) dynamic adjustment of the optimization problem and concurrently covering various multi-dimensional areas of the search space.

Furthermore, the study in [51] proposed an enhanced variant of the MPA, called the EMPA, designed for training Feedforward Neural Networks (FNNs). EMPA was intended to minimize classification, prediction, and approximation errors by adjusting connection weights and deviation values. It incorporates a ranking-based mutation operator to identify the strongest search agent, enhancing exploitation capabilities and preventing premature convergence. EMPA combines exploration and exploitation, providing stability and flexibility in achieving optimal solutions. Experimental results on seventeen datasets show that EMPA exhibits faster convergence, higher calculation accuracy, increased classification rates, and strong stability and robustness, improving its productivity and reliability in training FNNs.

Other modifications proposed are presented in Table IV. They include the use of mechanism for teaching and learning to balance search exploitation and exploration [47]; combining chaotic sequence parameter and adaptive mechanism for velocity update to better MPA's exploitation and exploration search [39]; modifying the classical MPA to produce a multi-objective MPA (MMPA), applying a strategic technique called dominance strategy based on exploration-exploitation (DSEE) to count the returned dominant solutions in every returned solution from which exploitation is carried out during the exploitation phase, incorporating Gaussian-based approach into MPA to produce M-MMPA-GM, and incorporating Nelder

Mead simplex at the beginning of the optimization process, building a front that helps MPA realise better solutions within the optimization flow [2]; using a local escaping operator (LEO) to improve MPA's searchability [20]; applying adaptive weights and OBL to enhance the performance of MPA [30]; combining chaotic sequence parameter and adaptive mechanism for velocity update to better MPA's exploitation and exploration search [39]; the use of CL approach to improve the search and transitioning within exploration and exploitation in MPA [34]; using linearly increased worst solutions (LIS) improvement strategy to address computational cost and accuracy issues associated with existing segmentation techniques, MPALS (MPA + LIS) and RUS are combined into a version called HMPA to serve as a solution to ISP [33]; incorporating logistic opposition-based learning (LOBL) into MPA to enhance the generation of various precise solutions with multiple population [32].; integrating MPA with CL approach and memory aspect of fractional calculus [42]; LA is used to enhance the artificial Jellyfish search algorithm (JS) and MPA, reducing

computational complexity while preserving their strengths [41]; MPA is integrated with spiral complex path search strategy based on Archimedes' spiral curve for perturbation, expanding the global exploration range and strengthening the algorithm's overall search capabilities [43]; it was also incorporated with pulse width modulation control boost converter to accurately track the MPP of a solar PV panel [65]; RL was integrated in MPA to improve its global searchability [44]; MPA was enhanced by incorporating a linear flight strategy in the middle stage to enhance predator interaction [56]; ranking-based mutation operator was used to identify the best search agent, to accelerate exploitation capabilities and preventing premature convergence in MPA [51]; DFS was used to tackle sudden transition between the Levy Flight and Brownian Motion [46]; teaching mechanism was also incorporated into MPA's first phase to promote its global search ability [47]; group-ranking of the predator populations, thorough learning approach implemented at stage 2 of MPA, and variable step-sizes control approach was applied [48]; etc.

TABLE IV. PROPOSED MPA MODIFICATIONS

Ref.	J. Quart.	Scopus CiteScore (2022)	Scopus IF	WOS IF	Algorithm	Article Type	Strategies for optimal search	Major Contributions	Research Domain
[64]	Q1	19.1	11.057	10.4	EMPA	Experiment/Result	Differential Evolution (DE) operator	DE operator is integrated into the exploration face of the standard MPA to escape local solution	PV Modelling
					MMPA			MMPA adopts classical MPA's search for MOPs. multi-objective modified MPA (M-MMPA) is a modification of the classical MPA to use DSEE strategy search phase for exploration and exploitation,	
					M-MMPA		multi-objective, dominance strategy based on exploration-exploitation (DSEE), Gaussian-based approach, and Nelder-Mead simplex	Gaussian-based mutation (GM) was integrated into M-MMPA to have a new model M-MMPA-GM, and Nelder-Mead simple method (NMM) was integrated to M-MMPA-GM to create a front for it to get to a better solution while maintaining the minimum possible time (NMM-M-MMPA-GM).	Engineering design
[2]	Q1	9.0	4.342	3.367	M-MMPA-GM	Experiment/Result		LEO is incorporated into classical MPA to tackle poor exploitation and exploration	
					M-MMPA-GM-NMM			optimization capabilities of the MPA were enhanced with adaptive weights and OBL, resulting in a Pareto front.	Image segmentation
[20]	Q1	12.3	8.664	8.038	LEO-MPA	Experiment/Result	Local Escaping Operator (LEO)	AMPA combines chaotic sequence parameters and adaptive mechanisms for velocity update to better MPA's exploitation and exploration search	Engineering design
[30]	Q1	14.3	9.028	8.7	BMPA	Experiment/Result	adaptive weights and OBL	CL approach that improves search and transitioning within exploration and exploitation is used on MPA	Antenna Signals
[39]	Q2	6.8	4.352	3.9	AMPA	Experiment/Result	chaotic sequence parameter and adaptive mechanism for velocity update	LIS is used to address computational cost and accuracy issues associated with existing segmentation techniques. MPALS and RUS are integrated into a version called HMPA to serve as a solution to ISP.	
[34]	Q1	11.9	7.811	5.431	MMPA	Experiment/Result	Comprehensive Learning (CL) approach	LOBL technique is incorporated to enhance the generation of various precise solutions with multiple populations.	Economic emission dispatch.
					MPALS		MPALS = MPA + LIS (linearly increased worst solutions improvement strategy.)		
[33]	Q1	23.0	11.674	8.139	HMPA	Experiment/Result	HMPA = MPALS + ranking-based updating strategy (RUS)		Image Segmentation
[32]	Q1	13.4	8.364	8.7	MMPA	Experiment/Result	Incorporates logistic opposition-based learning (LOBL)		Engineering design

Ref.	J. Quart.	Scopus CiteScore (2022)	Scopus IF	WOS IF	Algorithm	Article Type	Strategies for optimal search	Major Contributions	Research Domain
[42]	Q1	12.3	8.664	8.8	FOCLMPA	Experiment/Result	comprehensive learning (CL) approach	integrates MPA with the CL approach and memory aspect of fractional calculus.	Knowledge-based systems
[41]	Q1	12.3	8.664	8.8	LA-JS-MPA	Experiment/Result	Learning-Automata (LA)	LA is used to enhance the artificial Jellyfish search algorithm (JS) and MPA, reducing computational complexity while preserving their strengths. MPA is incorporated with a spiral complex path search strategy based on Archimedes' spiral curve for perturbation, expanding the global exploration range and strengthening the algorithm's overall search capabilities	Data clustering
[403]	Q2	4.5	3.143	2.7	FMMPA	Experiment/Result	Fusion multi-strategy	MPA is incorporated with a pulse width modulation control boost converter to accurately track the MPP of a solar PV panel.	Robot path planning
[65]	Q3	5.5	3.542	3.2	MPA	Experiment/Result	Pulse width modulation control boost converter	MPA is incorporated with a pulse width modulation control boost converter to accurately track the MPP of a solar PV panel.	Solar PV systems
[44]	Q1	12.3	8.635	8.0	Deep-MPA	Experiment/Result	reinforcement learning (RL)	RL is integrated with MPA to improve its global searchability.	Renewable energy system design
[56]	Q3	3.5	2.071	2.4	MSMPA	Experiment/Result	linear flight strategy	MPA is enhanced by incorporating a linear flight strategy in the middle stage to enhance predator interaction	Engineering design
[51]	Q2	4.7	3.308	3.5	EMPA	Experiment/Result	ranking-based mutation operator	the ranking-based mutation operator is used to identify the best search agent, enhancing exploitation capabilities and preventing premature convergence.	ANN classification
[46]	Q1	12.3	8.635	8.0	DFSMPA	Experiment/Result	Dynamic Foraging Strategy (DFS)	DFS is used to tackle sudden transitions between the Levy Flight and Brownian Motion	Real-world engineering
[47]	Q3	3.9	2.393	2.6	MTLMPA	Experiment/Result	Mechanism for Teaching & Learning	teaching mechanism is incorporated into MPA's first phase to promote its global search ability	Engineering Design
[48]	Q2	4.7	2.941	2.524	DAMPA	Experiment/Result	group-ranking of the predator populations	group-ranking of the predator populations, thorough learning approach implemented at stage 2 of MPA, and variable step-sizes control approach	Task scheduling in Cloud Computing

E. Recent Proposed Improvements in MPA

Recently, more articles have been published with many improvements still underway. Some of these proposals include a hybrid MPA and Particle Swarm Optimization (MPA-PSO), combining the global and local search abilities of PSO with the MPA [66], Multi-Population-based MPA (MultiPopMPA) which uses global, balanced, and local search strategies simultaneously throughout the search process [67], and a multi-

strategy MPA, Regularized ELM, and CFA, integrating multiple algorithms [68]. Furthermore, an improved MPA (IMPA) with Deep Gated Recurrent Unit (DGRU), a hybrid model combining IMPA and DGRU for better accuracy and generalization in profit prediction [69], and an improved MPA (IMPA), using adaptive weight adjustment and dynamic social learning mechanisms [70] have been proposed. A summary of the recent literature is presented in Table V.

TABLE V. RECENT PROPOSED MPA IMPROVEMENTS

Ref.	J. Quart.	Scopus CiteScore (2023)	Scopus IF	Proposed Algorithm	Article Type	Main Goal	Strategies of Optimal Search	Major Contribution	Research Domain
[66]	Q3	4.1	1.3	Hybrid MPA and Particle Swarm Optimization (MPA-PSO)	Experiment	To develop an optimal resource allocation strategy for vehicular edge computing networks	Combining the global and local search abilities of PSO with the MPA	Improved performance in resource allocation by leveraging the strengths of both MPA and PSO	Vehicular Edge Computing (VEC)
[67]	Q2	8.1	3.1	Multi-Population-based MPA (MultiPopMPA)	Experiment	To improve the search capabilities of the MPA by using a multi-population and multi-search strategy.	The algorithm uses global, balanced, and local search strategies simultaneously throughout the search process.	The proposed MultiPopMPA outperforms other metaheuristic algorithms in terms of precision, sensitivity, and F1-score metrics	AI, specifically in training ANN for classification tasks.
[68]	Q3	2.3	0.278	Multi-Strategy MPA, Regularized ELM, and CFA	Experiment	Accurate prediction of passenger flow to	Combining multiple algorithms to handle complexity and	High prediction accuracy and strong convergence performance with only 30	Passenger Flow Prediction.

Ref.	J. Quart.	Scopus CiteScore (2023)	Scopus IF	Proposed Algorithm	Article Type	Main Goal	Strategies of Optimal Search	Major Contribution	Research Domain
[71]	Q2	6.2	3.5	Quantum Theory-based MPA (QTbMPA)	Original Research	help local authorities with resource regulation. To develop an automated deep learning model for classifying brain tumors from MRI images Improve the search performance of the MPA for feature selection in schizophrenia classification using EEG signals.	uncertainty in passenger flow prediction. Bayesian optimization for hyperparameters and QTbMPA for feature selection.	iterations needed to reach the optimal solution Improved accuracy and sensitivity in brain tumor classification using a hybrid deep learning framework.	Medical image analysis
[72]	Q1	9.7	3.6	Chaotic-based MPA (CMPA)	Experiment	To enhance the performance of the Gorilla Troops Optimizer (GTO) by integrating high and low-velocity ratios inspired by the MPA.	Combining MPA with chaotic maps (logistic, tent, henon, sine, and tinkerbella maps).	The proposed SCMPA significantly outperforms other MPA variants in feature selection and classification accuracy.	Schizophrenia classification using EEG signals and metaheuristic algorithms.
[73]	Q1	7.5	3.8	Enhanced Gorilla Troops Optimizer (EGTO), with MPA	Experiment	To develop an optimal structured DCNN for automatic COVID-19 diagnosis using chest CT scans.	Balancing exploration and exploitation phases using high and low-velocity ratios	EGTO achieves superior performance in global optimization and engineering design problems compared to other algorithms.	Optimization and Engineering Design.
[74]	Q1	9.6	3.662	Modified MPA - convolutional neural networks (DCNNs).	Original Research	Improve profit prediction in financial accounting information systems	Utilizes a novel encoding scheme based on IP addresses, an Enfeebled layer for variable-length DCNN, and divides large datasets into smaller chunks for random evaluation.	The proposed DCNN-IPMPA model outperforms other benchmarks with high accuracy and competitive processing time.	Deep Learning and Medical Imaging.
[69]	Q1	9.6	5.0	Improved MPA (IMPA) with Deep Gated Recurrent Unit (DGRU).	Experiment	To improve gene selection methods for cancer classification using microarray data.	Dynamic flight behavior between Levy and Gaussian to enhance MPA's performance	Hybrid model combining IMPA and DGRU for better accuracy and generalization in profit prediction.	Financial accounting information systems and profit prediction.
[75]	Q2	11.4	4.5	Recursive Spider Wasp Optimizer MPA (RSWO-MPA)	Experiment	To enhance the MPA by addressing its limitations such as local optima traps, insufficient diversity, and premature convergence.	Combines ReliefF filter method with RSWO-MPA for efficient gene selection.	Achieves higher accuracy, selects fewer features, and exhibits more stability compared to other algorithms.	AI and Bioinformatics
[70]	Q1	7.5	3.8	Improved MPA (IMPA).	Experiment	Optimize the steam gasification process for converting palm oil waste into environmentally friendly energy	Adaptive weight adjustment and dynamic social learning mechanisms.	IMPA significantly improves optimization performance in engineering design problems by balancing exploration and exploitation.	Optimization algorithms in engineering design
[76]	Q1	9.6	5.0	Adaptive MPA (AMPA).	Experiment	Inversion of the permeability coefficient of a high core wall dam	Incorporation of AMPA into the SVM framework to enhance prediction precision and efficiency	Development of an intelligent optimization framework surpassing conventional machine learning techniques.	Renewable energy and intelligent systems.
[77]	Q2	5.3	2.5	MPA combined with a BP Neural Network.	Experiment	Enhance the low-voltage ride-through (LVRT) capability of grid-connected photovoltaic (PV) systems	Lévy and Brownian movements.	Comparison of three methods for seepage parameters inversion and demonstrating the advantage of the MPA.	Hydrology and Hydraulic Engineering.
[78]	Q2	3.5	3.4	MPA for optimized tuning of PI controllers	Experiment	To optimize process parameters in multi-process manufacturing to	MPA, Grey Wolf Optimization (GWO), and Particle Swarm Optimization (PSO).	MPA provides better results with higher convergence rates and improved system performance.	Electrical Engineering and Renewable Energy.
[79]	Q3	2.2	1.5	Improved MPA	Experiment		Utilizes reverse learning strategies and mixed control parameters to	Proposes a multi-process parameter optimization method using an improved MPA,	Mechanical Engineering and Manufacturing.

Ref.	J. Quart.	Scopus CiteScore (2023)	Scopus IF	Proposed Algorithm	Article Type	Main Goal	Strategies of Optimal Search	Major Contribution	Research Domain
[80]	Q3	4.1		MPA Aquila Optimizer (MAO)	Experiment	improve product quality To present a hybrid method combining MPA and AO for droop control in DC microgrids	enhance optimization capability Combining the strengths of MPA and AO to enhance exploration and exploitation.	addressing the severe coupling of multiple processes. Superior convergence ability and promising performance in droop control.	Electrical Engineering
[81]	Q3	2.4	1.2	MPA for robot path planning	Experiment	To design an optimal path for a robot to navigate from its starting point to its goal while avoiding obstacles	Heuristic search-based methods, potential field-based methods, sampling-based methods, hybrid methods, and evolutionary methods	The proposed method uses the Marine Predator Algorithm, which shows good performance in different situations.	Robot path planning and autonomous driving.
[82]	Q2	3.6	1.6	MPA- P-P-FOPID controller.	Experiment	To design a cascade P-P-FOPID controller optimized by the MPA for improving load frequency control in electric power systems.	The MPA is employed for its parameter-less, derivative-free, user-friendly, flexible, and simple nature.	The proposed controller demonstrated superior performance in reducing integral time absolute error (ITAE), settling time, and frequency and tie-line power deviations compared to other recent approaches. It also showed robustness against parametric uncertainties.	Electric power systems
[83]	Q1	19.9	6.2	improved binary MPA	Experiment	To develop an efficient offloading method that reduces energy consumption and meets time constraints in edge computing environments	The binary MPA is used for its effectiveness in solving optimization problems under constraints	The proposed method effectively meets deadlines while reducing energy consumption, even with an increasing number of users.	Edge computing.
[84]	Q2	4.3	2.7	Enhanced MPA (EMPA) - SVM	Experiment	To improve the accuracy and efficiency of IGBT switching power loss estimation using an optimized SVM model.	The EMPA is employed for its effectiveness in parameter optimization, leveraging its ability to handle complex, multi-dimensional search spaces.	The integration of EMPA with SVM results in a model that significantly enhances the accuracy and efficiency of power loss estimation in IGBT, outperforming traditional methods.	power electronics
[85]	Q1	11.2	6.2	MPA	Analytical	To identify and analyze the structural biases in the MPA using the BIAS Toolbox and Generalized Signature Test (GST).	The study employs the BIAS Toolbox and GST to detect and evaluate the structural biases within the MPA, revealing how these biases affect the algorithm's performance.	The article highlights significant structural biases in the MPA, which cause the population to revisit specific regions of the search space, leading to increased computational costs and slower convergence.	optimization algorithms
[86]	Q1	11.5	7.5	Improved Weighted MPA (WMPA)	Experiment	To enhance the accuracy and efficiency of SOM estimation by selecting the most relevant hyperspectral features using the improved WMPA.	The WMPA is optimized to improve feature selection by leveraging its ability to handle complex, multi-dimensional search spaces effectively.	The improved WMPA demonstrates higher accuracy and stability in predicting SOM content compared to traditional methods, providing a robust and efficient approach for SOM estimation.	agricultural and environmental monitoring
[87]	Q1	7.7	4.8	Enhanced Hybrid Aquila Optimizer with MPA (EHAOMPA)	Experiment	To enhance the performance of the Aquila Optimizer in solving combinatorial optimization problems by integrating it with the MPA.	The hybrid algorithm leverages the exploration capabilities of MPA and the exploitation strengths of AO to effectively navigate the search space and find optimal solutions.	The EHAOMPA demonstrates superior performance in various benchmark problems compared to traditional AO and other optimization algorithms, showing promise in solving industrial-constrained design problems and optimizing hyperparameters for	combinatorial optimization.

Ref.	J. Quart.	Scopus CiteScore (2023)	Scopus IF	Proposed Algorithm	Article Type	Main Goal	Strategies of Optimal Search	Major Contribution	Research Domain
[88]	Q1	14.8	7.2	improved MPA combined with Extreme Gradient Boosting (XGBoost)	Experiment	To enhance the accuracy and efficiency of shipment status time predictions using a hybrid approach that combines MPA and XGBoost.	The improved MPA incorporates opposition-based learning, chaos maps, and self-adaptive population strategies to optimize the parameters of the XGBoost model.	COVID-19 CT-image detection. The hybrid model demonstrates superior performance in predicting shipment status times compared to traditional methods, providing a robust and efficient solution for logistics and supply chain management. The algorithm successfully forms core backbone grids for the IEEE 39-node and IEEE 300-node systems, ensuring economic feasibility and optimal network connectivity while balancing active and reactive power demands.	logistics and supply chain management
[89]	Q2	4.3	2.4	improved MPA (multi-objective optimization).	Experiment	To enhance the resilience of power grid infrastructure by optimizing core backbone grid planning using a multi-objective 0–1 planning problem.	The improved MPA incorporates file management and an enhanced top predator selection mechanism to effectively explore the Pareto frontier for optimal solutions.	The research demonstrates that the MSC-KPCA-MPA-RF model achieves the best results, with a fitting coefficient of 0.9963 and a mean square error of 0.0047	power systems
[90]	Q2	4.3	2.7	MPA optimized random forest (RF) algorithm with laser-induced fluorescence (LIF) technology	Experiment	To develop a more efficient and accurate method for diagnosing transformer faults, overcoming the limitations of traditional methods.	The study employs principal component analysis (PCA) and kernel principal component analysis (KPCA) for dimensionality reduction, followed by the MPA-RF model for optimal fault diagnosis.		power systems and electrical engineering
[91]	Q1	9.5	2.6	MPA optimized pavement maintenance and rehabilitation (M&R) scheduling	Experiment	To develop a sustainable M&R scheduling optimization model that considers highway agency costs, environmental impacts, and social effects.	The MPA is used to handle the computational complexities of optimizing M&R scheduling for large-scale networks	The sustainable model reduces CO2 emissions by 6.5% and improves equity and safety indices by 40.7% and 2.5%, respectively, compared to conventional methods	pavement management systems and sustainable infrastructure engineering.
[92]	Q1	12.6	7.2	Clustering Wavelet Opposition-based MPA (CWOMPA) enhanced-MPA	Experiment	To improve optimization performance and feature selection in high-dimensional datasets, particularly in medical diagnosis.	CWOMPA incorporates fuzzy clustering, wavelet basis function, and adaptive opposition-based learning to enhance population diversity and prevent premature convergence	Demonstrates CWOMPA's superior performance in optimization and feature selection across various benchmark functions and medical datasets	Meta-heuristic optimization algorithms and feature selection in medical datasets.
[93]	Q1	5.7	2.6	Hybrid MPA-PSO to tackle the Energy Scheduling Problem (ESP).	Experiment	To optimize electricity bills, energy consumption, and user comfort by finding the best schedule for smart appliances	The proposed method enhances the searching capabilities of MPA using PSO components to improve schedules with poor fitness values	The research demonstrates the efficiency and high performance of the hybrid method in optimizing ESP objectives compared to other methods	Internet of Things (IoT) and smart grid technology
[94]	Q4	1.3	1.74	Modified MPA (MMPA) for automated atrial fibrillation detection using ECG signals.	Experiment	To develop a method for automatically detecting atrial fibrillation using transient single lead ECG readings	The algorithm utilizes Heart Rate Variability (HRV) and frequency analysis for feature extraction, followed by classification using SVM	The study's innovative contribution is the application of the MMPA for identifying atrial fibrillation in brief ECG data, achieving a maximum accuracy of 99.8%.	Biomedical Engineering and AI.
[95]	Q1	6.5	4.1	bidirectional gated recurrent unit (BiGRU) optimized MPA	Experiment	To analyze the influence of scraper geometry and roughness on the coating process using advanced predictive and simulation models.	The MPA-BiGRU pseudo-lattice Boltzmann (pseudo-LB) method is employed to simulate the coating flow without specific rheological equations.	The study finds that rectangle geometry is suitable for high coating speeds, while trapezium geometry is better for low speeds. Scraper roughness significantly affects the process with rectangle geometry.	materials science and engineering

Ref.	J. Quart.	Scopus CiteScore (2023)	Scopus IF	Proposed Algorithm	Article Type	Main Goal	Strategies of Optimal Search	Major Contribution	Research Domain
[96]	Q2	4.6	2.7	Improved MPA (IMPA)	Experiment	To optimize water resource allocation in Huaying City by balancing social, economic, and ecological benefits	The IMPA employs chaotic initialization for population diversity, golden sine algorithm for balanced exploration and exploitation, and quadratic interpolation for enhanced search accuracy.	The study demonstrates that IMPA outperforms other algorithms in terms of stability and accuracy for water resource optimization, providing a new approach for sustainable water management.	water resource management and optimization algorithms.
[97]	Q2	10.2	4.7	MPA	Comparative	To minimize the deficit of agricultural water supply by optimizing reservoir operations under baseline and climate change conditions.	The MPA uses random walk strategies (Brownian and Levy motions) and elite matrices to enhance exploration and exploitation phases.	Demonstrates that MPA outperforms GA in terms of reliability, resiliency, and vulnerability in reservoir operations.	Water Resource Management and Optimization Algorithms.
[98]	Q2	3.4	1.7	PRMPA-Spectral-SMOTE with improved MPA (IMPA).	Experiment	To enhance the classification performance of biomedical data, which is often high-dimensional and imbalanced	The algorithm uses minimal-redundancy maximal-relevance (mRMR) for feature selection, Spectral-SMOTE for data resampling, and an improved MPA for optimizing key parameters.	The method significantly improves the classification accuracy of biomedical data, outperforming other data resampling methods	biomedical data
[99]	Q1	9.8	3.4	Uniform MPA (UMPA), combines uniform design with the MPA	Experiment	To accurately and efficiently detect neural unit modules in brain networks, which can aid in disease detection and targeted therapy	UMPA leverages uniform design to ensure evenly distributed solutions and MPA for optimization, incorporating Lévy flight and Brownian movement strategies.	Integration of uniform design with MPA, resulting in improved performance in identifying neural unit modules compared to other methods.	brain network analysis
[100]	Q1	9.7	3.6	Reinforcement Learning MPA (RLMPA) to enhance global optimization.	Experiment	Improve Optimization-Address weak convergence, limited balance capacity, and optimization limitations in MPA by introducing RLMPA.	Three Location Update Strategies: Ranking paired mutually beneficial learning; Gaussian random walk learning; and Modified somersault foraging.	Enhanced Performance: RLMPA shows superior performance in global optimization, search efficiency, and convergence speed compared to 10 competitive algorithms.	engineering design

V. APPLICATIONS OF MPA IN VARIOUS DOMAINS

The MPA has found wide acceptance across many research domains. Focusing on areas with the widest coverage and most recent development, Engineering (28.2%), Computer Science (26.0%), Mathematics (10.9%), and Energy (7.1%) of the applications, the following summaries are presented.

A. Engineering

The highest of MPA's applications based on Table V are in real-world engineering designs [1]. In this domain, MPA has been used to solve real-world problems such as pressure vessel design, tension/compression spring design, and welded beam design [1], estimating the parameter of frequency-modulated sound wave (FM), speed spectrum radar Polly phase code design (SSRPP), and Lennard-Jones (LJ) potential problem [45]. These problems are constrained engineering benchmarks and were made with associated practical engineering examples. With the help of the death penalty approach, the constrained problems were converted to unconstrained ones. Another real-world

problem solved includes demand-controlled ventilation of the operating fan schedule, where a 2-zone (entry and exit) retail store stocked with a supply and exhaust fan for ventilation was examined. The main objective was to "reduce the fan's energy consumption using demand-controlled ventilation subject to airflow and the amount of carbon dioxide (CO₂)".

B. Computer Science

In this area, MPA has been used to develop an efficient offloading method that reduces energy consumption and meets time constraints in edge computing environments [83] and to develop an optimal resource allocation strategy for vehicular edge computing networks [66]. It has also been used to design an optimal structured deep convolutional neural network (DCNN) [74], in training ANN for classification tasks [67], incorporation into the SVM framework to enhance prediction precision and efficiency [76] and was combined with a BP Neural Network [77] for improved performance. Furthermore, on the Internet of Things (IoT) and smart grid technology, MPA was used to optimize electricity bills, energy consumption, and

user comfort by finding the best schedule for smart appliances [93].

C. Mathematics

The proposed variant of MPA such as DFSMPA was also applied to three sets of standard mathematical test functions and one set of real-world engineering optimization problems including (i) Classical functions such as unimodal, multimodal, and fixed multimodal functions. (ii) Contemporary numerical optimizations CEC-BC-2017 comprises 30 composition and hybrid functions. (iii) CEC06-2019 (100-Digits challenge). and (iv) Ten CEC-2020 problems applicable to real engineering optimization [46].

D. Energy

MPA has been applied in power systems and electrical engineering to develop a more efficient and accurate method for diagnosing transformer faults, overcoming the limitations of traditional methods [90], and enhancing the resilience of power grid infrastructure by optimizing core backbone grid planning using a multi-objective 0–1 planning problem [89]. In addition, a cascade P-P-FOPID controller optimized by MPA for improving load frequency control in electric power systems was also designed [82]. Furthermore, MPA and AO have been combined for drop control in DC microgrids [80]. In electrical engineering and renewable energy, MPA has been applied to enhance the low-voltage ride-through (LVRT) capability of grid-connected photovoltaic (PV) systems [78] and to optimize the steam gasification process for converting palm oil waste into environmentally friendly energy [76]. MPA has been used in power electronics to improve the accuracy and efficiency of IGBT switching power loss estimation [84].

Comprehensively, the research application domain and tasks include real-world and engineering design [2, 20, 32, 45–47, 55, 58, 70, 73, 77, 79, 85, 91, 95, 100–108], microgrid feature selection [7], antenna signals [39], selective harmonic elimination [62], power modulation [38], ligament deficiency detection [35], transistor's design [36], breast cancer diagnosis [37], task scheduling in cloud computing [48], time series prediction [49], medical image fusion and analysis [26, 71, 72, 74, 92], economic emission dispatch [34], wind renewable energy [60], mathematical computation [61, 87], image segmentation [30, 33, 109, 110], PV System and modelling [27, 64, 65, 111], Optimal power flow [28], Biological Processes [50], Microgrid system [29], optimal reactive power dispatch [24], ANN training and classification [51, 67], pipeline leakage detection [52], arrhythmia classification [31], load frequency controller design [53], knowledge-based systems [42], data clustering [41], robot path planning [43, 81], battery aging mechanism [54], li-ion batteries [63], transformer oil breakdown [59], renewable energy system design [44], dynamic clustering simulation [112], marine stabilized platforms [113], joint regularization semi-supervised ELM [114], oil layer prediction [115], EEG/ERP signal [116], urban green space type [117], SVM optimization [118], solar-powered BLDC motor design [119], network reconfiguration and distributed generator allocation [121], wind and solar energy [122], DNA storage [123], white blood cell classification [124], wireless sensor network coverage [125], hybrid heartbeats [126], distribution system [127], task scheduling in cloud computing [128], shrimp

freshness detection and classification [129], evolutionary computations [130], energy management system [131], hybrid active power filter [132], gene selection in cancer microarray classification [133], supercapacitor modelling [134], COVID-19 detection modelling [135], wind power forecasting [136], thermal error modelling of electrical spindle [137], electrical power system & renewable energy [76, 78, 80, 82, 84, 89, 90, 138], DC motors [139], feature selection in metabolomics [140], optimal power flow [141], fuel cell steady-state modelling [142], structural damage detection [143], production planning [144], wind energy systems [145], AI and Bioinformatics [75, 94, 98, 99], Vehicular Edge Computing (VEC)[66, 83], passenger flow prediction [68], Internet of Things (IoT) and smart grid technology [93, 120], financial accounting information systems [69], agricultural and environmental monitoring [86], logistics and supply chain management [88], and water resource management and optimization algorithms [96, 97].

VI. DISCUSSION

Although the classic MPA proposed by study [1] was for a single objective and possessed some shortcomings, multi-objective variants were later proposed and other subsequent improvements were made [133], [146–148]. It is worth noting that most of the improvements in the literature were based on enhancing MPA's initial population, exploitation, exploration, and convergence.

Firstly, opposition-based learning (OBL), a novel technique introduced by Tizhoosh, has been widely adopted by numerous researchers to improve the initial population quality of metaheuristic algorithms. The OBL has been used to produce a more widely distributed initial population for MPA.

Furthermore, many scholars applied varying OBL approaches to tackle MPA's limitations. These include integrating the OBL strategy with GWO into MPA [27], OBL and TL were also used to improve MPA and IMPA-ResNet50 using a modified MPA algorithm for CNN hyperparameter selection to improve output performance for classification [31], enhancing the optimization capabilities of MPA with adaptive weights and OBL [30], and logistic OBL (LOBL) technique was incorporated in study [32] to enhance the generation of various precise solutions with multiple populations. In study [106], quasi-learning (Q-learning) was introduced to help MPA fully utilize the information generated by previous iterations and subsequent ones, QOBL was introduced to support an increase in population diversity, reducing the risk of convergence to inferior local optima. In addition, the quasi-opposition learning and spiral search strategies were incorporated into QRSS-MPA to improve it [123].

Additional strategies adopted in the literature to control MPA's exploration-exploitation search include chaotic maps' exploitation capabilities alone. Several chaotic maps can be implemented to improve the exploration-exploitation process [108]. The chaotic map can be applied to balance the trade-off between the exploration and exploitation phases. The self-adaptive population method automatically adjusts the population size for each iteration. It helps to increase the convergence speed [101, 108]. In study [107], chaotic maps, opposition-based learning strategy (OBLs), and teaching-

learning-based optimization (TLBO) with strong exploitation operators were combined. MPA was first modified to have MMPA that leverages chaotic maps and OBLs in the initialization stage to generate high-quality individuals. Parameter-free teaching-learning-based optimization method with a strong exploitation operator was incorporated into MPA (MMPA-TLBO), which effectively trades off between the exploitation and exploration process. Furthermore, [114] applied a multi-strategy approach involving three strategies to improve the performance of MPA. It included a chaotic opposition learning strategy to generate a high-quality initial population, adaptive inertia weights, adaptive step control factors to improve exploration, utilization, and convergence speed, and a neighborhood-dimensional learning strategy to maintain population diversity.

The literature also used comprehensive learning (CL) to improve the performance of MPA. For instance, in study [34], the CL approach was used to improve search and transition within the exploration and exploitation of MPA. In study [42], MPA was integrated with the CL approach and the memory aspect of fractional calculus.

Other approaches that were implemented to improve the performance of MPA include the use of Dynamic Foraging Strategy (DFS) to tackle sudden transition between the Levy Flight and Brownian Motion by study [46], Differential Evolution (DE) operator was integrated into the exploration phase of the standard MPA by study [64] to escape local solution, and the use of teaching and learning mechanism was incorporated into MPA in study [47] where the teaching mechanism was integrated into the first phase of the MPA to promote its global search ability. In study [39], chaotic sequence parameters and an adaptive mechanism for velocity update were implemented to improve MPA's search exploitation and exploration. Also, group ranking of predator populations was proposed by study [48], incorporating a thorough learning approach implemented at stage 2 of MPA, and a variable step-size control approach.

Furthermore, time-delay polynomials were applied to improve the model prediction performance [49]. A local escaping operator (LEO) was incorporated into classical MPA [20] to tackle poor exploitation and exploration. In study [24], exploration was improved by replacing the Brownian motion's random walk of search agents with a high-tailed Weibull distribution. The exploitation stage in phase 3 was also modified to use PSO or MPA based on probability. The study in [51] used a ranking-based mutation operator to identify the most performing search agent, enhancing exploitation capabilities and preventing premature convergence.

In addition, [52] applied a strategy known as a "good point set" in the initial population stage for improvement by initializing a good point set and enhancing the convergence factor (CF) and Cauchy distribution. Learning automata (LA) was used in [41] to improve the artificial jellyfish search algorithm (JS) and MPA, reducing computational complexity while preserving their strengths. MPA is incorporated with a spiral complex path search strategy based on Archimedes' spiral curve for perturbations in study [43], expanding the global exploration range and strengthening the algorithm's overall search capabilities. In study [55], HOGO was implemented

which modifies the search transition in MPA to gradually shift from exploration to exploitation as iterations progress, utilizing the global best solution at each iteration. In study [56], MPA was enhanced by incorporating a linear flight strategy in the middle stage to enhance predator interaction. Reinforcement learning (RL) was also integrated into MPA [44] to improve global searchability.

VII. FUTURE RESEARCH DIRECTION

The application of the MPA is predominant in engineering and real-world design. However, researchers need to extend it to other disciplines and optimization problems. Additional variants of MPA, such as Constrained MPA (CMPA), Mixed-Integer MPA (MIMPA), and Parameter Less MPA (PMPA), warrant exploration, alongside dynamic applications like Mobile or dynamic MPA in robotics. Moreover, MPA shows potential in diverse areas like knowledge discovery, power systems, signal processing, DNA assembly, and medical diagnostics. However, variants like MpNMRA still face challenges such as potential entrapment in local optima and inefficient optimization of all test functions, necessitating improvements in control parameter selection and solution retention. In addition, the expansion of MPA into multi-objective problems, alongside enhanced stability and convergence analysis remains an area for future research. Other proposed methods, such as DAMPA and MTLMPA, also require comprehensive testing in various domains to assess their efficacy. Additionally, the scalability of dataset testing for algorithms like IMPA-ResNet50 and exploration of their performance in regression tasks, computational efficiency enhancements, and generalization to different CNN configurations are suggested. Further research efforts should aim to integrate MPA with deep learning and machine learning techniques, explore its potential in renewable energy systems – with emphasis on solar radiation forecasting, refine its application in real-world scenarios, and investigate its hybridization with other metaheuristic methods for improved optimization outcomes.

VIII. CONCLUSION

This systematic review of MPA presents a wide panorama concerning its theoretical formulation, practical implementations, and novel improvements. Current research synthesizes studies undertaken during the last five years that underline, among others, the flexibility and efficiency of the MPA approach as a metaheuristic optimization method but with peculiar efficacy in handling complex, high-dimensional, and multimodal optimization problems. From its principle of inspiration to its adaptive nature, MPA has always shown robust performance in diverse fields, ranging from real-world engineering design, image segmentation, and PV system modeling, indicating its wide acceptance and high applications. It also identifies the critical design parameters and their influence on the convergence and performance of the algorithm, thus contributing to the deeper theoretical understanding of the method.

This research significantly contributes to the optimization literature by systematically categorizing MPA variants based on their core improvements, including parameter tuning, hybridization, and other modification mechanisms. The review delineates the strengths and limitations of each approach by

comparing these variants across benchmark problems, thus providing a roadmap for future research. We also find gaps in the current literature, such as a need for more rigorous theoretical analysis regarding convergence properties and scalability in dynamic environments. These insights pave the way for developing more efficient, adaptive, and robust MPA variants that can address emerging challenges in optimization.

MPA provides several practical benefits over other optimization algorithms, including ease of implementation and minimal parameter-tuning requirements to escape local optima using a form of collective intelligence. Computationally efficient with adaptability toward real-time application, such as resource allocation and feature selection control system optimization is another added area of MPA. Besides, the ease with which the algorithm can be combined with other optimization techniques has favored its use in hybrid systems and further extended the usefulness of the MPA in solving complex, real-world problems. This review, therefore, highlights that the simplicity and flexibility of MPA make it a useful tool for practitioners from all walks of life in addressing optimization problems, both at the academic and applied levels.

This systematic review, therefore, underlines the continuous relevance and transformational potential of MPA as an optimization technique. By connecting the dots between theoretical developments and practical applications, it provides a comprehensive overview of the algorithm's capabilities and limitations, thus laying the ground for future innovations in the field. We expect this work to be a reference point for researchers and practitioners alike, encouraging new contributions that tap into MPA's unique strengths to solve ever more complex optimization problems.

REFERENCES

- [1] A. Faramarzi, M. Heidarinejad, S. Mirjalili, and A. H. Gandomi, "Marine Predators Algorithm: A nature-inspired metaheuristic," *Expert Syst. Appl.*, vol. 152, 2020.
- [2] M. Abdel-Basset, R. Mohamed, S. Mirjalili, R. K. Chakraborty, and M. Ryan, "An Efficient Marine Predators Algorithm for Solving Multi-Objective Optimization Problems: Analysis and Validations," *IEEE Access*, vol. 9, pp. 42817–42844, 2021.
- [3] L. Stripinis and R. Paulavičius, "DIRECTGO: A new DIRECT-type MATLAB toolbox for derivative-free global optimization," *ACM Trans. Math. Softw.*, vol. 48, no. 4, pp. 1–46, 2022.
- [4] F. Boukouvala, R. Misener, and C. A. Floudas, "Global optimization advances in mixed-integer nonlinear programming, MINLP, and constrained derivative-free optimization, CDFO," *Eur. J. Oper. Res.*, vol. 252, no. 3, pp. 701–727, 2016.
- [5] D. Lera and Y. D. Sergeyev, "GOSH: derivative-free global optimization using multi-dimensional space-filling curves," *J. Glob. Optim.*, vol. 71, pp. 193–211, 2018.
- [6] Y. D. Sergeyev, M. S. Mukhametzhanov, D. E. Kvasov, and D. Lera, "Derivative-free local tuning and local improvement techniques embedded in the univariate global optimization," *J. Optim. Theory Appl.*, vol. 171, pp. 186–208, 2016.
- [7] R. Rai, K. G. Dhal, A. Das, and S. Ray, "An inclusive survey on marine predators algorithm: variants and applications," *Arch. Comput. Methods Eng.*, pp. 1–40, 2023.
- [8] S. Mirjalili, "Genetic Algorithm," in *Evolutionary Algorithms and Neural Networks: Theory and Applications*, Cham: Springer International Publishing, 2019, pp. 43–55.
- [9] H.-P. Schwefel, *Numerical optimization of computer models*. John Wiley & Sons, Inc., 1981.
- [10] J. Koza, "On the programming of computers by means of natural selection," *Genet. Program.*, 1992.
- [11] R. Storn and K. Price, "Differential evolution--a simple and efficient heuristic for global optimization over continuous spaces," *J. Glob. Optim.*, vol. 11, pp. 341–359, 1997.
- [12] S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi, "Optimization by simulated annealing," *Science (80-.)*, vol. 220, no. 4598, pp. 671–680, 1983.
- [13] E. Rashedi, H. Nezamabadi-pour, and S. Saryazdi, "GSA: A Gravitational Search Algorithm," *Inf. Sci. (Ny)*, vol. 179, no. 13, pp. 2232–2248, 2009.
- [14] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-International Conference on Neural Networks*, 1995, vol. 4, pp. 1942–1948.
- [15] M. Dorigo, M. Birattari, and T. Stützle, "Ant colony optimization-artificial ants as a computational intelligence technique.," *IEEE Comput. Intell. Mag.*, 2006.
- [16] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *2009 World congress on nature & biologically inspired computing (NaBIC)*, 2009, pp. 210–214.
- [17] S. Mirjalili, S. Saremi, S. M. Mirjalili, and L. dos S. Coelho, "Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization," *Expert Syst. Appl.*, vol. 47, pp. 106–119, 2016.
- [18] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, 2014.
- [19] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems," *Adv. Eng. Softw.*, vol. 114, pp. 163–191, 2017.
- [20] M. Oszust, "Enhanced Marine Predators Algorithm with Local Escaping Operator for Global Optimization," *Knowledge-Based Syst.*, vol. 232, 2021.
- [21] W. H. Bangyal, J. Ahmad, H. T. Rauf, and S. Pervaiz, "An overview of mutation strategies in bat algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 8, pp. 523 – 534, 2018.
- [22] N. Ul Hassan, W. H. Bangyal, M. S. Ali Khan, K. Nisar, A. A. Ag Ibrahim, and D. B. Rawat, "Improved opposition-based particle swarm optimization algorithm for global optimization," *Symmetry (Basel)*, vol. 13, no. 12, p. 2280, 2021.
- [23] W. H. Bangyal, J. Ahmad, and H. T. Rauf, "An overview of mutation strategies in particle swarm optimization," *Int. J. Appl. Metaheuristic Comput.*, vol. 11, no. 4, pp. 16 – 37, 2020.
- [24] M. A. M. Shaheen, D. Youstri, A. Fathy, H. M. Hasanien, A. Alkuhayli, and S. M. Mueen, "A Novel Application of Improved Marine Predators Algorithm and Particle Swarm Optimization for Solving the ORPD Problem," *Energies*, vol. 13, no. 21, 2020.
- [25] M. A. Elaziz et al., "Enhanced Marine Predators Algorithm for identifying static and dynamic Photovoltaic models parameters," *ENERGY Convers. Manag.*, vol. 236, May 2021.
- [26] P.-H. Dinh, "A novel approach based on Three-scale image decomposition and Marine predators algorithm for multi-modal medical image fusion," *Biomed. Signal Process. Control*, vol. 67, 2021.
- [27] E. H. Houssein, M. A. Mahdy, A. Fathy, and H. Rezk, "A modified Marine Predator Algorithm based on opposition based learning for tracking the global MPP of shaded PV system," *Expert Syst. Appl.*, vol. 183, 2021.
- [28] R. A. Swief, N. M. Hassan, H. M. Hasanien, A. Y. Abdelaziz, and M. Z. Kamh, "Multi-Regional Optimal Power Flow Using Marine Predators Algorithm Considering Load and Generation Variability," *IEEE Access*, vol. 9, pp. 74600–74613, 2021.
- [29] A. H. Yakout, H. Kotb, H. M. Hasanien, and K. M. Aboras, "Optimal Fuzzy PIDF Load Frequency Controller for Hybrid Microgrid System Using Marine Predator Algorithm," *IEEE Access*, vol. 9, pp. 54220–54232, 2021.
- [30] Z. Xing and Y. He, "Many-objective multilevel thresholding image segmentation for infrared images of power equipment with boost marine predators algorithm," *Appl. Soft Comput.*, vol. 113, 2021.
- [31] E. H. Houssein, M. Hassaballah, I. E. Ibrahim, D. S. AbdElminaam, and Y. M. Wazery, "An automatic arrhythmia classification model based on

- improved Marine Predators Algorithm and Convolutions Neural Networks,” *Expert Syst. Appl.*, vol. 187, 2022.
- [32] Q. Fan, H. Huang, Q. Chen, L. Yao, K. Yang, and D. Huang, “A modified self-adaptive marine predators algorithm: framework and engineering applications,” *Eng. Comput.*, vol. 38, no. 4, pp. 3269–3294, 2022.
- [33] M. Abdel-Basset, R. Mohamed, and M. Abouhawwash, “Hybrid marine predators algorithm for image segmentation: analysis and validations,” *Artif. Intell. Rev.*, vol. 55, no. 4, pp. 3315–3367, 2022.
- [34] M. H. Hassan, D. Yousri, S. Kamel, and C. Rahmann, “A modified Marine predators algorithm for solving single- and multi-objective combined economic emission dispatch problems,” *Comput. Ind. Eng.*, vol. 164, 2022.
- [35] G. Wang, X. Zeng, G. Lai, G. Zhong, K. Ma, and Y. Zhang, “Efficient Subject-Independent Detection of Anterior Cruciate Ligament Deficiency Based on Marine Predator Algorithm and Support Vector Machine,” *IEEE J. Biomed. Heal. Informatics*, vol. 26, no. 10, pp. 4936–4947, 2022.
- [36] N. Kaur, M. Rattan, S. Singh Gill, G. Kaur, G. Kaur Walia, and R. Kaur, “Marine predators algorithm for performance optimization of nanoscale FinFET,” in *Materials Today: Proceedings*, 2022, vol. 66, pp. 3529–3533.
- [37] E. H. Houssein, M. M. Emam, and A. A. Ali, “An optimized deep learning architecture for breast cancer diagnosis based on improved marine predators algorithm,” *Neural Comput. Appl.*, vol. 34, no. 20, pp. 18015–18033, 2022.
- [38] A. H. Yakout, W. Sabry, A. Y. Abdelaziz, H. M. Hasanien, K. M. AboRas, and H. Kotb, “Enhancement of frequency stability of power systems integrated with wind energy using marine predator algorithm based PIDA controlled STATCOM,” *Alexandria Eng. J.*, vol. 61, no. 8, pp. 5851–5867, 2022.
- [39] E. O. Owoola, K. Xia, S. Ogunjo, S. Mukase, and A. Mohamed, “Advanced Marine Predator Algorithm for Circular Antenna Array Pattern Synthesis,” *SENSORS*, vol. 22, no. 15, Aug. 2022.
- [40] B. S. Yıldız, “Marine predators algorithm and multi-verse optimisation algorithm for optimal battery case design of electric vehicles,” *Int. J. Veh. Des.*, vol. 88, no. 1, pp. 1–11, 2022.
- [41] S. Barshandeh, R. Dana, and P. Eskandarian, “A learning automata-based hybrid MPA and JS algorithm for numerical optimization problems and its application on data clustering,” *KNOWLEDGE-BASED Syst.*, vol. 236, Jan. 2022.
- [42] D. Yousri, M. Abd Elaziz, D. Oliva, A. Abraham, M. A. Alotaibi, and M. A. Hossain, “Fractional-order comprehensive learning marine predators algorithm for global optimization and feature selection,” *Knowledge-Based Syst.*, vol. 235, 2022.
- [43] L. Yang, Q. He, L. Yang, and S. Luo, “A Fusion Multi-Strategy Marine Predator Algorithm for Mobile Robot Path Planning,” *Appl. Sci.*, vol. 12, no. 18, 2022.
- [44] E. H. Houssein, I. E. Ibrahim, M. Kharrich, and S. Kamel, “An improved marine predators algorithm for the optimal design of hybrid renewable energy systems,” *Eng. Appl. Artif. Intell.*, vol. 110, 2022.
- [45] R. Salgotra, S. Singh, U. Singh, S. Mirjalili, and A. H. Gandomi, “Marine predator inspired naked mole-rat algorithm for global optimization,” *Expert Syst. Appl.*, vol. 212, 2023.
- [46] B. Shen, M. Khishe, and S. Mirjalili, “Evolving Marine Predators Algorithm by dynamic foraging strategy for real-world engineering optimization problems,” *Eng. Appl. Artif. Intell.*, vol. 123, 2023.
- [47] Y. Ma, C. Chang, Z. Lin, X. Zhang, J. Song, and L. Chen, “Modified Marine Predators Algorithm hybridized with teaching-learning mechanism for solving optimization problems,” *Math. Biosci. Eng.*, vol. 20, no. 1, pp. 93 – 127, 2023.
- [48] D. Chen and Y. Zhang, “Diversity-Aware Marine Predators Algorithm for Task Scheduling in Cloud Computing,” *Entropy*, vol. 25, no. 2, 2023.
- [49] H. Zhu, L. Chong, W. Wu, and W. Xie, “A novel conformable fractional nonlinear grey multivariable prediction model with marine predator algorithm for time series prediction,” *Comput. Ind. Eng.*, vol. 180, 2023.
- [50] E. Öge, B. N. Yaman, and Y. B. ÇeSahin, “Optimization of biodegradation yield of reactive blue 49: An integrated approach using response surface methodology based marine predators algorithm,” *J. Microbiol. Methods*, vol. 206, p. 106691, 2023.
- [51] J. Zhang and Y. Xu, “Training Feedforward Neural Networks Using an Enhanced Marine Predators Algorithm,” *Processes*, vol. 11, no. 3, 2023.
- [52] Y. Hou, Y. Zhang, J. Lu, N. Hou, and D. Yang, “Application of improved multi-strategy MPA-VMD in pipeline leakage detection,” *Syst. Sci. Control Eng.*, vol. 11, no. 1, 2023.
- [53] A. Halmous, Y. Oubbati, M. Lahdeb, and S. Arif, “Design a new cascade controller PD-PID optimized by marine predators algorithm for load frequency control,” *Soft Comput.*, vol. 27, no. 14, pp. 9551–9564, 2023.
- [54] R. Xu, Y. Wang, and Z. Chen, “Data-Driven Battery Aging Mechanism Analysis and Degradation Pathway Prediction,” *BATTERIES-BASEL*, vol. 9, no. 2, Feb. 2023.
- [55] P. D. Kusuma and D. Adiputra, “Hybrid Marine Predator Algorithm and Hide Object Game Optimization,” *Eng. Lett.*, vol. 31, no. 1, pp. 262–270, 2023.
- [56] C. Qin and B. Han, “Multi-Stage Improvement of Marine Predators Algorithm and Its Application,” *C. - Comput. Model. Eng. Sci.*, vol. 136, no. 3, pp. 3097–3119, 2023.
- [57] F. Qin, A. M. Zain, and K.-Q. Zhou, “Harmony search algorithm and related variants: A systematic review,” *Swarm Evol. Comput.*, p. 101126, 2022.
- [58] D. Moher, A. Liberati, J. Tetzlaff, D. G. Altman, and P. Group*, “Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement,” *Ann. Intern. Med.*, vol. 151, no. 4, pp. 264–269, 2009.
- [59] S. S. M. Ghoneim, M. M. Alharthi, R. A. El-Sehiemy, and A. M. Shaheen, “Prediction of Transformer Oil Breakdown Voltage with Barriers Using Optimization Techniques,” *Intell. Autom. SOFT Comput.*, vol. 31, no. 3, pp. 1593–1610, 2022.
- [60] H. Xia, S. Zhang, R. Jia, H. Qiu, and S. Xu, “Blade shape optimization of Savonius wind turbine using radial based function model and marine predator algorithm,” *Energy Reports*, vol. 8, pp. 12366–12378, 2022.
- [61] K. Mehmood et al., “Nonlinear Hammerstein System Identification: A Novel Application of Marine Predator Optimization Using the Key Term Separation Technique,” *Mathematics*, vol. 10, no. 22, 2022.
- [62] H. Yigit, S. Urgan, and S. Mirjalili, “Comparison of recent metaheuristic optimization algorithms to solve the SHE optimization problem in MLI,” *NEURAL Comput. & Appl.*, vol. 35, no. 10, SI, pp. 7369–7388, Apr. 2023.
- [63] S. M. Abdelhafiz, M. E. Fouda, and A. G. Radwan, “Parameter Identification of Li-ion Batteries: A Comparative Study,” *ELECTRONICS*, vol. 12, no. 6, Mar. 2023.
- [64] M. Abd Elaziz et al., “Enhanced Marine Predators Algorithm for identifying static and dynamic Photovoltaic models parameters,” *Energy Convers. Manag.*, vol. 236, 2021.
- [65] S. K. Vankadara, S. Chatterjee, P. K. Balachandran, and L. Mihet-Popa, “Marine Predator Algorithm (MPA)-Based MPPT Technique for Solar PV Systems under Partial Shading Conditions,” *ENERGIES*, vol. 15, no. 17, Sep. 2022.
- [66] S. S. Abuthahir and J. S. P. Peter, “A Combined Marine Predators and Particle Swarm Optimization for Task Offloading in Vehicular Edge Computing Network,” *Int. J. Networked Distrib. Comput.*, 2024.
- [67] A. Özkiş, “A multi-population-based marine predators algorithm to train artificial neural network,” *Soft Comput.*, 2024.
- [68] P. Guo, “Big Data Multi-Strategy Predator Algorithm for Passenger Flow Prediction,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 5, pp. 800 – 810, 2024.
- [69] X. Li, M. Khishe, and L. Qian, “Evolving deep gated recurrent unit using improved marine predator algorithm for profit prediction based on financial accounting information system,” *Complex Intell. Syst.*, vol. 10, no. 1, pp. 595 – 611, 2024.
- [70] Y. Chun, X. Hua, C. Qi, and Y. X. Yao, “Improved marine predators algorithm for engineering design optimization problems,” *Sci. Rep.*, vol. 14, no. 1, 2024.
- [71] M. S. Ullah, M. A. Khan, A. Masood, O. Mzoughi, O. Saidani, and N. Alturki, “Brain tumor classification from MRI scans: a framework of hybrid deep learning model with Bayesian optimization and quantum theory-based marine predator algorithm,” *Front. Oncol.*, vol. 14, 2024.

- [72] Z. Garip, E. Ekinçi, K. Serbest, and S. Eken, "Chaotic marine predator optimization algorithm for feature selection in schizophrenia classification using EEG signals," *Cluster Comput.*, 2024.
- [73] M. H. Hassan, S. Kamel, and A. W. Mohamed, "Enhanced gorilla troops optimizer powered by marine predator algorithm: global optimization and engineering design," *Sci. Rep.*, vol. 14, no. 1, 2024.
- [74] B. Liu, X. Nie, Z. Li, S. Yang, and Y. Tian, "Evolving deep convolutional neural networks by IP-based marine predator algorithm for COVID-19 diagnosis using chest CT scans," *J. Ambient Intell. Humaniz. Comput.*, vol. 15, no. 1, pp. 451 – 464, 2024.
- [75] S. Osama, A. A. Ali, and H. Shaban, "Gene selection based on recursive spider wasp optimizer guided by marine predators algorithm," *Neural Comput. Appl.*, 2024.
- [76] X. Guo, Y. Bouteraa, M. Khishe, C. Li, and D. Martín, "Intelligent optimization of steam gasification catalysts for palm oil waste using support vector machine and adaptive transition marine predator algorithm," *Complex Intell. Syst.*, 2024.
- [77] J. Duan and Z. Shen, "Inversion of the Permeability Coefficient of a High Core Wall Dam Based on a BP Neural Network and the Marine Predator Algorithm," *Appl. Sci.*, vol. 14, no. 10, 2024.
- [78] H. H. Ellithy, H. M. Hasanien, M. Alharbi, M. A. Sobhy, A. M. Taha, and M. A. Attia, "Marine Predator Algorithm-Based Optimal PI Controllers for LVRT Capability Enhancement of Grid-Connected PV Systems," *Biomimetics*, vol. 9, no. 2, 2024.
- [79] X. Jiang, H. Zhan, J. Yu, and R. Wang, "Multi-stage manufacturing process parameter optimization method based on improved marine predator algorithm," *Eng. Res. Express*, vol. 6, no. 2, 2024.
- [80] W. Aribowo, H. Suryoatmojo, and F. A. Pamuji, "Novel hybrid of marine predator algorithm - Aquila optimizer for droop control in DC microgrid," *Int. J. Electr. Comput. Eng.*, vol. 14, no. 4, pp. 3703 – 3715, 2024.
- [81] Q. Wang and Y. Huang, "Novel path planning method using marine predator algorithm for mobile robot," *Arch. Control Sci.*, vol. 34, no. 1, pp. 225 – 242, 2024.
- [82] J. Hussain, R. Zou, S. Akhtar, and K. A. Abouda, "Design of cascade P-P-FOPID controller based on marine predators algorithm for load frequency control of electric power systems," *Electr. Eng.*, 2024.
- [83] S. Qiu, J. Zhao, X. Zhang, F. Chen, and Y. Wang, "Improved binary marine predator algorithm-based digital twin-assisted edge-computing offloading method," *Futur. Gener. Comput. Syst.*, vol. 155, pp. 437 – 446, 2024.
- [84] J. Liu, L. Li, and Y. Liu, "Enhanced marine predators algorithm optimized support vector machine for IGBT switching power loss estimation," *Meas. Sci. Technol.*, vol. 35, no. 1, 2024.
- [85] M. Kumar, K. Rajwar, and K. Deep, "Analysis of Marine Predators Algorithm using BIAS toolbox and Generalized Signature Test," *Alexandria Eng. J.*, vol. 95, pp. 38 – 49, 2024.
- [86] K. Tan, L. Zhu, and X. Wang, "A Hyperspectral Feature Selection Method for Soil Organic Matter Estimation Based on an Improved Weighted Marine Predators Algorithm," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–1, 2024.
- [87] S. Wang et al., "Boosting aquila optimizer by marine predators algorithm for combinatorial optimization," *J. Comput. Des. Eng.*, vol. 11, no. 2, pp. 37 – 69, 2024.
- [88] R. Özdemir, M. Taşyürek, and V. Aslantaş, "Improved Marine Predators Algorithm and Extreme Gradient Boosting (XGBoost) for shipment status time prediction," *Knowledge-Based Syst.*, vol. 294, 2024.
- [89] J. Geng et al., "Power system differentiation planning based on an improved marine predator algorithm," *Int. J. Low-Carbon Technol.*, vol. 19, pp. 1623 – 1632, 2024.
- [90] P. Yan, J. Wang, W. Wang, G. Li, Y. Zhao, and Z. Wen, "Transformer fault diagnosis based on MPA-RF algorithm and LIF technology," *Meas. Sci. Technol.*, vol. 35, no. 2, 2024.
- [91] H. Naseri, A. Golroo, M. Shokoochi, and A. H. Gandomi, "Sustainable pavement maintenance and rehabilitation planning using the marine predator optimization algorithm," *Struct. Infrastruct. Eng.*, vol. 20, no. 3, pp. 340 – 352, 2024.
- [92] K. Rezaei and O. S. Fard, "Multi-strategy enhanced Marine Predators Algorithm with applications in engineering optimization and feature selection problems," *Appl. Soft Comput.*, vol. 159, 2024.
- [93] S. N. Makhadmeh, M. A. Al-Betar, A. K. Abasi, A. Al-Redhaei, O. A. Alomari, and S. Kouka, "A Hybrid Marine Predators Algorithm with Particle Swarm Optimization Using Renewable Energy Sources for Energy Scheduling Problem-Based IoT," *Arab. J. Sci. Eng.*, 2024.
- [94] S. Ummadisetty and M. Tatineni, "A Novel Modified Marine Predator Algorithm (MMPA) based Automated Atrial Fibrillation Detection (AAFD) System using ECG Signals," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 1, pp. 708 – 719, 2024.
- [95] W. Wu, X. Liu, M. Gu, S. Ding, Y. Zhang, and X. Wei, "Scraper factors investigation on Al₂O₃ paste flow based on marine predators algorithm-bidirectional gated recurrent unit pseudo-lattice Boltzmann method for stereolithography molding," *Phys. Fluids*, vol. 36, no. 1, 2024.
- [96] Z. Wang, H. Zhao, X. Bao, and T. Wu, "Multi-objective optimal allocation of water resources based on improved marine predator algorithm and entropy weighting method," *Earth Sci. Informatics*, vol. 17, no. 2, pp. 1483 – 1499, 2024.
- [97] S. Moradi-Far, P.-S. Ashofteh, and H. A. Loáiciga, "Development of the marine predators algorithm for optimizing the performance of water supply reservoirs," *Environ. Dev. Sustain.*, 2024.
- [98] X. Qin, S. Zhang, X. Dong, H. Shi, and L. Yuan, "Classification of high-dimensional imbalanced biomedical data based on spectral clustering SMOTE and marine predators algorithm," *J. Intell. Fuzzy Syst.*, vol. 46, no. 4, pp. 8709 – 8728, 2024.
- [99] Y. Wang, J. Feng, J. Zhang, and Y. Chen, "Finding Community Modules for Brain Networks Combined Uniform Design with Marine Predators Algorithm," *IEEE Access*, pp. 1–1, 2024.
- [100] J. Wang, Z. Wang, D. Zhu, S. Yang, J. Wang, and D. Li, "Reinforcement learning marine predators algorithm for global optimization," *Cluster Comput.*, 2024.
- [101] M. Ramezani, D. Bahmanyar, and N. Razmjoooy, "A New Improved Model of Marine Predator Algorithm for Optimization Problems," *Arab. J. Sci. Eng.*, vol. 46, no. 9, pp. 8803–8826, 2021.
- [102] L. Chen, C. Hao, and Y. Ma, "A Multi-Disturbance Marine Predator Algorithm Based on Oppositional Learning and Compound Mutation," *Electron.*, vol. 11, no. 24, 2022.
- [103] C. Qin and B. Han, "A Novel Hybrid Quantum Particle Swarm Optimization With Marine Predators for Engineering Design Problems," *IEEE Access*, vol. 10, pp. 129322–129343, 2022.
- [104] M. Han, Z. Du, H. Zhu, Y. Li, Q. Yuan, and H. Zhu, "Golden-Sine dynamic marine predator algorithm for addressing engineering design optimization," *Expert Syst. Appl.*, vol. 210, 2022.
- [105] A. A. Dehkordi, B. Etaati, M. Neshat, and S. Mirjalili, "Adaptive Chaotic Marine Predators Hill Climbing Algorithm for Large-Scale Design Optimizations," *IEEE ACCESS*, vol. 11, pp. 39269–39294, 2023.
- [106] S. Zhao, Y. Wu, S. Tan, J. Wu, Z. Cui, and Y.-G. Wang, "QLMPA: A quasi-opposition learning and Q-learning based marine predators algorithm," *Expert Syst. Appl.*, vol. 213, 2023.
- [107] Z. Gao, Y. Zhuang, C. Chen, and Q. Wang, "Hybrid modified marine predators algorithm with teaching-learning-based optimization for global optimization and abrupt motion tracking," *Multimed. Tools Appl.*, vol. 82, no. 13, pp. 19793–19828, 2023.
- [108] S. Kumar et al., "Chaotic marine predators algorithm for global optimization of real-world engineering problems," *Knowledge-Based Syst.*, vol. 261, 2023.
- [109] M. Abd Elaziz, D. Mohammadi, D. Oliva, and K. Salimifard, "Quantum marine predators algorithm for addressing multilevel image segmentation," *Appl. Soft Comput.*, vol. 110, 2021.
- [110] L. Abualigah, N. K. Al-Okbi, M. A. Elaziz, and E. H. Houssein, "Boosting Marine Predators Algorithm by Salp Swarm Algorithm for Multilevel Thresholding Image Segmentation," *Multimed. Tools Appl.*, vol. 81, no. 12, pp. 16707–16742, 2022.
- [111] M. Abdel-Basset, D. El-Shahat, R. K. Chakraborty, and M. Ryan, "Parameter estimation of photovoltaic models using an improved marine predators algorithm," *Energy Convers. Manag.*, vol. 227, 2021.

- [112] N. Wang, J. S. Wang, L. F. Zhu, H. Y. Wang, and G. Wang, "A Novel Dynamic Clustering Method by Integrating Marine Predators Algorithm and Particle Swarm Optimization Algorithm," *IEEE Access*, vol. 9, pp. 3557–3569, 2021.
- [113] Z. Du, X. Chen, Q. Zhang, and Y. Yang, "An extended state observer-based sliding mode control method for hydraulic servo system of marine stabilized platforms," *Ocean Eng.*, vol. 279, Jul. 2023.
- [114] W. Yang, K. Xia, T. Li, M. Xie, and F. Song, "A multi-strategy marine predator algorithm and its application in joint regularization semi-supervised ELM," *Mathematics*, vol. 9, no. 3, pp. 1–34, 2021.
- [115] P. Lan, K. Xia, Y. Pan, and S. Fan, "An Improved GWO Algorithm Optimized RVFL Model for Oil Layer Prediction," *ELECTRONICS*, vol. 10, no. 24, Dec. 2021.
- [116] S. Yadav, S. K. Saha, R. Kar, and D. Mandal, "EEG/ERP signal enhancement through an optimally tuned adaptive filter based on marine predators algorithm," *Biomed. Signal Process. Control*, vol. 73, 2022.
- [117] J. Yan, H. Liu, S. Yu, X. Zong, and Y. Shan, "Classification of Urban Green Space Types Using Machine Learning Optimized by Marine Predators Algorithm," *Sustain.*, vol. 15, no. 7, 2023.
- [118] H. Jia, K. Sun, Y. Li, and N. Cao, "Improved marine predators algorithm for feature selection and SVM optimization," *KSII Trans. Internet Inf. Syst.*, vol. 16, no. 4, pp. 1128–1145, 2022.
- [119] R. K. G. Radhakrishnan, U. Marimuthu, P. K. Balachandran, A. M. M. Shukry, and T. Senju, "An Intensified Marine Predator Algorithm (MPA) for Designing a Solar-Powered BLDC Motor Used in EV Systems," *SUSTAINABILITY*, vol. 14, no. 21, Nov. 2022.
- [120] Y. Djenouri, A. Belhadi, G. Srivastava, J. C.-W. Lin, and A. Yazidi, "Interpretable intrusion detection for next generation of Internet of Things," *Comput. Commun.*, vol. 203, pp. 192–198, Apr. 2023.
- [121] A. M. Shaheen, A. M. Elsayed, R. A. El-Sehiemy, S. Kamel, and S. S. M. Ghoneim, "A modified marine predators optimization algorithm for simultaneous network reconfiguration and distributed generator allocation in distribution systems under different loading conditions," *Eng. Optim.*, vol. 54, no. 4, pp. 687–708, 2022.
- [122] N. H. Khan, R. Jamal, M. Ebeed, S. Kamel, H. Zeinoddini-Meymand, and H. M. Zawbaa, "Adopting Scenario-Based approach to solve optimal reactive power Dispatch problem with integration of wind and solar energy using improved Marine predator algorithm," *AIN SHAMS Eng. J.*, vol. 13, no. 5, Sep. 2022.
- [123] Q. Yin, Y. Zheng, B. Wang, and Q. Zhang, "Design of Constraint Coding Sets for Archive DNA Storage," *IEEE-ACM Trans. Comput. Biol. Bioinforma.*, vol. 19, no. 6, pp. 3384–3394, 2022.
- [124] R. Ahmad, M. Awais, N. Kausar, and T. Akram, "White Blood Cells Classification Using Entropy-Controlled Deep Features Optimization," *DIAGNOSTICS*, vol. 13, no. 3, Feb. 2023.
- [125] Q. He, Z. Lan, D. Zhang, L. Yang, and S. Luo, "Improved Marine Predator Algorithm for Wireless Sensor Network Coverage Optimization Problem," *Sustain.*, vol. 14, no. 16, 2022.
- [126] E. H. Houssein, D. S. Abdelminaam, I. E. Ibrahim, M. Hassaballah, and Y. M. Wazery, "A Hybrid Heartbeats Classification Approach Based on Marine Predators Algorithm and Convolution Neural Networks," *IEEE Access*, vol. 9, pp. 86194–86206, 2021.
- [127] N. Dharavat, S. K. Sudabattula, and V. Suresh, "Optimal Integration of Distributed Generators (DGs) Shunt Capacitors (SCs) and Electric Vehicles (EVs) in a Distribution System (DS) using Marine Predator Algorithm," *Int. J. Renew. ENERGY Res.*, vol. 12, no. 3, pp. 1637–1650, Sep. 2022.
- [128] R. Gong, D. Li, L. Hong, and N. Xie, "Task scheduling in cloud computing environment based on enhanced marine predator algorithm," *Clust. Comput. J. NETWORKS Softw. TOOLS Appl.*, 2023.
- [129] K. Prema and J. Visumathi, "A Novel Marine Predators Optimization based Deep Neural Network for Quality and Shelf-Life Prediction of Shrimp," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, pp. 65–72, 2023.
- [130] F. Cuevas, O. Castillo, and P. Cortés-Antonio, "Generalized Type-2 Fuzzy Parameter Adaptation in the Marine Predator Algorithm for Fuzzy Controller Parameterization in Mobile Robots," *Symmetry (Basel)*, vol. 14, no. 5, 2022.
- [131] A. H. Elmetwaly, A. A. ElDesouky, A. I. Omar, and M. A. Saad, "Operation control, energy management, and power quality enhancement for a cluster of isolated microgrids," *AIN SHAMS Eng. J.*, vol. 13, no. 5, Sep. 2022.
- [132] S. Ali, A. Bhargava, A. Saxena, and P. Kumar, "A Hybrid Marine Predator Sine Cosine Algorithm for Parameter Selection of Hybrid Active Power Filter," *MATHEMATICS*, vol. 11, no. 3, Feb. 2023.
- [133] Q. Fu, Q. Li, and X. Li, "An improved multi-objective marine predator algorithm for gene selection in classification of cancer microarray data," *Comput. Biol. Med.*, vol. 160, 2023.
- [134] D. Yousri, A. Fathy, and H. Rezk, "A new comprehensive learning marine predator algorithm for extracting the optimal parameters of supercapacitor model," *J. Energy Storage*, vol. 42, 2021.
- [135] M. Abdel-Basset, R. Mohamed, M. Elhoseny, R. K. Chakraborty, and M. Ryan, "A Hybrid COVID-19 Detection Model Using an Improved Marine Predators Algorithm and a Ranking-Based Diversity Reduction Strategy," *IEEE Access*, vol. 8, pp. 79521–79540, 2020.
- [136] M. A. A. Al-qaness, A. A. Ewees, H. Fan, L. Abualigah, and M. A. Elaziz, "Boosted ANFIS model using augmented marine predator algorithm with mutation operators for wind power forecasting," *Appl. Energy*, vol. 314, 2022.
- [137] Z. Li, B. Wang, B. Zhu, Q. Wang, and W. Zhu, "Thermal error modeling of electrical spindle based on optimized ELM with marine predator algorithm," *Case Stud. Therm. Eng.*, vol. 38, 2022.
- [138] A. H. Yakout, M. A. Attia, and H. Kotb, "Marine Predator Algorithm based Cascaded PIDA Load Frequency Controller for Electric Power Systems with Wave Energy Conversion Systems," *Alexandria Eng. J.*, vol. 60, no. 4, pp. 4213–4222, 2021.
- [139] W. Aribowo, B. Suprianto, R. Rahmadian, M. Widyartono, A. L. Wardani, and A. Prapanca, "Optimal tuning fractional order PID based on marine predator algorithm for DC motor," *Int. J. Power Electron. Drive Syst.*, vol. 14, no. 2, pp. 762–770, 2023.
- [140] M. Abd Elaziz, A. A. Ewees, D. Yousri, L. Abualigah, and M. A. A. Al-qaness, "Modified marine predators algorithm for feature selection: case study metabolomics," *Knowl. Inf. Syst.*, vol. 64, no. 1, pp. 261–287, 2022.
- [141] M. Z. Islam et al., "Marine predators algorithm for solving single-objective optimal power flow," *PLoS One*, vol. 16, no. 8, 2021.
- [142] A. H. Yakout, H. M. Hasanien, and H. Kotb, "Proton Exchange Membrane Fuel Cell Steady State Modeling Using Marine Predator Algorithm Optimizer," *AIN SHAMS Eng. J.*, vol. 12, no. 4, pp. 3765–3774, Dec. 2021.
- [143] L. V. Ho et al., "A hybrid computational intelligence approach for structural damage detection using marine predator algorithm and feedforward neural networks," *Comput. & Struct.*, vol. 252, Aug. 2021.
- [144] P. D. Kusuma and R. A. Nugrahaeni, "Stochastic Marine Predator Algorithm with Multiple Candidates," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 4, pp. 241–251, 2022.
- [145] R. G. Mohamed, M. Ao. E. H. of L.-S. W. F. U. M. P. A. Ebrahim, Z. M. Alaas, and M. M. R. Ahmed, "Optimal Energy Harvesting of Large-Scale Wind Farm Using Marine Predators Algorithm," *IEEE Access*, vol. 10, pp. 24995–25004, 2022.
- [146] K. Zhong, G. Zhou, W. Deng, Y. Zhou, and Q. Luo, "MOMPA: Multi-objective marine predator algorithm," *Comput. Methods Appl. Mech. Eng.*, vol. 385, Nov. 2021.
- [147] Q. Fu et al., "An improved multi-objective marine predator algorithm for gene selection in classification of cancer microarray data," *Expert Syst. Appl.*, vol. 9, no. 4, pp. 340–350, 2022.
- [148] P. Jangir, H. Buch, S. Mirjalili, and P. Manoharan, "MOMPA: Multi-objective marine predator algorithm for solving multi-objective optimization problems," *Evol. Intell.*, vol. 16, no. 1, pp. 169–195, Feb. 2023.

The Current Challenges Review of Deep Learning-Based Nuclei Segmentation of Diffuse Large B-Cell Lymphoma

Gei Ki Tang¹, Chee Chin Lim², Faezahtul Arbaeyah Hussain³, Qi Wei Oung⁴, Aidy Irman Yazid⁵, Sumayyah Mohammad Azmi⁶, Haniza Yazid⁷, Yen Fook Chong⁸

Faculty of Electronic Engineering and Technology, Universiti Malaysia Perlis, Arau, Perlis, Malaysia^{1, 2, 4, 7}

Sport Engineering Research Centre, Universiti Malaysia Perlis, Arau, Perlis, Malaysia^{2, 8}

Hospital Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia^{3, 5, 6}

Department of Pathology, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan Malaysia^{3, 5, 6}

Advanced Communication Engineering (ACE) Centre of Excellence, Universiti Malaysia Perlis, Arau, Perlis, Malaysia⁴

Abstract— Diffuse Large B-Cell Lymphoma stands as the most prevalent form of non-Hodgkin lymphoma worldwide, constituting approximately 30 percent of cases within this diverse group of blood cancers affecting the lymphatic system. This study addresses the challenges associated with the accurate DLBCL segmentation and classification, including difficulties in identifying and diagnosing DLBCL, manpower shortage, and limitations of manual imaging methods. The study highlights the potential of deep learning to effectively segment and classify DLBCL types. The implementation of such technology has the potential to extract and preprocess image patches, identify, and segment the nuclei in DLBCL images, and classify DLBCL severity based on segmented nuclei counting.

Keywords—Deep learning; Diffuse Large B-Cell Lymphoma (DLBCL); lymphoma cancer; HoVerNet

I. INTRODUCTION

Diffuse Large B-Cell Lymphoma (DLBCL) stands as the most prevalent form of non-Hodgkin lymphoma (NHL), comprising approximately 30% of all cases within this diverse group of blood cancers affecting the lymphatic system. DLBCL is characterized by the rapid proliferation of malignant B-cells in lymph nodes, bone marrow, and other lymphatic tissues. DLBCL can afflict individuals of any age, with a predilection for those over 60 [1]. Given its life-threatening nature and variable clinical outcomes, precise diagnosis assumes paramount importance in the management of DLBCL. The identification and quantification of cell nuclei within tissue samples emerge as crucial for assessing tumor characteristics and grading, thus guiding treatment decisions. Deep learning-based nuclei segmentation offers a promising solution, potentially enhancing diagnostic accuracy and efficiency, to streamline this labor-intensive and time-consuming task.

Accurate diagnosis and staging of DLBCL are critical for determining optimal treatment and prognosis. Nuclei segmentation and classification, a pivotal component of DLBCL tissue image analysis, allow for the identification and quantification of tumor cells. Conventional nuclei segmentation and classification methods are often time-consuming, labor-intensive, and error-prone, making the need for deep learning

models readily apparent. However, the complexity of DLBCL samples, such as tissue heterogeneity, staining changes, and complex cell interactions, requires complex and accurate deep learning models to address these challenges [1].

This paper explores the current challenges in nuclei segmentation and classification for DLBCL using deep learning methods. It provides an extensive review of the techniques, preprocessing methods, and segmentation approaches applied to DLBCL analysis. Furthermore, the paper highlights advancements in the field and identifies gaps for future exploration, aiming to inspire further research and innovation in digital pathology.

This paper is structured as follows: Section II reviews related work, highlighting the advancements and gaps in deep learning-based nuclei segmentation and classification for DLBCL. Section III describes the methods, including preprocessing techniques, model architectures, and evaluation metrics. Section IV presents the results and discussions, comparing the performance of state-of-the-art deep learning methods. Finally, Section V concludes the study, summarizing key findings and future research directions. This structure aims to provide readers with a comprehensive understanding of the challenges and contributions in the field.

II. RELATED WORK: ADVANCEMENTS IN DIAGNOSIS AND STAGING OF DIFFUSE LARGE B-CELL LYMPHOMA

A. Symptoms, Risk Factors, and Causes of Diffuse Large B-Cell Lymphoma

DLBCL can cause many different symptoms. One of the most common signs is the painless swelling of lymph nodes. These are lumps under the skin, usually found in places like the neck, armpit, or groin. This might happen, along with other signs. For example, losing weight for no reason, always feeling tired, or sweating a lot at night. Also, patients with DLBCL may experience fever from time to time, which further signals the response of the body to lymphoma. Sometimes, DLBCL might affect abdominal organs, leading to symptoms like abdominal pain or swelling. The symptoms show up based on where and how big the bothered lymph nodes or organs are. DLBCL can also have other unique symptoms. These vary based on where

in the body they show up. For example, if it is in the chest or lungs, the patient might have trouble breathing, coughing, or having chest pain. Those with DLBCL in their gastrointestinal tract could feel stomach pain, nausea, vomiting, and changes in bowel habits. When DLBCL affects the brain or spinal cord, it might lead to headaches, behavior changes, or even seizures. In some cases, DLBCL could cause problems with the skin. This could look like a rash or bumps. These skin changes might give doctors extra clues during physical examinations.

There are several risk factors and causes that have been identified. Advancing age is a prominent factor, as DLBCL is more prevalent in individuals over the age of 60, and the incidence increases with age. This age-related susceptibility suggests a cumulative effect over time, possibly linked to cellular changes and immune system alterations associated with ageing. A compromised immune system, whether due to medical conditions such as Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome (HIV/AIDS) or the use of immunosuppressive drugs post-organ transplant, is another significant risk factor [2]. The impaired immune surveillance in these scenarios may create an environment conducive to the uncontrolled growth of lymphoid cells, fostering the development of DLBCL.

The Epstein-Barr virus (EBV), belonging to the herpesvirus family, has been linked to an increased risk of DLBCL, particularly in immunocompromised individuals [3]. The ability of EBV to infect B-cells and potentially contribute to the transformation of these cells underscores its role in the lymphoma genic process. Besides, genetic factors also contribute, with a family history of lymphomas potentially elevating the risk. While the specific genetic mechanisms are not fully elucidated, ongoing research aims to uncover the intricate interplay of genetic and environmental factors in DLBCL development [2]. Other factors, such as autoimmune diseases and certain chemical exposures, have been explored for their potential roles in lymphomagenesis, though their associations remain complex and multifaceted.

B. Diagnosis Method and Staging of Diffuse Large B-Cell Lymphoma

DLBCL is typically diagnosed by removing a swollen lymph node or taking a sample of tissue from it and examining it under a microscope. This involves a minor procedure known as a biopsy, which is usually performed under local anesthesia or through a minor operation. Following the biopsy, expert pathologists use special staining, a test called flow cytometry, and chromosome analysis to determine the exact variant of DLBCL. DLBCL is also diagnosed using blood tests and imaging tests. Blood tests can help determine the overall health of the patient and detect any abnormalities in the blood cells. To determine the location and extent of the disease, imaging tests such as Magnetic Resonance Imaging (MRI) scans, Computed Tomography (CT) scans, Positron Emission Tomography (PET) scans, and ultrasounds are used. Apart from that, the current standard diagnosis method for DLBCL includes two tests: Fluorescence In-Situ Hybridization (FISH) tests and Immunohistochemistry (IHC) tests [4].

The staging of DLBCL depends on the extent of the disease and the organs involved. The Ann Arbor staging system is

commonly used to stage DLBCL for lymphoma. It classifies the disease into four stages of involvement, namely Stages I, II, III, and IV [5]. Stage I has been described because the cancer is found in the lymphatic zone or in only one organ. Stage II is indicated because cancer is found in two or more lymph nodes on one side of the lung or in one limb and in one or more lymph nodes on the same side of the lung. Furthermore, cancer found in different parts of the lymph nodes or on either side of the diaphragm is stage III. Lastly, stage IV means that the cancer has spread to one or more organs outside the lymphatic system, such as the liver, lungs, or bones.

1) *Biopsy of Diffuse Large B-Cell Lymphoma*: A biopsy is a crucial diagnostic method for DLBCL, which allows for the presence or absence of certain genetic alterations. In this procedure, a sample of the affected lymph node or tissue is extracted and examined under a microscope [6]. Immunohistochemical staining and molecular tests are then employed to detect the expression of MYC gene rearrangements. The presence of MYC rearrangements classifies the lymphoma as MYC+, indicating a potentially more aggressive form. Conversely, if there is no evidence of MYC rearrangements, the lymphoma is classified as MYC-.

2) *Blood tests*: Blood tests are essential in diagnosing DLBCL. A complete blood count (CBC) assesses various blood components, including white blood cells. Elevated white blood cell counts may indicate the presence of lymphoma. Being that B-cells grow and become mature in the bone marrow, the CBC count tests for anemia, thrombocytopenia, and/or leukopenia indicate the extent of bone marrow involvement in DLBCL.

Furthermore, the LDH level is a helpful predictor of treatment response, recurrence, and the severity of DLBCL. Increased levels of potassium, phosphorus, and uric acid combined with a drop in calcium could be signs of tumor lysis syndrome, a condition that can develop during chemotherapy. LDH levels, a specific blood marker, can be indicative of lymphoma activity. Elevated LDH levels may suggest a more aggressive disease. While these blood tests do not directly determine MYC status, abnormal results can prompt further investigations, including imaging studies and biopsies.

3) *Imaging tests*: Imaging tests are essential components of the diagnostic process for DLBCL, providing valuable information about the extent of the disease and its characteristics. MRI uses strong magnetic fields and radioactive waves to generate detailed images of the internal structures of the body. In DLBCL diagnosis, MRI is useful for assessing the involvement of lymph nodes and surrounding tissues, aiding in the accurate staging of the disease [7]. It offers high-resolution images that aid in determining the size, location, and characteristics of lymphoma masses. Besides, CT scans employ X-rays from multiple angles to create detailed, cross-sectional images of the body [7], [8]. CT scans are valuable in identifying and measuring lymph nodes affected by DLBCL. They help in assessing the extent of the disease, determining the stage, and identifying whether the lymphoma has spread to other organs or tissues.

Other than that, ultrasound is also one of the imaging techniques used to diagnose DLBCL. Ultrasound employs high-frequency sound waves to create real-time images of internal organs and tissues. While less commonly used than other imaging modalities, ultrasound can assist in evaluating abnormalities in lymph nodes. It is particularly useful for examining superficial lymph nodes and assessing potential changes in organ structures caused by lymphoma. Other than that, a PET scan is suitable to detect the progression of tumors and cancer. A PET uses a radiotracer to show the differences between healthy tissues and cancerous tissues. A radiotracer is injected into the patient, and the cancerous cells absorb more of the radiotracer. PET will detect the radiation given off by the tracer and produce color-coded images of the body that show both healthy and cancerous tissues. A special camera from a PET scan detects the radiation emitted by these cells. PET scans highlight areas with increased metabolic activity, helping to identify active lymphoma sites [7]. This is crucial for determining the extent of DLBCL, assessing response to treatment, and locating residual or recurrent lymphoma. In short, these imaging methods provide valuable information about the size, location, and characteristics of lymphoma lesions, aiding in the accurate diagnosis, staging, and ongoing management of DLBCL.

C. Current Standard Diagnosis Method

The current standard diagnosis method for DLBCL involves Fluorescence In-Situ Hybridization (FISH) tests and Immunohistochemistry (IHC) tests. The FISH test is a technique that is essential for identifying genetic abnormalities such as MYC translocations in DLBCL and offering insights into the severity of the disease. Besides, the IHC test is a method that involves the expression of specific proteins and helps in characterizing DLBCL subtypes, such as GCB subtypes and ABC subtypes. By integrating the FISH and IHC, it provides a holistic evaluation, combining genetic and protein expression data to guide accurate diagnosis, subtype classification, and personalized treatment planning for individuals with DLBCL.

1) *Fluorescence In-Situ Hybridization (FISH) Test:* Fluorescence In-Situ Hybridization (FISH) method is a molecular technique used to detect and locate the presence or absence of specific DNA sequences on chromosomes [4]. It uses fluorescent light that binds to only those parts of the chromosome that show a high degree of sequence complementarity. The FISH method is used to identify specific genetic abnormalities, such as the rearrangements of the MYC, BCL2, and BCL6 genes, to diagnose DLBCL [4]. The purpose of FISH in diagnosing DLBCL is to provide a more accurate diagnosis, which can guide treatment decisions. It enables the precise detection of genetic abnormalities that could be driving diseases like DLBCL [22]. One of the studies that used FISH methods for DLBCL analysis was conducted by Blanc Durand et al. [14]. The authors worked with pre-therapy FDG-PET/CT scans from 733 DLBCL patients.

FISH analysis, which can be performed using dual-color and dual-fusion cleavage probe methods, is a highly sensitive and accurate technique for detecting oncogene amplification in

human tissue samples. However, due to the high variability in MYC breakpoints, it may not identify all MYC abnormalities. Furthermore, the FISH method has some limitations and shortcomings. This method may not be universally applicable to all diseases due to its labor-intensive and demanding nature, making it a time-consuming procedure. The need for expensive techniques, especially when using fluorescence microscopy, is specific and sensitive, emphasizing the importance of elucidating the genetic status of DLBCL.

2) *Immunohistochemistry (IHC) Test:*

Immunohistochemistry (IHC) is a method used to visually detect the presence of specific proteins in cells or tissues. It involves the use of antibodies that bind to these proteins and a detection system that uses a colored or fluorescent dye to visualize the binding [23]. IHC studies have evolved, emerging as the most widely used test to characterize cancers and identify hidden metastatic sites, particularly in lymph nodes. The method is based on the specific binding of antibodies to antigens, allowing the detection and specific localization of molecules in cells and tissues. The primary analysis is typically conducted using a light microscope [9]. IHC plays a crucial role in cancer diagnosis, especially when specific tumor antigens are overexpressed in certain malignancies. Notably, IHC offers significant advantages, particularly in settings with limited resources and drawbacks. IHC provides qualitative information about the presence or absence of specific antigens but does not quantify the expression levels.

In DLBCL, IHC is used to identify abnormal protein expression of certain genes. For instance, IHC can be used as a screening test to identify cases of DLBCL and identify overexpression of the BCL2 protein, which is associated with a poor prognosis in DLBCL [23]. The purpose of IHC in diagnosing DLBCL is to provide a more accurate diagnosis, which can guide treatment decisions. It is an inexpensive and rapid test that can identify abnormal protein expression in mutated genes [23]. Furthermore, the intensity of marker expression can have prognostic implications. This limitation may impact the precision of the diagnosis. Apart from that, DLBCL comprises different subtypes with varying clinical behaviors. IHC alone may not always reliably distinguish between GCB subtypes and ABC subtypes. Gene expression profiling or additional molecular tests might be required for a more accurate subtype classification [10]. Also, IHC primarily provides information on protein expression but may not directly reveal underlying genetic alterations, such as gene mutations or chromosomal abnormalities.

III. RELATED WORK: IMAGE PROCESSING METHODS AND ARTIFICIAL INTELLIGENCE ALGORITHMS

A. *Image Processing Methods*

1) *Image patches:* Image patches are small, square regions of an image used for feature extraction. It plays a crucial role in identifying the regions of interest (ROIs) in medical images of DLBCL patients. These patches are needed for image processing tasks and algorithms such as image analysis, feature extraction, and applications involving AI algorithms. Besides,

it is normally used at the local level to analyze and manipulate image data and enable us to concentrate on specific areas of interest, such as structures and textures. This approach yields more detailed and accurate results, allowing feature extraction and providing reliable decisions. The size and shape of patches can vary depending on the task and requirements. It can range from a single pixel to a predefined window that features multiple pixels. The segmentation techniques may organize identical areas and distinguish them from the background by analyzing the color, appearance, or pixel's intensity, which allows for tasks like object detection and recognition.

Patch size and resolution are determined by specific applications and dataset requirements. For example, El Hussien et al. [16] obtained an overall mean Dice score of 0.825 from a quantitative assessment that included 15 manually annotated patches of 256×256 pixels. Furthermore, they conducted a study in which 10 Chronic Lymphocytic Leukemia (CLL), 12 accelerated Chronic Lymphocytic Leukemia (aCLL), and 8 Richter's Transformation (RT) digitally stained hematoxylin and eosin (H&E) slides from a lymph node excisional biopsy were chosen at random. The study used Aperio AT2 scanners to scan the slides, with an optical resolution of 20× magnification. These slides came from various patients, and a total of 25, 28, and 21 ROIs were from CLL, aCLL, and RT, respectively.

Wójcik et al. [17] employed 37,665 H&E-stained images obtained at 40× magnifications from a solitary WSI of DLBCL lymph nodes. Each image tile, sized at 512×512, underwent segmentation, with bounding boxes outlining individual nuclei, although no cell labels were provided. The images are standardized to 448×448 pixels, with additional randomly cropped tiles to augment the training dataset. Li et al. [18] focused on DLBCL tissue sections, capturing pathologic images initially at 400× original magnifications. The study began with 500 images obtained from labelled H&E-stained sections of lymph nodes. Apart from that, Swiderska-Chadaj et al. [20] digitized 42 H&E slides of DLBCL using a Panoramic 250 Flash II scanner at an objective magnification of 20×. These slides, with a pixel size of 0.24 μm, comprised an external validation set, allowing assessment across different hospital protocols.

Bándi, P. et al. [21] collected 100 WSI from various medical centers, comprising 10 tissue samples across different staining categories. Image patches were extracted from annotated areas using mask images according to the different pixel spacings, being 62.5, 250, and 10000 μm at respective resolutions of 0.5, 2, and 8 μm. Other than that, the research by Shankar, V. et al. [27] involves the H&E-stained tissue cores, pinpointed by hemapathologists using Qupath, for extracting image patches. These patches were obtained at 40× magnifications from each core. Based on the study by Swiderska-Chadaj, Z. et al. [29], the training dataset was derived from H&E-stained specimens. The image patches, each sized at 512×512 pixels, were extracted from slides at 5× magnification level with a pixel size of 1 μm for optimal analysis. Perry, C. et al. [31] involved a self-supervised phase, which included the slides to be scanned

at either 20× or 40× magnifications. The 40× images were converted to 20× for analysis. The WSI was divided into smaller patches used as model input by using patches of size 384×384 pixels.

Studies by El Hussien et al. [16], Wójcik et al. [17], and others have demonstrated the utility of image patches in DLBCL analysis, achieving high mean Dice scores and effectively capturing relevant features. However, while image patches excel in local analysis, they may struggle with capturing global context, which is crucial for a comprehensive understanding and diagnosis of DLBCL. Besides, weaknesses such as the absence of cell labels in datasets and variability in staining and scanning methods across the study by Wójcik et al. [17] pose challenges to consistency and accuracy. Opportunities lie in the potential for standardizing imaging and analysis protocols, augmenting training datasets with synthetic images, and applying transfer learning to enhance model performance. Conversely, threats include inter-laboratory variations that may limit model generalizability, computational resource constraints, and data privacy concerns related to patient-derived images. These factors collectively underscore the complexities and prospects of advancing the field of digital pathology for hematological malignancies.

2) *Preprocessing methods*: Preprocessing is a technique that is required to prepare image data for model input. For example, the fully connected layer of a CNN required that all images be stored in arrays of identical size. Model preprocessing may also reduce the training period and accelerate model inference. If the input images are very large, diminishing the size of the images will drastically reduce the time required to train the model without affecting model performance significantly. Basically, the preprocessing steps include orientation, resize, random flips, grayscale, and other different exposures that inhibit unforeseen distortions or improve certain characteristics of images essential to the deep learning pre-trained model.

Hamdi, M. et al. [11] used Gaussian filter to smooth the images, a Laplacian filter for edge detection, color normalization for standardization, resizing to a consistent resolution, and the use of Gradient Vector Flow for additional feature extraction. Besides, Vrabac et al. [12] focused on the employment of various preprocessing techniques to prepare histopathological images for analysis. The authors arranged the images in tissue microarrays (TMAs) and performed cell nucleus extraction from H&E-stained images. Additionally, they extracted features such as maximum area, minimum area, hull area, perimeter nucleus, maximum angle, ellipse perimeter, and ellipse area to capture morphological characteristics [12]. Basu and his team developed novel preprocessing methods for DLBCL classification. Although the exact preprocessing steps were not specified, the authors introduced in their attention map feature transformer, feature fusion techniques, and a specific loss function to improve the performance of the DLBCL classification model [13].

In the study conducted by Blanc-Durand et al. [14], they underwent a series of preprocessing steps. These steps included resampling, padding, cropping, and scaling of PET and CT

image data to ensure consistency. Additionally, adaptive thresholding was applied to segment images, and various features related to tumor characteristics, such as tumor heterogeneity, textural features, total tumor surfaces, and spatial dispersion, were extracted to provide a comprehensive set of features for classification. Ferrández and colleagues employed preprocessing methods tailored to medical imaging [15]. Gaussian filtering was applied to enhance image quality, and metabolic tumor volume (MTV) and standard uptake value (SUV) were computed to quantify metabolic activity. The authors also considered features related to tumor dissemination and textural features to capture the heterogeneity of DLBCL tumors.

El Hussien, S. et al. [16] focused on the preprocessing methods that involved annotating ROIs, calculating the ratio of the segmented nuclear contour area to its convex area, and measuring hull areas within these annotated regions. Besides, Graham, S. et al. [19] involved preprocessing methods such as Otsu thresholding, pixel intensity manipulation, color adjustments, and extraction of textural features to capture various characteristics within the images. Furthermore, Ferrández, M. C. et al. [24] studied how their preprocessing workflow incorporated normalization techniques, filtering procedures, max-pooling layer utilization, and rectified linear unit (ReLU) operations, possibly aimed at enhancing image quality and extracting relevant features. Other than that, Mohlman, J. S. et al. [25] involved the preprocessing methods, which are normalization techniques and edge detection methods, and utilized a deep network-based pixel-level concept, indicating a complex approach to feature extraction.

The other studies, including Farinha, F. et al. [26], Shankar, V. et al. [27], Jiang, C. et al. [28], Swiderska-Chadaj, Z. et al. [29], Steinbuss, G. et al. [30], Perry, C. et al. [31], Lisson, C. S. et al. [32], have been using the same preprocessing techniques. Their methodologies involved various preprocessing techniques, such as normalization, quality control thresholds, machine learning algorithms, filtering, and feature selection, aiming to enhance image quality, extract informative features, and facilitate accurate DLBCL classification.

Hamdi et al. [11], Vrabac et al. [12], and others have employed various preprocessing techniques tailored to DLBCL analysis, aiming to enhance image quality and extract informative features. While preprocessing can improve model performance and accelerate training, aggressive preprocessing may distort image features, leading to erroneous analysis results. The various preprocessing techniques used in DLBCL image analysis, such as feature extraction and Gaussian smoothing, highlight a reliable strategy for improving diagnostic accuracy. Despite their differences, these methods work together to provide a classification process that is more precise and effective, demonstrating the dynamic interaction between pathology-specific technology and medical knowledge.

3) *Data augmentation*: Data augmentation is a method for improving performance. It entails changing the color, brightness, or contrast of the existing training data by cropping, flipping, rotating, scaling, or changing the color, brightness, or contrast. Data augmentation can increase the variety and scope

of the training data. This can minimize excessive overfitting and make the model more resilient to different inputs. To implement data augmentation techniques for CNN training data, Python libraries such as TensorFlow, Keras, and OpenCV are used.

Basu, S. et al. [13] augmented their datasets through diverse techniques such as image rotations, horizontal and vertical flips, zoom scaling, as well as horizontal and vertical shifts. Wójcik, P. et al. [17] incorporated cell patch embedding, patch aggregation, random resizing, cropping, color jittering, and random flipping, aimed at organizing and augmenting image data for analysis. Other than that, Graham, S. et al. [19] used data augmentation methods such as flipping, rotation Gaussian blur, and median blur for enhanced feature variability. Bándi, P. et al. [21] also applied diverse augmentation techniques like horizontal mirroring, rotation, scaling, color, and contrast adjustments, additive Gaussian noise, and Gaussian blur for image enhancement and feature variability.

Mohlman, J. S. et al. [25] augmented their datasets through random horizontal flipping of images and random alteration of contrast, while Swiderska-Chadaj, Z. et al. [29] includes data augmentations such as brightness, contrast, saturation, rotation, gaussian noise, and gaussian blur. Besides, the data augmentation for the study of Perry, C. et al. [31] includes color jittering and channel shuffle. The study of Lisson, C. S. et al. [32] includes data augmentations such as random flip, gaussian blur, and gaussian noise.

Basu et al. [13], Wójcik et al. [17], and others have utilized data augmentation techniques to enhance the variability of DLBCL image data and improve model performance. However, the effectiveness of data augmentation depends on the appropriateness of the augmentation strategies and the quality of the generated samples. Moreover, excessive augmentation may introduce synthetic artifacts or distortions that do not accurately represent real-world variability. By simulating a variety of variations in medical images, these techniques improve generalization to new data and increase the size of the training dataset. However, the creation of high-quality samples and the choice of suitable tactics are prerequisites for the success of data augmentation. Excessive or improper augmentation can result in synthetic artefacts or distortions that may not accurately represent clinical scenarios and lead to inaccurate predictions, despite being necessary for the robustness of the model. As a result, the ability of data augmentation to provide realistic and clinically relevant variations without compromising the diagnostic integrity of the images serves as a gauge for its efficacy.

B. Artificial Intelligence Algorithm

The process of transferring information, data, and cognitive abilities to machines is known as AI. The primary objective of AI is to create independent machines with human-like thought and behavior. Through learning and problem-solving, these machines can mimic human behavior and carry out tasks. For resolving complex issues, most AI systems mimic natural intelligence. A branch of AI called machine learning employs statistical techniques to let machines learn from experience. Deep learning is a branch of machine learning that processes

information for specific analysis and subsequent action by modelling parts of the human brain with multi-layer neural networks. Hence, AI can be expressed more simply as the overall system, with machine learning and deep learning being its subsets. Deep learning is a subset machine learning, which employs neural networks to learn from massive amounts of data. The relationship between AI, machine learning, and deep learning are shown in Fig. 1.

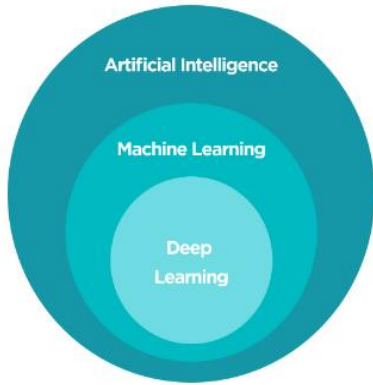


Fig. 1. Hierarchical relationship between artificial intelligence, machine learning, and deep learning.

C. Machine Learning Approaches

Machine learning can be used to extract features from medical images and classify them as either healthy or cancerous. These algorithms can be trained on large datasets of medical images to learn how to identify those that are indicative of cancer. Once trained, these algorithms can be used to analyze and classify new medical images.

1) *Multilayer Perceptron (MLP)*: An artificial neural network type called a Multilayer Perceptron (MLP) is frequently used for machine learning tasks like regression and classification. MLP can be used in the image processing and classification of DLBCL to identify features in medical images and categorize them as either benign or malignant. MLP is a multi-layered input layer, one or more hidden layers, and an output layer, which make up the feedforward neural network [40] [41], [42]. Every single node in the network's hierarchy is connected to it, and every connection has a weight. During training, the weights are changed to minimize the discrepancy between the expected and actual outputs. The structure of MLP in machine learning is shown in Fig. 2.

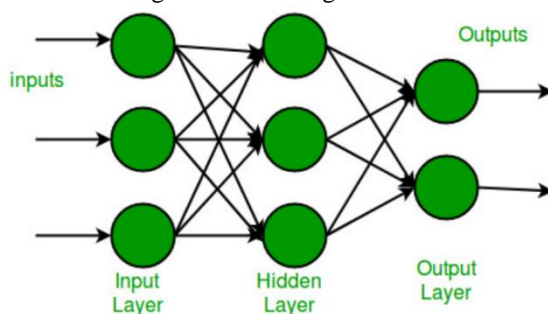


Fig. 2. Structure of Multilayer Perceptron (MLP) in Machine Learning [40].

The MLP emerges as a pivotal tool across several studies in medical imaging and data analysis. Carreras et al. [33], [34], [37] extensively employed MLP, alongside other statistical methods, in various scales and contexts. In their investigations involving 414 and 100 cases, respectively, they utilized MLP along with Mann-Whitney U tests, Kaplan-Meier analysis, and multivariate Cox regression to discern hazard ratios and risks. Additionally, Wagner et al. [35] and Chen et al. [36] explored different facets of image processing; while Wagner utilized grayscale images and specific filtering techniques like Rudin-Osher-Fatemi (ROF), Chen focused on feature extraction from biopsy specimens using solidity features and ROI annotation. Bhattamisra et al. [38] and Achi et al. [39] emphasized the role of MLP in handling vast geometric data and image analysis, respectively.

2) *Radial Basis Function (RBF)*: In machine learning, the Radial Basis Function (RBF) is a kernel function that is used to identify a regression line or non-linear classifier. RBF is capable of being used to extract characteristics from clinical pictures and categorize them as either benign or malignant in DLBCL image manipulation and classification. RBF compares two inputs according to how far apart they are in a high-dimensional space. The Gaussian kernel, also referred to as the RBF kernel, can be found in the Eq. (1).

$$K(x, x') = e^{-\gamma ||x-x'||^2} \quad (1)$$

where,

$$K(x, x') = \text{Radial basis function}$$
$$\gamma = \text{Width of the kernel}$$

The RBF stands out as a key computational approach utilised in several studies, notably alongside the MLP in medical data analysis. Carreras et al. [33], [34], [37] incorporated RBF networks in conjunction with MLPs to process and interpret diverse datasets. Specifically, they employed RBF alongside MLP in their analyses involving various statistical tests such as Mann-Whitney U tests, Kaplan-Meier analysis, and multivariate Cox regression, elucidating hazard ratios and risks across different case volumes. The utilisation of RBF underscored its relevance in enhancing the MLP's performance and classification and prediction taste, contributing to the robustness of models used in medical imaging and genomic analysis. The application of RBF within MLP architectures demonstrated its capacity to handle complex data structures, aiding in the extraction of valuable insights from medical datasets.

D. Deep Learning Approaches

In recent decades, there has been a lot of interest in the advanced field of ML known as DL. It has been extensively employed in numerous applications and has proven to be a successful ML technique for a few challenging problems. DL algorithms, such as CNN, are particularly effective for image processing and classification tasks. CNNs can learn to identify complex patterns in medical images and classify them with high accuracy. For example, CNN can be trained to identify specific features in medical images of DLBCL patients, such as the size and shape of a cancerous cell and use this information to

classify the images as either healthy or cancerous. Identification and manual classification are challenging tasks, particularly in the medical field. Thus, using different architectures to improve the classification of images requires the application of DL. The goal of image classification is to effectively identify and categorise the biomedical characteristics that have important benefits for many research and development domains.

Among deep learning architectures, such as U-Net, ResNet, and HoverNet have emerged as popular choices for nuclei segmentation in medical imaging. Recent studies have compared the performance of these architectures in the context of DLBCL segmentation and classification [11]-[39]. For example, U-Net demonstrated moderate success in segmenting nuclei but required extensive preprocessing and data augmentation to achieve consistent results. ResNet-based models showed improved feature extraction capabilities but were prone to overfitting with smaller datasets. HoVerNet excelled in cases involving nuclear overlap and heterogeneity but at the cost of increased computational complexity. These findings highlight the need for continued exploration and optimisation of deep learning methods tailored specifically to the nuances of DLBCL.

E. Convolutional Neural Network (CNN) Architecture

CNN are a subset of neural networks that are particularly adept at processing data using network-like topologies such as images. The binary representation that represents visual data is what makes up a digital image. It is made up of a grid-like arrangement of pixels with pixel values to indicate the colour and brightness of each pixel.

In the CNN architecture, it typically has three layers. First, there is the convolution layer. The convolutional layer is the fundamental component of a CNN, carrying the majority of the network's computational load. It works by performing a dot product operation between a limited area of the input image called the receptive field and a learnable matrix called a kernel. The kernel functions across the height and width of the picture during the forward pass. It is less extensive systematically but greater in depth than the image. This motion creates a 2D activation map that illustrates the response of the kernel at every spatial location. The total area of this activation map is determined by the sliding motion of the kernel, also known as the stride. The formula for the convolutional layer is expressed as in Eq. (2):

$$W_{out} = \frac{W-F+2P}{S} + 1 \quad (2)$$

where,

W_{out} = Output volume size

W = Squared input image

F = Receptive field size

P = Amount of zero padding

S = Stride

Second is the pooling layer. By calculating a summary statistic from the outputs in the vicinity, the pooling layer substitutes the network's output at specific points. This aids in

shrinking the representation's spatial size, which lowers the quantity of calculations and weights needed. Each of the sections of the representation is processed independently for the pooling operation. The formula for the pooling layer is expressed in Eq. (3):

$$W_{out} = \frac{W-F}{S} + 1 \quad (3)$$

where:

W_{out} = Output volume size

W = Squared input image

F = Receptive field size

S = Stride

Nonetheless, the most widely used method is max pooling, which provides the neighbourhood's maximum output. Lastly, the third layer involved is a fully connected layer (FC layer). As in a regular fully convolutional neural network (FCNN), all neurons in this layer are fully connected to all neurons in the layer that comes before and after. Hence, it can be calculated using the standard method of matrix multiplication and the bias effect. The relationship between the input data and the output is mapped with the aid of the FC layer.

Based on DLBCL nuclei segmentation in CNN, it normally uses a pre-trained model, epoch, optimiser, learning rate, and decay rate. A single run through the complete training dataset is referred to as an epoch. The model is trained on a new batch of dataset samples during each epoch. One hyperparameter that indicates the number of times the model is trained on the complete dataset is the number of epochs. Besides, during training, an optimiser is an algorithm that modifies the neural network's weights. The optimiser determines the way to modify the weights to minimise the loss by utilising the weights' gradients and the loss function. Also, learning rate refers to the weights of the neural network, which are updated to a certain extent based on the hyperparameter. A low learning rate can lead to more stable training, as a high learning rate will trigger the algorithm to converge more rapidly, potentially leading to unstable learning. Apart from that, decay rate controls the amount of learning rate that decreases following each epoch. The learning rate may decrease too rapidly or too slowly, depending on the decay rate, while a high decay rate may decrease the learning rate in a rapid way. Also, a pre-trained model in CNN is a model that is ready to use as the basis for a new model since it has been trained on a sizable dataset. When a new model needs a good setting, pre-trained models can help it perform better. Based on the studies [11-32], there are a few pre-trained models that are being utilised, such as DenseNet-201 [13], ResNet-50 [12], HoVerNet [12], [16], [17], [19], and U-Net [14], [15], [20], [24], [26], [28], [29]. The summaries for methods on DLBCL by using CNN architectures and ML are tabulated in Table I and Table II (Appendix).

1) *HoVerNet*: DLBCL nuclei segmentation is the process of highlighting nuclei in pathology images. HoVerNet, a specialised network, excels at this task by incorporating multiple branches for segmentation and classification into a unified framework. It takes advantage of nuclear pixel distances

from their centres of mass, which is critical for segmenting clustered nuclei found in DLBCL images. The network's dedicated up sampling branch aids in the classification of different nuclear types. The efficacy of HoVerNet in DLBCL nuclei segmentation stems from its ability to handle complex arrangements, which contributes to improved pathology image analysis.

Based on the study by Vrabac et al. [12], they employed HoVerNet as their chosen pre-trained model. Besides, Hussein et. al. [16] utilised HoVerNet as their pre-trained model for the analysis of CLL, aCLL, and RT cases. Wójcik et al. [17] employed HoVerNet as their chosen pre-trained model, training it for 800 epochs. Lastly, Graham et al. [19] use HoVerNet as their pre-trained model, training it for 50 epochs with the Adam optimiser.

2) *U-Net*: U-Net, a well-known CNN architecture, is widely used in image segmentation, including the segmentation of DLBCL nuclei. U-Net extracts image features from the encoder and reconstructs a segmentation map from the decoder, which is made up of an encoder and decoder network linked by a bottleneck layer. U-Net can be differentiated into two-dimensional U-Net (2D U-Net) and three-dimensional U-Net (3D U-Net).

Based on the study by Blanc-Durand et al. [14], they utilised the 3D U-Net architecture as the pre-trained model for their pre-therapy FDG-PET/CT scans from 733 patients with DLBCL. The optimizer used was Adam. Similarly, Ferrández et al. [15] employed the 3D U-Net architecture as the pre-trained model. The optimisation was conducted using the Adam optimiser, with an epoch setting of 200. The learning rate was set at 0.00005, alongside a decay rate of 0.000001. Furthermore, Swiderska-Chadaj et al. [20] utilised U-Net as their chosen pre-trained model. In addition, Ferrández et al. [24] employed 3D U-Net as their chosen pre-trained model. The optimiser used was Adam, with an epoch setting of 200, a learning rate of 0.00005, and a decay rate of 0.000001. Other than that, Farinha, F. et al. [26] employed U-Net as their chosen pre-trained model. The epoch setting of 150, a learning rate of 0.0001. Lastly, Jiang, C. et al. [28] utilised 3D U-Net model and was trained for 1000 epochs with a learning rate of 0.01 and Nesterov momentum set to 0.99. According to the work by Swiderska-Chadaj, Z. et al. [29], the U-Net model was pre-trained using the Adam optimizer for 500 epochs with a learning rate of 0.0005.

3) *ResNet-50*: ResNet-50, which is made up of many residual blocks, aids in the learning of complex characteristics that are required for accurate nuclei segmentation and classification and performs well in DLBCL nuclei segmentation. The depth of its architecture allows for the capture of complicated nuclear characteristics, improving segmentation precision in DLBCL pathology images. Its residual connections promote gradient flow, which aids in the learning of complicated nuclear patterns, which is important in DLBCL analysis. Based on the study by Vrabac et al. [12], they employed ResNet-50 as their chosen pre-trained model.

4) *DenseNet-201*: DenseNet-201 has a distinct architecture that is advantageous for DLBCL nuclei segmentation. Each layer in DenseNet-201 receives input from all preceding layers, resulting in dense connections throughout the network. This connectivity pattern allows for efficient information flow, which is important for capturing complex nuclear features in DLBCL pathology images. The densely connected blocks of DenseNet-201 allow for feature reuse, increasing model efficiency while effectively separating nuclei. This architecture reduces information loss, which is especially useful when segmenting densely clustered nuclei, which is common in DLBCL samples. According to Basu et al. [13], they used DenseNet-201 as their pre-trained model, optimising it using the Adam optimiser. The learning rate employed was 0.0001 for training.

In the realm of DLBCL nuclei segmentation, various deep learning architectures have been employed to enhance the accuracy and efficiency of pathology image analysis. HoVerNet stands out for its ability to segment clustered nuclei through its multi-branch framework, proving effective in complex arrangements. U-Net, particularly in its 3D form, is widely adopted for its feature extraction and reconstruction capabilities, with several studies optimizing it for large patient datasets. ResNet-50's depth captures intricate nuclear characteristics, while DenseNet-201's dense connectivity ensures comprehensive feature capture and efficient information flow. These architectures, through their unique strengths, contribute significantly to the progress in digital pathology, offering promising avenues for improved diagnostic methods in DLBCL.

While both machine learning and deep learning approaches have shown promise in DLBCL image analysis, they have distinct advantages and limitations. Machine learning techniques, such as MLPs and RBF networks, offer interpretability and ease of implementation but may struggle with capturing complex patterns in high-dimensional data. On the other hand, deep learning approaches, particularly CNNs, excel in learning hierarchical representations directly from raw data but require large amounts of annotated data and computational resources for training.

In short, the choice between machine learning and deep learning approaches in DLBCL image analysis depends on factors such as dataset size, computational resources, and the complexity of the underlying patterns. Integrating both approaches and exploring hybrid models may offer a promising avenue for future research in DLBCL diagnosis and treatment.

IV. DISCUSSIONS

A. Role of Preprocessing and Augmentation

Preprocessing is a cornerstone of successful deep learning applications in DLBCL analysis. Techniques such as Gaussian smoothing, color normalization, and artifact removal ensure consistent image quality across datasets. Hamdi et al. [11] employed Gaussian and Laplacian filters to enhance features, while Graham et al. [19] utilized Otsu thresholding and pixel intensity adjustments to refine segmentation inputs.

Data augmentation further aids in addressing dataset limitations by artificially increasing sample diversity. Techniques such as rotation, flipping, and noise addition have been widely applied. For instance, Wójcik et al. [17] and Swiderska-Chadaj et al. [20] incorporated augmentation strategies to improve model effectiveness and generalizability. However, excessive augmentation risks introducing synthetic artifacts, which could affect real-world applicability.

B. Performance of Deep Learning Models

The performance of HoVerNet and U-Net was evaluated against other deep learning methods, including DenseNet-201 and ResNet-50. HoVerNet and U-Net have emerged as important architectures for nuclei segmentation in DLBCL. HoVerNet's multi-branch framework enables simultaneous segmentation and classification, excelling in cases involving clustered and overlapping nuclei. Studies such as Graham et al. [19] demonstrated its ability to achieve a high Dice score of 0.869 by leveraging nuclear pixel distances and incorporating classification branches. U-Net, on the other hand, employs an encoder-decoder structure that effectively extracts and reconstructs features, as highlighted by Blanc-Durand et al. [14], who used a 3D variant of U-Net for PET/CT imaging.

While these models show promise, their performance heavily depends on preprocessing pipelines. For example, resizing, normalization, and augmentation methods were critical for improving model accuracy in studies like Hamdi et al. [11] and Ferrández et al. [15]. Despite their strengths, challenges such as overfitting, dataset variability, and computational demands remain significant.

C. Challenges and Limitations

Several challenges persist in applying deep learning to DLBCL segmentation such as variability in data quality. The differences in staining protocols and imaging equipment introduce inconsistencies that can affect model performance. Studies like Vrabac et al. [12] underscore the need for standardized preprocessing pipelines. Besides, limited annotated data is one of the limitations. The scarcity of labeled datasets restricts model training and evaluation. Transfer learning and synthetic data generation offer potential solutions but require further refinement. Lastly, computational complexity poses challenges in applying deep learning to DLBCL. Advanced architectures such as HoVerNet demand significant computational resources, which may limit their accessibility in resource-constrained settings.

D. Opportunities and Future Directions

Advancements in artificial intelligence present opportunities to overcome existing challenges. Generative adversarial networks (GANs) can be used to augment datasets with realistic synthetic images, while hybrid models that combine U-net and HoVerNet architectures could leverage the strengths of both. Additionally, transfer learning can facilitate model adaptation across diverse datasets, improving generalizability.

Future research should focus on developing lightweight architectures for resource-limited environments. Besides, establishment of standardized datasets and evaluation metrics for fair benchmarking should be carried on. This advancement

could significantly enhance the diagnostic accuracy and efficiency of DLBCL analysis, ultimately improving patient outcomes.

V. CONCLUSION

In conclusion, deep learning approaches have demonstrated their potential as useful and efficient algorithms for segmentation and classification of DLBCL. Based on the literature review, there are some related studies that have been done on the deep learning-based nuclei segmentation of DLBCL. These studies have demonstrated the effectiveness of deep learning in improving DLBCL diagnosis. Thus, it is believed that further exploration and enhancement of the nuclei segmentation and classification will provide a wide alternative way to count and diagnose the severity level of DLBCL. Deep learning offers an alternative to traditional methods, opening opportunities for further research and practical applications.

INSTITUTIONAL REVIEW BOARD STATEMENT

This study was conducted in accordance with the Declaration of Helsinki and approved by the Jawatankuasa Etika Penyelidikan Manusia Universiti Sains Malaysia (JEPeM-USM), on April 2, 2023. (Approval Reference: USM/JEPeM/22110749).

ACKNOWLEDGMENT

The author thanks the Faculty of Electronic Engineering Technology at Universiti Malaysia Perlis for providing the opportunity to explore deep in this research. The author would like to acknowledge the support from the Ministry of Higher Education (MoHE) Malaysia through the Fundamental Research Grant Scheme (FRGS) under a grant number of FRGS/1/2023/ICT02/UNIMAP/02/3. In most cases, sponsor and financial support acknowledgements. Universiti Sains Malaysia, RU Top Down grant 1001/PPSP/8070016.

REFERENCES

- [1] Diffuse Large B-Cell Lymphoma - Lymphoma Research Foundation. (n.d.). <https://lymphoma.org/understanding-lymphoma/aboutlymphoma/nhl/dlbcl/>
- [2] M. A. Lopez, "Diffuse large B-cell lymphoma risk factors," Rare Disease Advisor, <https://www.rarediseaseadvisor.com/hcp-resource/diffuse-large-b-cell-lymphoma-risk-factors/>.
- [3] C. C. medical professional, "Diffuse large B-cell lymphoma," Cleveland Clinic, <https://my.clevelandclinic.org/health/diseases/24405-diffuse-large-b-cell-lymphoma>.
- [4] Larson, D.P., Peterson, J.F., Nowakowski, G.S. et al. (2020). A practical approach to FISH testing for MYC rearrangements and brief review of MYC in aggressive B-cell lymphomas. *J Hematopathol* 13, 127–135, doi: 10.1007/s12308-020-00404-w.
- [5] "Diffuse large B cell lymphoma: Outlook, stages, treatment," Medical News Today, <https://www.medicalnewstoday.com/articles/diffuse-large-b-cell-lymphoma>.
- [6] Liu, Y., & Barta, S. K. (2019). Diffuse large B-cell lymphoma: 2019 update on diagnosis, risk stratification, and treatment. *In American Journal of Hematology* (Vol. 94, Issue 5, pp. 604–616). Wiley-Liss Inc, doi: 10.1002/ajh.25460.
- [7] M. A. Lopez, "Diffuse large B-cell lymphoma diagnosis," Rare Disease Advisor, <https://www.rarediseaseadvisor.com/disease-info-pages/diffuse-large-b-cell-lymphoma-diagnosis/#:~:text=Laboratory%20Testing%20for%20Diagnosing%20DLBCL&text=The%20comprehensive%20metabolic%20panel%20assesses,elevation%20indicates%20the%20tumor%20burden>.

- [8] M. Shipra Gandhi, "Diffuse large B-cell lymphoma (DLBCL) workup," Approach Considerations, Flow Cytometry and Genetic Studies, Imaging Studies, <https://emedicine.medscape.com/article/202969-workup?form=fpf#c1>.
- [9] Magaki, S., Hojat, S. A., Wei, B., So, A., & Yong, W. H. (2019). An introduction to the performance of immunohistochemistry. In *Methods in Molecular Biology* (Vol. 1897, pp. 289–298). Humana Press Inc, doi: 10.1007/978-1-4939-8935-5_25.
- [10] Nguyen, L., Papenhausen, P., & Shao, H. (2017). The Role of c-MYC in B-Cell Lymphomas: Diagnostic and molecular aspects. *Genes*, 8(4), 2–22, doi: 10.3390/genes8040116.
- [11] Hamdi, M., Senan, E. M., Jadhav, M. E., Olayah, F., Awaji, B., & Alalayah, K. M. (2023). Hybrid Models Based on Fusion Features of a CNN and Handcrafted Features for Accurate Histopathological Image Analysis for Diagnosing Malignant Lymphomas. *Diagnostics*, 13(13), doi: 10.3390/diagnostics13132258.
- [12] Vrabac, D., Smit, A., Rojansky, R., Natkunam, Y., Advani, R. H., Ng, A. Y., Fernandez-Pol, S., & Rajpurkar, P. (2021). DLBCL-Morph: Morphological features computed using deep learning for an annotated digital DLBCL image set. *Scientific Data*, 8(1), doi: 10.1038/s41597-021-00915-w.
- [13] Basu, S., Agarwal, R., & Srivastava, V. (2022). Deep discriminative learning model with calibrated attention map for the automated diagnosis of diffuse large B-cell lymphoma. *Biomedical Signal Processing and Control*, 76, 103728, doi: 10.1016/j.bspc.2022.103728.
- [14] Blanc-Durand, P., Jégou, S., Kanoun, S., Berriolo-Riedinger, A., Bodet-Milin, C., Kraeber-Bodéré, F., Carlier, T., le Gouill, S., Casasnovas, R. O., Meignan, M., & Itti, E. (2021). Fully automatic segmentation of diffuse large B cell lymphoma lesions on 3D FDG-PET/CT for total metabolic tumour volume prediction using a convolutional neural network. *European Journal of Nuclear Medicine and Molecular Imaging*, 48(5), 1362–1370, doi: 10.1007/s00259-020-05080-7.
- [15] Ferrández, M. C., Golla, S. S. V., Eertink, J. J., de Vries, B. M., Wieggers, S. E., Zwezerijnen, G. J. C., Pieplensbosch, S., Schilder, L., Heymans, M. W., Zijlstra, J. M., & Boellaard, R. (2023). Sensitivity of an AI method for [18F]FDG PET/CT outcome prediction of diffuse large B-cell lymphoma patients to image reconstruction protocols. *EJNMMI Research*, 13(1), doi: 10.1186/s13550-023-01036-8.
- [16] el Hussein, S., Chen, P., Medeiros, L. J., Hazle, J. D., Wu, J., & Khoury, J. D. (2022). Artificial intelligence-assisted mapping of proliferation centers allows the distinction of accelerated phase from large cell transformation in chronic lymphocytic leukemia. *Modern Pathology*, 35(8), 1121–1125, doi: 10.1038/s41379-022-01015-9.
- [17] Wójcik, P., Naji, H., Simon, A., Büttner, R., & Božek, K. (2023). Learning Nuclei Representations with Masked Image Modelling. <http://arxiv.org/abs/2306.17116>
- [18] Li, D., Bledsoe, J. R., Zeng, Y., Liu, W., Hu, Y., Bi, K., Liang, A., & Li, S. (2020). A deep learning diagnostic platform for diffuse large B-cell lymphoma with high accuracy across multiple hospitals. *Nature Communications*, 11(1), doi: 10.1038/s41467-020-19817-3.
- [19] Graham, S., Vu, Q. D., Raza, S. E. A., Azam, A., Tsang, Y. W., Kwak, J. T., & Rajpoot, N. (2018). HoVer-Net: Simultaneous Segmentation and Classification of Nuclei in Multi-Tissue Histology Images. <http://arxiv.org/abs/1812.06499>
- [20] Swiderska-Chadaj, Z., Hebeda, K. M., van den Brand, M., & Litjens, G. (2021). Artificial intelligence to detect MYC translocation in slides of diffuse large B-cell lymphoma. *Virchows Archiv*, 479(3), 617–621, doi: 10.1007/s00428-020-02931-4.
- [21] Bándi, P., Balkenhol, M., van Ginneken, B., van der Laak, J., & Litjens, G. (2019). Resolution-agnostic tissue segmentation in whole-slide histopathology images with convolutional neural networks. *PeerJ*, 2019(12), doi: 10.7717/peerj.8242.
- [22] "Test Details - Diffuse Large B-Cell Lymphoma (DLBCL) FISH Panel." <https://knightdxlabs.ohsu.edu/home/test-details?id=Diffuse+Large+B-Cell+Lymphoma+%28DLBCL%29+FISH+Panel>
- [23] V. Kasireddy, "Double Hit Lymphomas: Role of Immunohistochemistry in the Era of Florescent in-Situ Hybridization," *Blood*, Dec. 02, 2016, doi: 10.1182/blood.V128.22.5.405.5405.
- [24] Ferrández, Maria C. ; Golla, Sandeep S.V. ; Eertink, Jakoba J. et al. (2023). An artificial intelligence method using FDG PET to predict treatment outcome in diffuse large B cell lymphoma patients. *Scientific Reports*, 13(1), doi: 10.1038/s41598-023-40218-1
- [25] Mohlman, J. S., Leventhal, S. D., Hansen, T., Kohan, J., Pascucci, V., & Salama, M. E. (2020). Improving Augmented Human Intelligence to Distinguish Burkitt Lymphoma from Diffuse Large B-Cell Lymphoma Cases. *American Journal of Clinical Pathology*, 153(6), 743–759, doi: 10.1093/ajcp/aqaa001.
- [26] Farinha, F., & Ioannidis, N. (n.d.). Artifact Removal and FOXP3+ Biomarker Segmentation for Follicular Lymphomas.
- [27] Shankar, V., Yang, X., Krishna, V., Tan, B. T., Rojansky, R., Ng, A. Y., Valvert, F., Briercheck, E. L., Weinstock, D. M., Natkunam, Y., Fernandez-Pol, S., & Rajpurkar, P. (n.d.). LymphoML: An interpretable artificial intelligence-1 based method identifies morphologic features that 2 correlate with lymphoma subtype 3 4, doi: 10.1101/2023.03.14.23287143.
- [28] Jiang, C., Chen, K., Teng, Y., Ding, C., Zhou, Z., Gao, Y., Wu, J., He, J., Kelei He, & Zhang, J. (n.d.). Deep learning-based tumour segmentation and total metabolic tumour volume prediction in the prognosis of diffuse large B-cell lymphoma patients in 3D FDG-PET images, doi: 10.1007/s00330-022-08573-1.
- [29] Swiderska-Chadaj, Z., Hebeda, K., van den Brand, M., & Litjens, G. (2020). Predicting MYC translocation in HE specimens of diffuse large B-cell lymphoma through deep learning. 36, doi: 10.1007/s00428-020-02931-4.
- [30] Steinbuss, G., Kriegsmann, M., Zgorzelski, C., Brobeil, A., Goeppert, B., Dietrich, S., Mechttersheimer, G., & Kriegsmann, K. (2021). Deep learning for the classification of non-hodgkin lymphoma on histopathological images. *Cancers*, 13(10), doi: 10.3390/cancers13102419.
- [31] Perry, C., Greenberg, O., Haberman, S., Herskovitz, N., Gazy, I., Avinoam, A., Paz-Yaacov, N., Hershkovitz, D., & Avivi, I. (2023). Image-Based Deep Learning Detection of High-Grade B-Cell Lymphomas Directly from Hematoxylin and Eosin Images. *Cancers*, 15(21), 5205, doi: 10.3390/cancers15215205.
- [32] Lisson, C. S., Lisson, C. G., Mezger, M. F., Wolf, D., Schmidt, S. A., Thaiss, W. M., Tausch, E., Beer, A. J., Stilgenbauer, S., Beer, M., & Goetz, M. (2022). Deep Neural Networks and Machine Learning Radiomics Modelling for Prediction of Relapse in Mantle Cell Lymphoma. *Cancers*, 14(8), doi: 10.3390/cancers14082008.
- [33] Carreras, J., Kikuti, Y. Y., Miyaoka, M., Hiraiwa, S., Tomita, S., Ikoma, H., Kondo, Y., Ito, A., Nakamura, N., & Hamoudi, R. (2021). A Combination of Multilayer Perceptron, Radial Basis Function Artificial Neural Networks and Machine Learning Image Segmentation for the Dimension Reduction and the Prognosis Assessment of Diffuse Large B-Cell Lymphoma. *AI 2021*, Vol. 2, Pages 106-134, 2(1), 106–134, doi: 10.3390/ai2010008.
- [34] Carreras, J., Roncador, G., & Hamoudi, R. (2022). Artificial Intelligence Predicted Overall Survival and Classified Mature B-Cell Neoplasms Based on Immuno-Oncology and Immune Checkpoint Panels. *Cancers*, 14(21), doi: 10.3390/cancers14215318.
- [35] Wagner, M., Hänsel, R., Reinke, S., Richter, J., Altenbuchinger, M., Braumann, U. D., Spang, R., Löffler, M., & Klapper, W. (2019). Automated macrophage counting in DLBCL tissue samples: A ROF filter based approach. *Biological Procedures Online*, 21(1), doi: 10.1186/s12575-019-0098-9.
- [36] Chen, P., el Hussein, S., Xing, F., Aminu, M., Kannapiran, A., Hazle, J. D., Medeiros, L. J., Wistuba, I. I., Jaffray, D., Khoury, J. D., & Wu, J. (2022). Chronic Lymphocytic Leukemia Progression Diagnosis with Intrinsic Cellular Patterns via Unsupervised Clustering. *Cancers*, 14(10), doi: 10.3390/cancers14102398.
- [37] Carreras, J., Kikuti, Y. Y., Miyaoka, M., Hiraiwa, S., Tomita, S., Ikoma, H., Kondo, Y., Ito, A., Shiraiwa, S., Hamoudi, R., Ando, K., & Nakamura, N. (2020). A Single Gene Expression Set Derived from Artificial Intelligence Predicted the Prognosis of Several Lymphoma Subtypes; and High Immunohistochemical Expression of TNFAIP8 Associated with Poor Prognosis in Diffuse Large B-Cell Lymphoma. *AI (Switzerland)*, 1(3), 342–360, doi: 10.3390/ai1030023.

[38] Bhattamisra, S. K., Banerjee, P., Gupta, P., Mayuren, J., Patra, S., & Candasamy, M. (2023). Artificial Intelligence in Pharmaceutical and Healthcare Research. In Big Data and Cognitive Computing (Vol. 7, Issue 1). MDPI, doi: 10.3390/bdcc7010010.

[39] Achi, H. el, Belousova, T., Chen, L., Wahed, A., Wang, I., Hu, Z., Kanaan, Z., Rios, A., & Nguyen, A. N. D. (2019). Automated Diagnosis of Lymphoma with Digital Pathology Images Using Deep Learning. www.annclinlabsci.org

[40] "Multi-Layer Perceptron Learning in Tensorflow," GeeksforGeeks, Nov. 03, 2021. <https://www.geeksforgeeks.org/multi-layer-perceptron-learning-in-tensorflow/>

[41] Kumar Agarwal, A., Angeline Ranjithamani, D., Velayudham, A., Shunmugam, A., & Ismail, M. B. (2021). Machine Learning Technique for the Assembly-Based Image Classification System.

[42] Mohammed Ismail.B, S.Mahaboob Basha, & B.Eswara Reddy. (2015). Improved fractal image compression using range block size. 2015 IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS) : 2-3 November, 2015, KIIT University, Bhubaneswar, Odisha, India. IEEE.

APPENDIX

TABLE I. SUMMARY OF PRE-PROCESSING AND SEGMENTATION METHODS ON DLBCL

Authors	No of Samples	Pre-Processing Methods	Features Extraction	Data Augmentation	CNN Architecture	Results
Hamdi, M. et al. [11]	15,000 H&E stained whole-slide images.	- Gaussian filter - Laplacian filter - Normalisation - Resize	Yes	Yes	Pre-trained model: - MobileNet-VGG16 - VGG16-AlexNet - MobileNet-AlexNet	MobileNet-VGG16 - AUC: 99.43% - Accuracy: 99.8% - Precision: 99.77% - Sensitivity: 99.7% - Specificity: 99.8%
Vrabac, D. et al. [12]	209 DLBCL cases.	N/A	- Maximum area - Minimum area - Hull area - Perimeter nucleus - Maximum angle - Ellipse perimeter - Ellipse area	N/A	Pre-trained model: - ResNet-50 - HoVerNet	C-index (95% CI) of 0.635 (0.574,0.691)
Basu, S. et al. [13]	1,000 pathologic tissue slides images of DLBCL and non-DLBCL.	N/A	- Attention map feature transformer - Feature fusion	- Image rotation - Horizontal and vertical flip - Zoom scaling - Vertical and horizontal shifts	Pre-trained model: - DenseNet-201 Optimiser: Adam Learning rate: 0.0001	- Accuracy: 98.31 ± 0.5 - Sensitivity: 98.27 ± 0.58 - Specificity: 98.35 ± 0.69
Blanc-Durand, P. et al. [14]	Pre-therapy FDG-PET/CT scans from 733 patients with DLBCL.	- Resampling - Padding - Cropping - Scaling of the PET and CT image data - Adaptive thresholding	- Tumour heterogeneity - Textural features - Total Tumour surfaces - Spatial dispersion	N/A	Pre-trained Model: - 3D U-Net Optimiser: Adam	- Mean DSC: 0.73 ± 0.20 (Median: 0.79) - Jaccard coefficients: 0.68 ± 0.21
Ferrández, M. C. et al. [15]	20 DLBCL patients on a dataset of 296 maximum intensity projection (MIP) images.	- Gaussian filter	- Metabolic tumour volume (MTV) - Standard uptake value (SUV) - Dissemination - Textural features	N/A	Pre-trained Model: - 3D U-Net Optimiser: Adam Epochs: 200 Learning rate: 0.00005 Decay rate: 0.000001	- Training: 0.81 (0.02) - Validation: 0.75 (0.07)

Authors	No of Samples	Pre-Processing Methods	Features Extraction	Data Augmentation	CNN Architecture	Results
el Hussein, S. et al. [16]	10 CLL, 12 aCLL, and 8 RT digitally stained H&E slides from a lymph node excisional biopsy.	N/A	- ROI annotation - Ratio of segmented nuclear contour area to its convex - Hull area	N/A	Pre-trained Model: - HoVerNet	- Accuracy: 0.658 (± 0.115)
Wójcik, P. et al. [17]	37,665 H&E stained DLBCL images of size 448×448, divided into 28×28 square patches.	N/A	- Cell Patch Embedding - Patch Aggregation	- Random resize and crop - Colour jittering - Random flip	Pre-trained Model: - HoVerNet Epochs: 800	- F1 Score: 0.939 for Epithelial cells.
Li, D. et al. [18]	Hospital A: 500 DLBCL & 505 non-DLBCL human samples Hospital B: 163 DLBCL & 184 non-DLBCL human samples Hospital C: 204 DLBCL & 198 non-DLBCL human samples	N/A	-Types of lymphomas and hematopoietic tumours - Colour - Morphology - Quality	N/A	- Deep Neural Network Classifiers and pathologists were compared.	- Recall: 100% - Precision: 96% - F1 score: 98%
Graham, S. et al. [19]	24,319 annotated nuclei within 41 colorectal adenocarcinoma image tiles.	N/A	- Nuclear pixel branch - Hover branch - Nuclear classification branch	- Flip - Rotation - Gaussian blur - Median blur	Pre-trained Model: - HoVerNet Optimiser: Adam Epochs: 50 Learning rate: 10^{-4}	- DICE score: 0.869
Swiderska-Chadaj, Z. et al. [20]	H&E-stained slides of 287 DLBCL cases from 11 hospitals.	N/A	N/A	N/A	Pre-trained Model: - U-Net	- AUC: 0.83 (External) - Sensitivity: 0.95 (External) - Specificity: 0.53 (Internal)
Bándi, P. et al. [21]	100 whole-slide images from 10 different tissues.	- Otsu thresholding	- Pixel intensity - Colour - Textural features	- Horizontal mirroring - 90° rotation - Scaling - Colour adjustment - Contrast adjustment - Additive Gaussian noise - Gaussian blur	- FCNN Optimiser: Adam Epochs: 16 Learning rate: 10^{-4} Activation: ReLU	- Dice scores: 0.9775 to 0.9891
Ferrández, M. C. et al. [24]	373 DLBCL patients	- Normalisation - Filtering	N/A	N/A	Pre-trained Model: - 3D U-Net Optimiser: Adam Epochs: 200	- AUC: 0.72 - Sensitivity: 0.59 - Specificity: 0.8

Authors	No of Samples	Pre-Processing Methods	Features Extraction	Data Augmentation	CNN Architecture	Results
					Learning rate: 0.00005 Decay rate: 0.000001 Activation: ReLU	
Mohlman, J. S. et al. [25]	10,818 images from Burkitt Lymphoma (BL) and DLBCL.	- Normalisation - Edge detection	- Notion of deep network pixel level	- Random horizontal flipping of images - Random alteration of contrast	Epochs: 200 Learning rate: 6.5×10^{-5}	Accuracy: 94%
Farinha, F. et al. [26]	2886×2886 high resolution images and patched into 36 patches of equal size (481×481)	N/A	N/A	N/A	Pre-trained Model: - U-Net Optimiser: Adam Epochs: 150 Learning rate: 0.0001 Activation: ReLU	- Linear regression, R2: 0.4688
Shankar, V. et al. [27]	670 lymphoma cases	- Normalisation - Patch-based quality control (PQC) threshold	- Minimum / maximum Feret diameters - Convex hull area - Circulatory - Elongation - Convexity	N/A	Pre-trained Model: - StarDist	Diagnostic accuracy: 64.3%
Jiang, C. et al. [28]	414 DLBCL patients collected from two independent centres in 3D FDG-PET images.	- Threshold - Normalisation	- Convolution number	N/A	Pre-trained Model: - 3D U-Net Epochs: 1000 Learning rate: 0.01 Nesterov momentum: 0.99	- PFS: 64.5% - OS: 73.4%
Swiderska-Chadaj, Z. et al. [29]	91 patients with H&E-stained specimens	N/A	N/A	- Brightness - Contrast - Saturation - Rotation - Gaussian noise - Gaussian blur	Pre-trained Model: - U-Net Optimiser: Adam Epochs: 500 Learning rate: 0.0005	- AUC: 0.77 - Sensitivity: 0.88 - Specificity: 0.66
Steinbuss, G. et al. [30]	84,139 image patches from 629 patients	- Patch-based quality control (PQC) threshold	N/A	N/A	Pre-trained Model: - Efficient-Net	- High accuracy above 95% - Lower BACC with multiple misclassification

Authors	No of Samples	Pre-Processing Methods	Features Extraction	Data Augmentation	CNN Architecture	Results
					Optimiser: Adam Epochs: 50 Learning rate: 10^{-5} to 10^{-6}	- Overall BACC up to 95.56%
Perry, C. et al. [31]	32 biopsies from 30 patients	N/A	N/A	- Colour Jittering - Channel shuffle	- Multiple Instance Learning (MIL) Optimiser: Adam Epochs: 20 Learning rate: 0.0001	- AUC: 0.95 - Sensitivity: 87% - Specificity: 100%
Lisson, C. S. et al. [32]	30 patients with histologically proven mantle cell lymphoma who underwent contrast-enhanced CT or PET/CT scans	- Filtering-based feature selection	- 3D volumetric radiomic features	- Random Flip - Gaussian Blur - Gaussian Noise	Pre-trained Model: - 3D SE ResNet - 3D DenseNet Optimiser: Adam Epochs: 100 Learning rate: 0.001	- Overall accuracy of predicting relapse: 64%

TABLE II. SUMMARY OF SEGMENTATION METHODS BY MACHINE LEARNING

Authors	No. of Samples	Features Extraction	Segmentation Techniques	Machine Learning	Machine Learning Library	Results
Carreras, J. et al. [33]	414 cases of DLBCL.	- Mann-Whitney U test - Kaplan-Meier - Multivariate Cox Regression - Hazard ratios / risks	- Trainable Weka Segmentation Method	- Multilayer Perceptron (MLP) - Radial Basis Function (RBF)	- XGBoost	- MLP was more "efficient" than RBF.
Carreras, J. et al. [34]	100 to 293 cases from the lymphoma series of Tokai University Hospital.	- Pearson Chi-Square - Fisher's exact tests - Nonparametric Mann-Whitney U test - Kruskal-Wallis H test - Kaplan-Meier - Log-rank tests - Univariate and multivariate Cox Regression	- Weka Method	- Multilayer Perceptron (MLP) - Radial Basis Function (RBF)	- XGBoost	- Overall accuracy: 100%
Wagner, M. et al. [35]	50 test images for whole tissue samples of DLBCL.	- Grayscale conversion	- Rudin-Osher-Fatemi (ROF) filtering	- Mask R-CNN	N/A	- Manual count: 0.9297
Chen, P. et al. [36]	193 biopsy specimens from 135 patients.	- Solidity feature	- ROI annotation	N/A	- XGBoost	- Accuracy: 0.925 - AUC: 0.978

Authors	No. of Samples	Features Extraction	Segmentation Techniques	Machine Learning	Machine Learning Library	Results
Carreras, J. et al. [37]	100 cases from Western countries diagnosed from nodal DLBCL.	<ul style="list-style-type: none">- Gaussian blur- Hessian- Membrane projections- Sobel filter- Difference of Gaussians	N/A	<ul style="list-style-type: none">- Multilayer Perceptron (MLP)	N/A	<ul style="list-style-type: none">- Successful AI approach in DLBCL
Bhattamisra, S. K. et al. [38]	20,863 genes as the input layer and lymphoma subtypes as the output layer.	N/A	N/A	<ul style="list-style-type: none">Multilayer Perceptron (MLP)	N/A	<ul style="list-style-type: none">- 58 genes predicted survival with high accuracy.- 10 genes were associated with poor survival and 5 genes with favourable survival.
Achi, H. el et al. [39]	Digital WSIs of H&E-stained slides of 128 cases with a total of 2,560 images	<ul style="list-style-type: none">- Data augmentation (Random cropping, image rotation, image inversion)- Max-pooling layers	N/A	N/A	<ul style="list-style-type: none">- Support Vector Machine- Neural Network	<ul style="list-style-type: none">- Overall accuracy: 95%

User Interface Design of Digital Test Based on Backward Chaining as a Measuring Tool for Students' Critical Thinking

I Putu Wisna Ariawan^{1*}, P. Wayan Arta Suyasa², Agus Adiarta³,

I Komang Gede Sukawijana⁴, Nyoman Santiyadnya⁵, Dewa Gede Hendra Divayana⁶

Department of Mathematics Education, Universitas Pendidikan Ganesha, Singaraja, Bali, Indonesia¹

Department of Informatics Education, Universitas Pendidikan Ganesha, Singaraja, Bali, Indonesia^{2,6}

Department of Electrical Education, Universitas Pendidikan Ganesha, Singaraja, Bali, Indonesia^{3,4,5}

Abstract—Assessing students' critical thinking skills is challenging due to the limitations of current measurement tools. Therefore, there is a need for a digital testing instrument that can effectively evaluate students' critical thinking abilities. The proposed digital test should be designed to present questions in a tiered manner, using a backward chaining approach that starts with general questions and progresses to more detailed ones. However, developing this measurement instrument requires careful planning. One of the initial steps in this process is to create a user interface design. The purpose of this study was to show the quality of the design of the user interface of a digital test based on backward chaining as a measuring tool for students' critical thinking in a differentiated learning atmosphere. Design development used the Borg and Gall model and only focused on three stages. These stages include design planning, initial testing, and revision for the initial testing results. Data collection was through initial testing of the design. The tool used to collect data was a questionnaire. Respondents involved in the initial testing were 34 people. The location for the study was at several IT vocational high schools spread across six regencies in Bali. The data analysis technique compared the percentage comparison of the quality of the user interface design with the quality standards of the user interface design and referred to a five scale. The results of the study showed that the design quality of the digital test user interface based on backward chaining was included in the good category, as indicated by a quality percentage of 88.94%. Specifically, the impact of the results on the field of educational evaluation is to make it easier for evaluators to make accurate measurements. In general, the effect of this study on the field of informatics engineering education is the existence of innovations in realizing a test to measure critical thinking in the domain of differentiated learning.

Keywords—User interface design; digital test; backward chaining; critical thinking; differentiated learning

I. INTRODUCTION

The rolling of the “*Merdeka Belajar*” (independent learning) policy provides students with the freedom to follow the learning process according to the differences in their needs and learning environment. This concept is called differentiated learning [1]. Currently, differentiated learning is a trend in the learning process at the IT Vocational School level.

The existence of differences in the way each student learns according to their characteristics and needs through differentiated learning certainly can encourage them to improve their critical thinking skills in dealing with the problems they face [2]. Differentiated learning can awaken students' critical thinking skills, but teachers find obstacles in its implementation. These obstacles are mainly related to measuring students' critical thinking skills. Based on these obstacles, it is necessary to find a breakthrough as a measuring tool in the form of a digital test that can easily measure students' critical abilities. The expected digital test can package sequentially backward test questions (backward chaining) from general question types to questions with more detailed or specific types so that later, the teacher can explore the student's critical thinking to solve the questions presented. The research question referring to the obstacles and breakthroughs initiated is “What is the form of the user interface design of a backward chaining-based digital test used to measure students' critical thinking in a differentiated learning atmosphere?”

The specific objective of this study is to show a backward chaining-based digital test that has good quality and accurately measures students' critical thinking skills in differentiated learning. The urgency of this study is to obtain a user interface design from a backward chaining-based digital test that effectively assesses students' critical thinking skills in differentiated learning, especially in Mathematics subjects at IT Vocational School in Bali.

Hizqiyah et al. conducted research on the development of digital problem-solving skills test instruments [3]. However, a gap in their research is that they did not demonstrate the test items graded sequentially from general types to more specific types. Ndibalema's research [4], on the other hand, focuses on a form of formative assessment conducted online. The key difference between Ndibalema's study and this current research lies in the type of evaluation used; Ndibalema's work leans towards formative assessment, while this research encompasses both formative and summative assessments. Additionally, Jaskova's research explores student satisfaction with online tests taken at home [5]. A limitation identified in Jaskova's study is the lack of information regarding the user interface design of the online tests administered at home. Noor's research highlights the use of Kahoot as a digital quiz tool [6]. However, a limitation of

*Corresponding author

this study is that it does not provide details about the design of the Kahoot user interface as a digital quiz. On the other hand, the research conducted by Domínguez-Figaredo & Gil-Jaurena examines the impact of familiarity on digital assessments in online education [7]. A limitation of their study is that it fails to present the specific format of the digital tests utilized in the assessment.

Based on the research question and specific objectives of this study, it is essential to determine the form and quality of the backward chaining-based digital test used to assess students' critical thinking skills in differentiated learning.

II. LITERATURE REVIEW

Some of the research behind this study includes a 2020 study by Ariawan, Giri, and Divayana on the development of a CIPP evaluation application based on Simple Additive Weighting [8] obtained visualization results from a CIPP evaluation application based on Simple Additive Weighting that can measure the effectiveness of learning at health science colleges in Bali online. The obstacle is that the application is not for a large-scale implementation. A 2021 study on the dissemination and implementation of a CIPP evaluation application based on Simple Additive Weighting at several health science colleges in Bali conducted by Divayana, Ariawan, and Giri [9] showed the success of implementing a CIPP evaluation application based on Simple Additive Weighting at several health science colleges in Bali. The obstacle is that the evaluation application does not yet use the integrated evaluation aspects of the Balinese local wisdom concept, so the measurement of students' knowledge domains in the learning process cannot be measured optimally and in depth according to student characteristics. A 2022 study by Ariawan et al. showed the development of a Formative-Summative evaluation model based on Tri Pramana by inserting Weighted Product calculations [10]. The general description of the results obtained in the 2022 study is a Formative-Summative evaluation model design based on Tri Pramana by inserting Weighted Product calculations so it can determine the aspects that determine the quality of e-learning implementation. In the 2023 study conducted by Ariawan et al., a Tri Pramana-Weighted Product-Based Formative-Summative Model Evaluation Application has been obtained and has been field tested [11]. Further research for 2024 is the realization of a digital test user interface design based on backward chaining as a measuring tool for students' critical thinking in the nuances of learning differentiation.

Divayana et al. research [12] showed test instrument items to measure students' cognitive abilities in implementing distance learning. The validity of the content and reliability of the instrument is good. However, there is no packaging of test questions sequentially graded backward from general question types to more detailed or specific types. Easa and Blonder's research [13] showed an instrument to measure or evaluate teacher and student beliefs about differentiated learning in Chemistry. The constraint of Easa and Blonder's research was that it did not show an evaluation instrument sequentially graded backward from general to specific questions. Kholid et al.'s research [14] showed the implementation of diagnostic assessments in differentiated learning modules for English subjects. The constraint of Kholid et al.'s research is that it has

not shown the form of a diagnostic assessment instrument for differentiated learning sequentially from complex things to more specific things.

III. METHOD

Several elements are presented in the methodology section of this research: 1) research approach; 2) subjects, objects, and research locations; 3) data collection instruments; and 4) data analysis techniques.

A. Research Approach

The research approach is development. The focus of development for this 2024 research was on three stages, including design development, initial trials, and revisions to the results of the initial trials (main product revision). The model used in the development process is Borg and Gall [15],[16],[17],[18]. The three stages of development referred to the researcher's desire/goal to realize a digital test user interface design based on backward chaining as a measuring tool for students' critical thinking in a differentiated learning atmosphere. The research stages carried out by the researcher can be seen in Fig. 1.

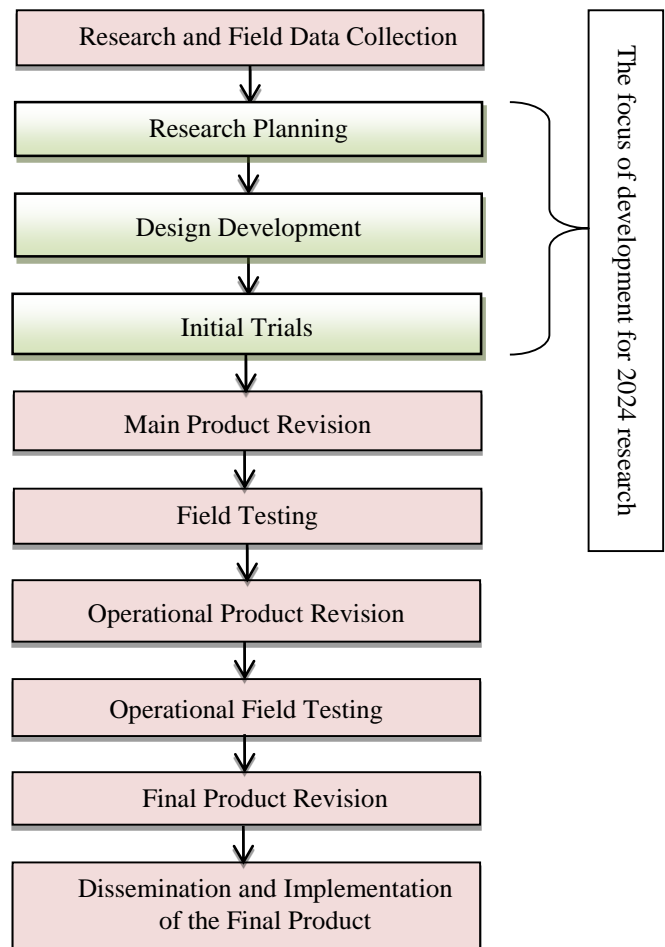


Fig. 1. The research stages that refer to the borg and gall design.

B. Subject, Object, and Location of Research

Subjects involved in the initial trial phase of the digital test user interface design based on backward chaining, including

education evaluation experts, informatics experts, and several teachers at IT Vocational School in Bali. The number of informatics education experts was two experts, the number of education evaluation experts was two experts, and the number of IT Vocational School teachers in Bali involved was 30 teachers. The selection of research subjects utilized a purposive sampling technique, involving individuals who possess in-depth knowledge and clear objectives regarding the object of study. The object of this research is the design of the digital test user interface based on backward chaining as a measuring tool for students' critical thinking in a differentiated learning atmosphere. The research location was at several IT Vocational School spread across six agencies in Bali.

C. Data Collection Instruments

The data collection tool used in this study is a questionnaire. All questions used in the questionnaire are related to the digital test user interface design based on backward chaining as a measuring tool for students' critical thinking in a differentiated learning atmosphere. The number of questions in the questionnaire was ten items. The details of the ten questions are explained in the discussion section of this paper. These ten questions are valid and reliable based on the instrument trials conducted by two education experts and two informatics experts.

D. Data Analysis Techniques

After being collected, quantitative data examination was a descriptive approach and descriptive percentage calculation. The technique for analyzing the initial trial data in this study was quantitative descriptive. It was to compare the percentage of the level of quality of the digital test user interface design based on backward chaining with the standard of user interface design quality that refers to a scale of five. The formula used to determine the percentage of the quality level of the digital test user interface design based on backward chaining is in equation (1) [19],[20],[21], then the quality standard that refers to a scale of five can be seen in Table I [22],[23],[24].

$$P = (f/N) \times 100\% \tag{1}$$

Notes:

f = Total acquisition value

N = maximum total value

TABLE I. QUALITY STANDARDS OF USER INTERFACE DESIGN OF DIGITAL TEST BASED ON BACKWARD CHAINING REFERRING TO FIVE SCALE CATEGORY

Percentage of Quality	Quality Category	Recommendations
90-100 %	Excellence	No Revision Required
80-89 %	Good	No Revision Required
65-79 %	Moderate	Revision
55-64 %	Less	Revision
0-54 %	Poor	Revision

IV. RESULTS AND DISCUSSION

A. Results

The results obtained on three stages of development focused on this research, including results at the design planning stage, the initial trial stage, and the initial trial revision stage. The data obtained based on the results at several stages of development in question are as follows.

1) Design Development: The design development stage produced a digital test user interface design based on backward chaining to measure students' critical thinking in a differentiated learning atmosphere. It was using the Balsamiq Mockups application. The form of the design intended is in Fig. 2.

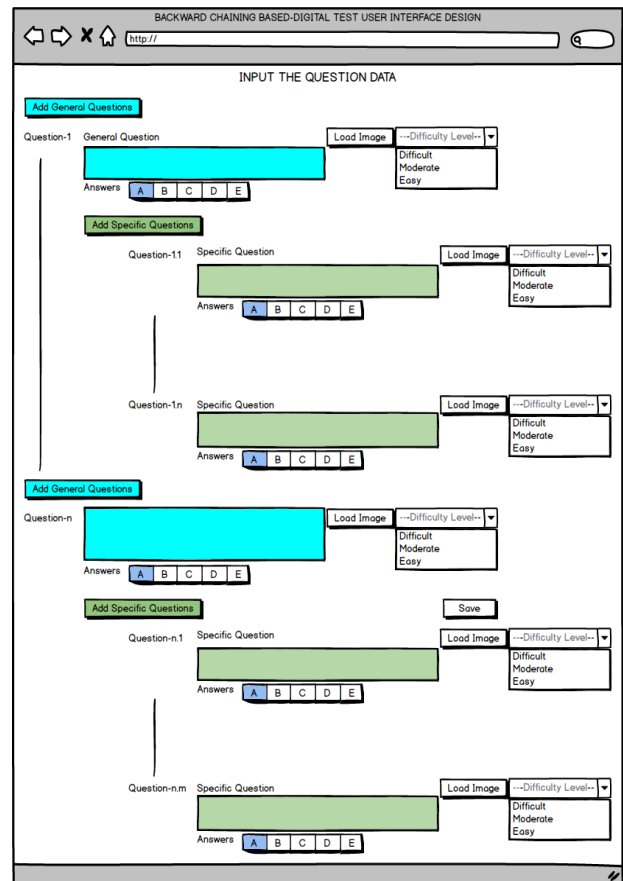


Fig. 2. The user interface design to enter question data.

Fig. 2 shows the user interface design of the form that functions to enter question data. There are several attributes in the form. The "add general questions" button to add general questions. The "add specific questions" button is the specific question from the available general questions. There is a "Load image" button to enter questions containing images. There is a "difficult level" combo box to select the level of difficulty question. Several "answers" buttons are answer choices for the available questions.

No_Rules	General_Questions	Specific_Questions	Action
R.001	Question-1	Question-1.1	Backward Chaining
R.002	Question-1	Question-1.2	Backward Chaining
R.003	Question-1	Question-1.3	Backward Chaining
R.004	Question-1	Question-1.4	Backward Chaining
R.005	Question-1	Question-1.5	Backward Chaining

Fig. 3. The user interface design for questions arrangement based on backward chaining.

Fig. 3 shows the user interface design of the form that manages questions by referring to the backward chaining method. There are several features available on the form. The function of the “no. rules” textbox is to enter the rules number. This number is unique so that no one can duplicate it. There is a “general questions” combo box to select general questions. There is a “specific questions” combo box to specific questions. There is an “arrangement process” combo box to the question arrangement process referring to normal conditions or conditions that refer to the backward chaining concept. The “process” button to run the arrangement process. The “save” button to save the arranged question data. There is a data storage database. The data storage consists of several fields, including No_rules, General_Questions, Specific_Questions, and Action.

features on this form. A text box to enter students’ names and study programs. The combo box for selecting the test type. There is a text area used to display questions. An “answers” button that is useful as a choice of answers to the available questions. There is a “next” button to go to the next question.

Fig. 4. The user interface design for the question-answering facility is based on the test type and employs a backward chaining approach for packaging.

Fig. 4 shows the user interface design of the form that functions as a place to answer questions. There are several

Fig. 5. The user interface design to display final score.

Fig. 5 shows the user interface design of the form to display the final score. This design shows a text area that functions to display questions. An “answers” button that is useful as an answer choice for the available questions. The “finish” button is to end the process of answering questions. The “score” button is to calculate the final score. The “save” button to save the final score.

2) *Initial Trials*: Four experts and 30 teachers of IT vocational schools in Bali conducted an initial trial of the digital test user interface design based on backward chaining. The questionnaire for the initial trial consisted of 10 questions. The results of the initial trial are in Table II.

TABLE II. RESULT OF INITIAL TRIALS TO DIGITAL TEST USER INTERFACE DESIGN BASED ON BACKWARD CHAINING AS A MEASURING TOOL FOR STUDENTS' CRITICAL THINKING IN DIFFERENTIATED LEARNING NUANCES

Experts	Items-										Σ	Percentage of Quality (%)
	1	2	3	4	5	6	7	8	9	10		
Expert-1	5	4	4	5	4	5	5	4	5	5	46	92
Expert-2	5	5	5	4	5	4	4	4	4	5	45	90
Expert-3	5	4	5	5	4	4	5	4	4	5	45	90
Expert-4	4	5	5	5	5	5	4	5	5	4	47	94
Teacher-1	5	4	4	5	4	5	4	5	4	4	44	88
Teacher-2	5	5	4	4	5	5	4	5	4	5	46	92
Teacher-3	5	4	4	5	4	4	5	5	4	5	45	90
Teacher-4	5	4	5	5	5	4	5	4	5	4	46	92
Teacher-5	4	4	5	5	4	4	4	5	5	4	44	88
Teacher-6	5	4	4	5	4	5	4	5	4	4	44	88
Teacher-7	5	5	4	4	4	4	5	4	5	4	44	88
Teacher-8	5	4	4	5	5	4	4	5	5	4	45	90
Teacher-9	5	4	5	4	4	4	5	4	4	5	44	88
Teacher-10	4	4	5	4	4	5	5	5	4	5	45	90
Teacher-11	5	4	4	5	4	5	5	4	4	4	44	88
Teacher-12	4	4	5	4	4	4	5	4	5	4	43	86
Teacher-13	4	5	5	5	5	4	4	4	5	5	46	92
Teacher-14	4	5	5	4	4	4	5	4	4	5	44	88
Teacher-15	4	4	5	4	4	5	4	4	4	5	43	86
Teacher-16	5	4	4	4	5	4	4	4	4	5	43	86
Teacher-17	4	4	5	4	4	5	4	5	4	4	43	86
Teacher-18	4	5	4	4	5	4	5	4	4	4	43	86
Teacher-19	4	4	5	4	5	4	4	5	4	5	44	88
Teacher-20	5	4	4	5	5	4	5	5	4	5	46	92
Teacher-21	4	4	5	4	4	5	4	4	5	5	44	88
Teacher-22	4	5	5	5	4	5	4	5	5	4	46	92
Teacher-23	4	5	5	4	4	4	4	5	4	5	44	88
Teacher-24	4	4	5	4	5	4	5	5	4	5	45	90
Teacher-25	5	4	4	5	5	4	5	4	5	5	46	92
Teacher-26	4	4	5	4	4	5	4	5	5	5	45	90
Teacher-27	4	5	5	5	4	5	4	5	5	4	46	92
Teacher-28	4	5	5	4	4	4	4	5	4	4	43	86
Teacher-29	4	4	5	4	5	4	5	5	4	4	44	88
Teacher-30	4	4	4	4	4	4	4	4	4	4	40	80
Average												88.94

Respondents/assessors provided several suggestions in the initial trial stage. Improvements to the digital test user interface design based on backward chaining used several of these suggestions. Some of them are in Table III.

TABLE III. RESPONDENTS' SUGGESTIONS IN THE INITIAL TRIAL

No	Experts	Suggestions
1	Expert-1	Add the test date to the answer sheet form.
2	Expert-2	Add a list of test participants' scores.
3	Expert-3	There needs to be a facility to display a recapitulation of test results.
4	Expert-4	There needs to be a test date.
5	Teacher-12	There needs to be a facility to view a recapitulation of test results.
6	Teacher-16	There needs to be a test completion time duration.
7	Teacher-18	There needs to be a facility to view the scores of each test participant.
8	Teacher-28	It is better to prepare a facility to display the test implementation date.
9	Teacher-30	There needs to be a facility to display the test completion time duration.

3) *Revision of Initial Trial Results:* Revision of the user interface design of the digital test based on backward chaining based on several respondents' suggestions in the initial trial. It is necessary to make revisions, especially those related to the test implementation date referring to the suggestions of expert-1, expert-4, and teacher-28. They were creating a user interface design to display the test implementation date. The improved design form is in Fig. 6.

It is necessary to make revisions, especially those related to the recapitulation of test results referring to the suggestions of expert-2, expert-3, teacher-12, and teacher-18. They were creating a user interface design to display the recapitulation of test results. The form of the improved design is in Fig. 7.

It is necessary to make revisions, especially those related to the display test completion time referring to the suggestions of expert-16 and teacher-30. They were creating a user interface design to display the display test completion time. The form of the improved design is in Fig. 8.

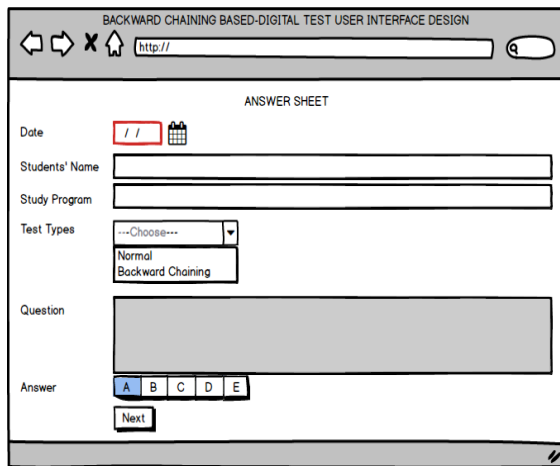


Fig. 6. User interface design to display test implementation date.

Fig. 6 shows the user interface design display for the test execution date. Fig. 6 shows the improvement. There is a date time picker “date” to show the test execution date.

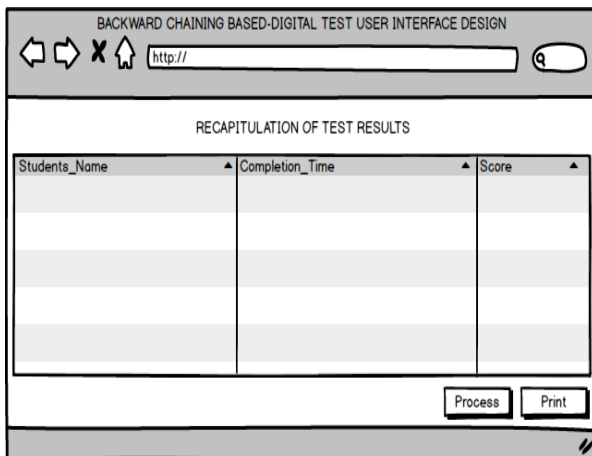


Fig. 7. User interface design to display test result recapitulation.

Fig. 7 shows the user interface design display for the recapitulation of test results. Fig. 7 shows that improvement. There is a database that shows the recapitulation of test results. They are Students_Name, Completion_Time, and Score.

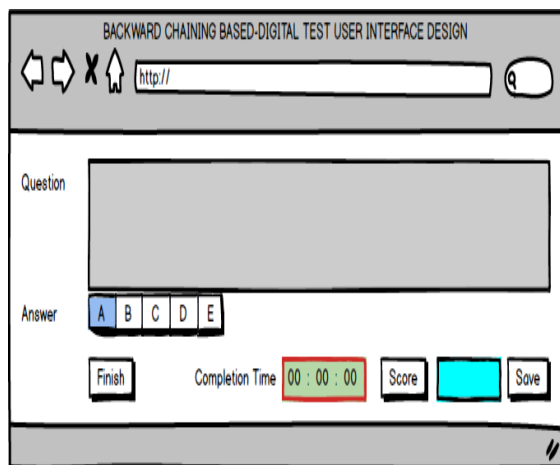


Fig. 8. User interface design to display test completion time.

Fig. 8 shows the user interface design display for the test completion time. Fig. 8 shows the improvement. There is a timer that shows the completion time.

B. Discussion

Referring to the percentage of quality shown in Table II, the digital test user interface design based on backward chaining is good quality. It is because of 88.94%, if checked through the quality standards shown in Table I, then it is true that the quality of the user interface design is good. The reference in providing assessments by respondents in the initial trial, resulting in the data shown in Table II, is in the form of ten questions.

Item-1 is about the suitability of the user interface design form for inputting question data. Item-2 is about the suitability of the general questions form. Item-3 is about the suitability of the specific questions form. Item-4 is about the sequence of relationships between general questions and specific questions. Item-5 is about the ease of creating questions containing image elements. Item-6 is about the ease of setting the level of difficulty of the questions. Item-7 is about the suitability of the user interface design form for backward chaining-based questions arrangement. Item-8 is about the suitability of the backward chaining concept in arranging questions based on the sequence of relationships between general and specific questions. Item-9 is about the suitability of the user interface design form for question answering facilities. Item-10 is about the suitability of the user interface design form for displaying the final score.

This study answers several constraints in the research of Ariawan, Giri, and Divayana [8], the research of Divayana, Ariawan, and Giri [9], the research of Divayana et al. [12], the research of Easa and Blonder [13], and the research of Kholid et al. [14]. The results of this research have been able to show well the design of the user interface of a digital test based on backward chaining packages of the test questions in a sequential manner backward from the general type of questions to questions with more detailed or specific types. In principle, the results of this research also have similarities with several studies of Putra et al. [25], research of Samrgandi [26], research of Darmawan et al. [27] by showing the existence of a user interface design for a measurement/test application.

The novelty of this research is the concept application of backward chaining in artificial intelligence to the preparation of digital test questions in educational evaluation. Based on the internalization of artificial intelligence into educational evaluation, the test formed is a measuring tool for students critical thinking in a differentiated learning atmosphere. However, this research also has constraints. The constraint of this research is that it has not formed a physical application for direct application in the field. It is only limited to the user interface design.

V. CONCLUSION

In general, the findings of this study effectively demonstrate the quality of the user interface design for a digital test based on backward chaining, which serves as a tool for measuring students' critical thinking in a differentiated learning context. A key innovation/novelty of this study is the arrangement of test questions using the artificial intelligence method known as

backward chaining. This approach organizes the questions systematically, progressing from general to specific types, thereby facilitating a deeper exploration of students' critical thinking abilities. Future work that needs to overcome the obstacles of this study is to create a physical application in the form of a backward chaining-based digital test that is ready for field testing. The impact of the results of this study on educational evaluation science is to make it easier for evaluators to conduct tests to measure students' critical thinking. The impact of research results on informatics engineering education, in general, is to show innovations in digital-based test development to determine the critical thinking in differentiated learning.

ACKNOWLEDGMENT

The authors would like to thank the Chair of the Research and Community Service Institute of Universitas Pendidikan Ganesha that providing opportunities and funding in carrying out this research on time based on research grant number: 893/UN48.16/LT/2024. Besides that, the authors express their gratitude to the Rector of Universitas Pendidikan Ganesha who gave a chance to the authors to complete this research.

REFERENCES

- [1] R. Febriana, S. Sugiman, and A. Wijaya, "Analysis of the implementation of differentiated learning in the implementation of the independent curriculum in middle school mathematics lessons," *International Journal of Humanities Education and Social Sciences*, vol. 3, no. 2, pp. 640-650, 2023.
- [2] N. Hidayat, Y. Ruhiat, N. Anriani, and S. Suryadi, "The impact of differentiated learning, adversity intelligence, and peer tutoring on student learning outcomes," *IJORE: International Journal of Recent Educational Research*, vol. 5, no. 3, pp. 537-548, 2024.
- [3] I. Y. N. Hizqiyah, A. Widodo, S. Sriyati, and A. Ahmad, "Development of a digital problem solving skills test instrument: Model rasch analysis," *Jurnal Penelitian Pendidikan IPA*, vol. 9, no. 4, pp. 1658-1663, 2023.
- [4] P. Ndibalema, "Online assessment in the era of digital natives in higher education institutions," *International Journal of Technology in Education (IJTE)*, vol. 4, no. 3, pp. 443-463, 2021.
- [5] J. Jaskova, "Digital testing during the pandemic crisis: university students' opinions on computer-based tests," *International Journal for Innovation Education and Research*, vol. 9, no.1, pp. 36-53, 2021.
- [6] P. Noor, "Kahoot! as a digital quiz in learning english: Graduate students' perspectives," *Journal of English Teaching and Research*, vol. 8, no. 2, pp. 124-132, 2023.
- [7] D. Domínguez-Figaredo, and I. Gil-Jaurena, "Effects of familiarity with digital assessment in online education," *Distance Education*, pp. 1-16, 2024.
- [8] I. P. W. Ariawan, M. K. W. Giri, and D. G. H. Divayana, "Simulation of SAW-based CIPP evaluation model calculation in determining improvement priority for e-learning services," In *4th International Conference on Vocational Education and Training (ICOVET)*, pp. 24-29, 2020.
- [9] D. G. H. Divayana, I. P. W. Ariawan, and M. K. W. Giri, "CIPP-SAW application as an evaluation tool of e-learning effectiveness," *International Journal of Modern Education and Computer Science (IJMECS)*, vol. 13, no. 6, pp. 42-59, 2021.
- [10] I. P. W. Ariawan, W. Sugandini, I. M. Ardana, I. M. S. D. Arta, and D. G. H. Divayana, "Design of formative-summative evaluation model based on tri pramana-weighted product," *Emerging Science Journal*, vol. 6, no. 6, pp. 1477-1491, 2022.
- [11] I. P. W. Ariawan, W. Sugandini, I. M. Ardana, G. A. D. Sugiharni, A. W. O. Gama, and D. G. H. Divayana, "Forms and field trials of a digital evaluation tool: integrating F-S model, WP method, and balinese local wisdom for effective e-learning," *Journal of Applied Data Sciences*, vol. 5, no. 2, pp. 441-454, 2024.
- [12] D. G. H. Divayana, I. G. Sudirtha, and I. K. Suartama, "Digital test instruments based on wondershare-superitem for supporting distance learning implementation of assessment course," *International Journal of Instruction*, vol. 14, no. 4, pp. 945-964, 2021.
- [13] E. Easa, and R. Blonder, "The development of an instrument for measuring teachers' and students' beliefs about differentiated instruction and teaching in heterogeneous chemistry classrooms," *Chemistry Teacher International*, vol. 5, no. 2, pp. 125-141, 2023.
- [14] B. Kholid, A. Rahman, and L. A. Irawan, "Implementing diagnostic assessment in designing differentiated learning for english language learning at the junior high schools," *Journal of Language and Literature Studies*, vol. 4, no. 2, pp. 445-458, 2024.
- [15] E. Faridah, I. Kasih, S. Nugroho, and T. Aji, "The effectiveness of blended learning model on rhythmic activity courses based on complementary work patterns," *International Journal of Education in Mathematics, Science and Technology*, vol. 10, no. 4, pp. 918-934, 2022.
- [16] K. Rusmulyani, I. M. Yudana, I. N. Natajaya, and D. G. H. Divayana, "E-Evaluation based on CSE-UCLA model refers to glickman pattern for evaluating the leadership training program," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, pp. 279-294, 2022.
- [17] D. G. H. Divayana, "Development of duck diseases expert system with applying alliance method at bali provincial livestock office" *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 5, no. 8, 2014.
- [18] N. M. Ratminingsih, L. P. P. Mahadewi, and D. G. H. Divayana, "ICT-Based Interactive Game in TEYL: Teachers' Perception, Students' Motivation, and Achievement," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 13, no. 9, pp. 190-203, 2018.
- [19] G. A. D. Sugiharni, "The development of interactive instructional media oriented to creative problem solving model on function graphic subject," *Journal of Educational Research and Evaluation*, vol. 2, no. 4, pp. 183-189, 2018.
- [20] A. Adiarta, I. M. Sugiarta, K. K. Heryanda, I. K. G. Sukawijana and Dewa Gede Hendra Divayana, "User interface design of sevima edlink platform for facilitating tri kaya parisudha-based asynchronous learning," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 15, no. 12, pp. 795-804, 2024.
- [21] D. G. H. Divayana, "Development of ANEKA-Weighted Product evaluation model based on Tri Kaya Parisudha in computer learning on vocational school," *Cogent Engineering*, vol. 5, no. 1, pp. 1-33, 2018.
- [22] B. Suratno, and J. Shafira, "Development of user interface/user experience using design thinking approach for GMS service company," *Journal of Information Systems and Informatics*, vol. 4, no. 2, pp. 469-494, 2022.
- [23] D. G. H. Divayana, A. Adiarta, and I. G. Sudirtha, "Instruments development of tri kaya parisudha-based countenance model in evaluating the blended learning," *International Journal of Engineering Pedagogy (iJEP)*, vol. 9, no. 5, pp. 55-74, 2019.
- [24] D. G. H. Divayana, P. Wayan Arta Suyasa, N.K. Widiartini, "An innovative model as evaluation model for information technology-based learning at ICT vocational schools," *Heliyon*, vol. 7, no. 2, pp. 1-13, 2021.
- [25] Z. F. F. Putra, H. Ajie, and I. A. Safitri, "Designing a user interface and user experience from piring makanku application by using figma application for teens," *International Journal of Information System & Technology*, vol. 5, no. 3, 308-315, 2021.
- [26] N. Samrgandi, "User interface design & evaluation of mobile applications," *International Journal of Computer Science and Network Security*, vol. 21, no. 1, pp. 55-63, 2021.
- [27] I. Darmawan, M. S. Anwar, A. Rahmatulloh, and H. Sulastris, "Design thinking approach for user interface design and user experience on campus academic information systems," *International Journal on Informatics Visualization*, vol. 6, no. 2, pp. 327-334, 2022.

Early Alzheimer's Disease Detection Through Targeting the Feature Extraction Using CNNs

D Prasad¹, K Jayanthi², Pradeep Tilakan³

Dept. of Electronics and Communication Engineering, Puducherry Technological University, Puducherry, India^{1,2}

Dept. of Psychiatry, Pondicherry Institute of Medical Sciences, Puducherry, India³

Abstract—Alzheimer's Disease (AD) is a persistent, irreversible, and degenerative neurological disorder of the brain that currently has no effective therapy. This condition is identified by pathological abnormalities in the hippocampal area, which may develop up to 10 years prior to the onset of clinical symptoms. Timely detection of pathogenic abnormalities is essential to impede the worsening of AD. Recent studies on neuroimaging have shown that the use of Deep Learning techniques to analyze multimodal brain scans may effectively and correctly detect AD. The main goal of this work is to design and develop an Artificial Intelligence (AI) based diagnostic framework that can accurately and promptly detect AD by analyzing Structural Magnetic Resonance Imaging (SMRI) data. This study presents a novel approach that combines a Directed Acyclic Graph 3D-CNN with an SVM classifier for timely detection and identification of AD by analyzing the Regions of Interest (RoI) like cerebral spinal fluid, white and gray matter, and the hippocampus in SMRI images. The proposed hybrid model combines Deep Learning for feature extraction and Machine Learning techniques for classification. The obtained results demonstrate its superior performance compared to earlier methods in accurately identifying individuals with early mild cognitive impairment (EMCI) from those with normal cognition (NC) using the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. The model attains a classification accuracy of 97.67%, with precision at 94.12%, and sensitivity at 98.60%.

Keywords—Alzheimer's Disease (AD); convolutional neural networks (CNN); Support Vector Machine (SVM); Directed Acyclic Graph (DAG); Late Mild Cognitive Impairment (LMCI); Alzheimer's Disease Neuroimaging Initiative (ADNI)

I. INTRODUCTION

Dementia is a broad word that encompasses many cognitive impairments that hinder daily functioning, impacting memory, thinking, language, and problem-solving skills. AD is the main source of dementia with distinct pathological features in the brain, responsible for around 80% of cases [1]. AD is defined by permanent neurodegeneration and currently lacks the potential for treatment. Neurodegenerative illnesses provide significant challenges in countries with a mostly aging population. It is the sixth primary cause of mortality and has a substantial global impact, mostly affecting the older demographic [2]. MR imaging is a diagnostic modality that uses T1-weighted images to identify and examine the morphological and structural irregularities associated with brain atrophy [3]. Therefore, MR imaging plays a crucial part in screening and diagnosing of AD [4, 5]. The incidence of the ailment has surpassed original forecasts due to the increasing

older population and the simultaneous commencement of their diagnosis [6]. This necessitates due attention to effectively handle Alzheimer's diagnosis and treatment.

While the exact process behind the progression of Alzheimer's is still not fully understood, existing knowledge indicates that the illness may be broadly categorized into three separate stages i.e. Preclinical, MCI, and AD [7, 8]. There are no noticeable symptoms of AD in the preclinical stage. However, subtle changes begin to occur in the brain. These pathological changes can start many years, even decades, before any cognitive symptoms appear [9]. The second stage is characterized by MCI, where individuals start to notice slight but measurable changes in their cognitive abilities, particularly memory. These changes are more significant than what is expected from normal aging, although they do not reach a level of severity that hinders one's ability to carry out everyday activities [10]. In the last stage of AD, the cognitive decline becomes sufficiently pronounced to disrupt everyday activities. Individuals may encounter memory impairment, disorientation, and challenges with tasks that require planning or decision-making, may struggle with recognizing familiar people or places, and may experience a decline in physical abilities such as walking, swallowing, and bladder and bowel control [11].

The field of Machine Learning (ML) and Deep Learning (DL) has gathered considerable attention in over the past few years for its ability to accurately detect and isolate possible features of dementia illness, by accurately identifying the minute morphological changes in brain structure by analyzing MRI data [12, 13]. DL methodologies have proven that the area of AD detection has seen notable advancements via the use of CNNs [4, 14]. CNNs shall effectively extract structural characteristics from T1 MRI data with a large number of dimensions, leading to more precise tailored diagnoses. Furthermore, the implementation of an ensemble approach is gaining more significance in the field of medical image evaluation [15, 16]. AD has a quick and profound impact on the hippocampus, making it one of the most damaged brain areas and making it vital for the prompt detection of AD. The hippocampus is composed of several subdomains, each exhibiting distinct characteristics. A comprehensive evaluation of neurodegenerative disorders in medical applications heavily relies on the subfields of the hippocampus [17]. Scholars have proposed that the analysis of form and volume characteristics of hippocampal subfields in many MRI scans provides advantages in the prompt identification and assessment of AD [18, 19].

*Corresponding Author

The process of classifying hippocampus characteristics entails extracting them from either 2D or 3D MRI images using 2D and 3D CNNs [20,21]. When doing a comparison between 3D convolutions performed on a whole MRI and 2D convolutions performed on slices, it is evident that the former can capture crucial 3D structural information that is vital for distinction [22]. The brain MRI data has a lot of dimensions, thus, three-dimensional CNNs [23] are computationally difficult and need a longer training time compared to two-dimensional CNNs. The above-said facts served as the motivation for this work. The goal of this research is to facilitate the timely detection of AD using the popular SVM classifier with more emphasis on the input features fed to it. This is accomplished by employing a new DAG CNN approach to perform feature extraction. In this study, both 2D and 3D DAG-CNN is used. In summary, a hybrid of CNN combined with SVM classifier is employed to perform early detection of AD. The next section gives a detailed comment on the literature work done in this direction.

II. LITERATURE REVIEW

Hongbo Xu et al. [24] proposed a CNN that utilizes multi-scale attention to diagnose AD by analyzing hippocampal subfields. This study employs two datasets, procured from Peking University of China and ADNI. These datasets consist of a combined sample of 283 NC patients and 241 AD cases. The network can easily extract 3D data characteristics from three different planes of hippocampus subfields as input. This improves computational efficiency and reduces network complexity. Experimental methods have shown notable classification performance in identifying AD, eliminating the need for manual feature extraction. Bo Liu et al. [25] employ MRI scans of the hippocampus and an attention mechanism (DenseNet-AM) to improve classification accuracy. The empirical findings illustrate that the DenseNet-AM is 92.8% accurate, the sensitivity is 97.1%, and the specificity is 89.6% when distinguishing between instances of cognitive normalcy and AD. Malik et al. [26] presented a novel methodology known as the intuitionistic fuzzy random vector functional link network (IFRVFL), which utilizes brain imaging data to diagnose AD. This study aims to improve existing approaches by incorporating a fuzzy weighted approach into the IFRVFL model to improve its capability to withstand challenges. This methodology considers the importance of individual data samples while minimizing the influence of outliers and noise. Experimental studies indicate that in comparison to cases of Alzheimer's dementia (AD), the IFRVFL model has a higher level of efficacy in identifying both MCI and early identification of AD in clinical settings. Shuqiang Wang et al. [27] conducted a research where they introduced a new approach that combines 3D-DenseNets to automatically diagnose Alzheimer's (AD) and moderate cognitive impairment by analyzing 3D brain magnetic resonance images. A comprehensive assessment was conducted for evaluating the performance of the suggested model using the ADNI dataset with 833 patients, and it was determined to be superior. The suggested approach enhances the transmission of information across layers by integrating several connections and a weighted-based combining approach is employed to integrate diverse topologies. A promising result was seen in the

automation of dementia illness identification by employing an ensemble strategy that incorporates dense connections and a weighted-based fusion method. Reddy et al. [28] provide a deep hybrid framework in their research, which employs boosting approaches to classify 3D MRI images of Alzheimer's. The research primarily focuses on early diagnosis and utilizes the categorization of subcategories of MCI. The system uses ResNet 50 and VGG16 to extract structural information from MRI volumes followed by using Extreme Gradient Boosting (XGBoost) for classification. Pallawi et al. [29] employed a Transfer Learning approach to distinguish between various phases of Alzheimer's with an enhanced EfficientNetB0 model via MRI images obtained from the Kaggle dataset. To tackle the issue of inadequate data, data augmentation techniques were used. Consequently, the model effectively categorized several classes with a precision rate of 95.78%, exceeding the efficacy of existing methodologies. Rui Guo et al. [30] provide a novel approach known as graph-based fusion (GBF) in their research. This technique utilizes imaging, genomic, and clinical data to effectively identify degenerative illnesses. By combining a multi-graph fusion module with an imaging-genetic combining module to efficiently extract unique information from many data modalities. The effectiveness of the GBF approach is shown by trials done on benchmarks about the identification of degenerative illnesses, in contrast to known graph-based methods. In their study, Xiaowei Yu et al. [31] has focused on developing a supervised deep tree model (SDTree) to forecast the advancement of AD at an individual level. The proposed SDTree methodology employs a nonlinear reversed graph embedding method inside a hierarchical tree framework within a latent space for enhanced prediction. This technique encompasses the whole spectrum of Alzheimer's progression and enables the generation of predictions for future instances. Furthermore, the attainment of a resilient depiction of the tree is accomplished by using node clustering in locations with high population density. Moreover, a novel methodology is suggested for multi-class classification by using a supervised deep tree architecture that integrates class labels to guide the acquisition of tree structure.

The reviewed works focus on several facets of AD diagnosis, using novel methodologies such as multi-scale attention (Hongbo Xu et al.) and attention processes (Bo Liu et al.) to improve efficiency and precision. Innovative approaches like IFRVFL (Malik et al.) proficiently manage noise and outliers, while 3D-DenseNets (Shuqiang Wang et al.) and hybrid frameworks integrating ResNet and XGBoost (Reddy et al.) emphasize early identification of MCI categorization. Transfer Learning (Pallawi et al.) attains high precision by data augmentation, whereas graph-based fusion (Rui Guo et al.) amalgamates multi-modal data to enhance accuracy. The SDTree model (Xiaowei Yu et al.) provides a comprehensive framework for predicting AD development.

The research highlights many constraints in the categorization systems used for AD in the literature. Hongbo Xu et al. [24] and Pallawi et al. [29] demonstrate how dataset variety limits model generalizability to larger populations. In approaches like Malik et al. [26], Shuqiang Wang et al. [27], and Rui Guo et al. [30], computational complexity is a major

issue. Dense networks, fuzzy logic systems, and graph-based solutions need plenty of resources, limiting scalability. Many researches, such as Bo Liu et al. [25] and Pallawi et al. [29], focus on specific brain areas or use single-modal data, missing the opportunity to increase accuracy via multi-modal integration. The techniques of Reddy et al. [28] and Xiaowei Yu et al. [31] enhance architectural complexity, which reduces interpretability and hinders clinical applicability. Finally, research like Bo Liu et al. [25] and Malik et al. [26] lack rigorous evaluations compared to state-of-the-art methodologies or real-world clinical datasets, leaving practical robustness untested.

Conventional methods often need the extraction of features by hand, a process that may be time-consuming and potentially overlook crucial attributes. In addition, current models may have difficulties in dealing with noise, resulting in a decrease in classification accuracy. Additionally, there are difficulties in accurately detecting the initial phases of AD and differentiating them between various phases of cognitive decline. The issue of inadequate data might ultimately restrict the capacity to train resilient models, hence affecting their overall effectiveness.

To overcome these limitations, the researchers in this study have developed methods like DAG CNNs for automatically extracting features from the SMRI images, by avoiding the need for manual feature extraction and capturing more relevant characteristics. Advanced architectures like DAG CNNs enhance accuracy by better analyzing complex data, such as brain MRI scans. This research also used data augmentation methods to overcome the problem of inadequate data by artificially expanding the amount and variety of the dataset. This results in improved training and more efficient models. The hybrid models used in this study combine the strengths of multiple methods like DL for extracting the significant features and ML techniques for performing classification to further improve classification performance.

A. Novelty

The proposed methodology is innovative in integrating DAG 3D-CNN with SVM to facilitate the prompt detection of AD, utilizing the advantages of both methodologies." Although CNNs are proficient in feature extraction, SVM classifiers are

recognized for their resilience in managing small sample data and high-dimensional feature spaces, which are typical issues in medical imaging datasets. This hybrid method offers a distinctive means to enhance classification efficacy in AD diagnosis.

The DAG architecture for CNNs facilitates a versatile route for feature propagation and allows for more profound feature investigation, circumventing the vanishing gradient issue. Our method provides a customized solution to the unique problems of volumetric medical data by building the architecture specifically for these issues, distinguishing it from conventional CNN architectures used in analogous applications.

The researchers have chosen DAG-CNN architecture because to its capacity for parallel processing of features across several scales, enhancing the network's proficiency in capturing the spatial hierarchies present in 3D medical pictures. This structure enhances generalization by mitigating overfitting, since the modular architecture allows for selective feature aggregation.

III. METHODOLOGY

Fig. 1 demonstrates a system specifically developed for the prompt identification and categorization of AD. This approach uses CNN and SVM. The method starts by obtaining the brain's SMRI data from ADNI, and KAGGLE datasets. These images are essential for discerning structural alterations corresponding to AD since they record intricate details about the brain's composition. Following that, the SMRI images go through pre-processing, a crucial stage that employs methods including skull stripping, normalization, shrinking, and noise reduction. Pre-processing ensures that the pictures are normalized, strengthening key characteristics while decreasing noise and unnecessary details, thereby improving the accuracy of further investigations. Once the data has been pre-processed, it is then split into two distinct training and validation sets. These two datasets are then used for training and assessing efficacy of the model, during the training process, therefore mitigating overfitting and ensuring the model's ability to effectively generalize to novel data.

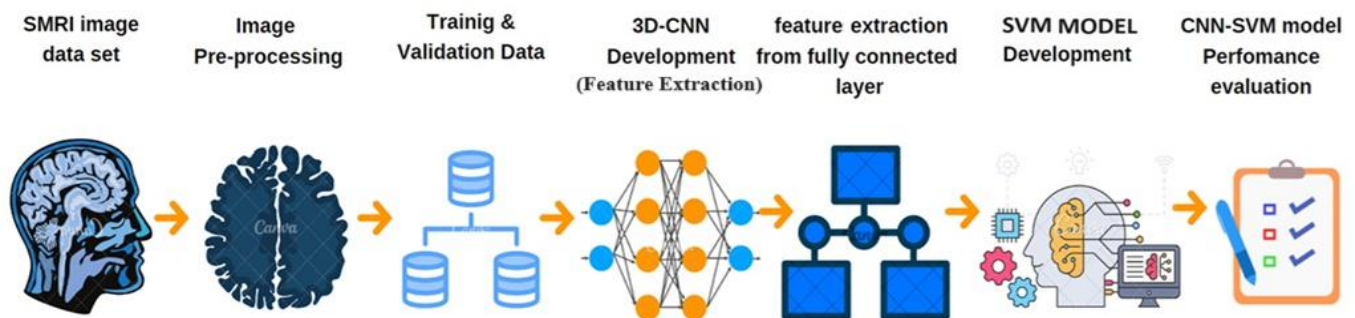


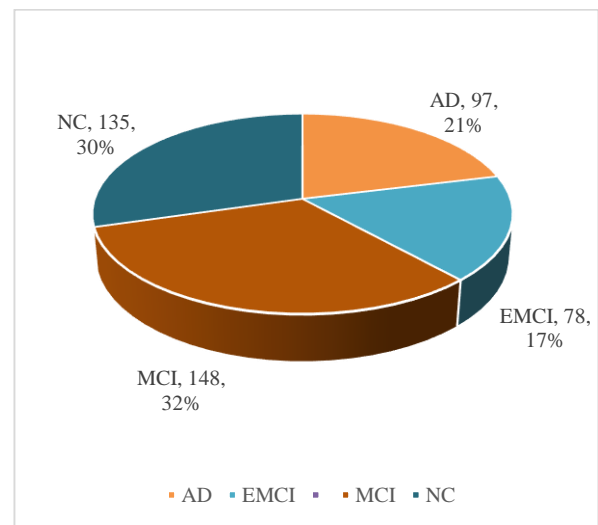
Fig. 1. Block diagram for the methodology.

The crux of this method is in the development of a CNN network, which excels at processing volumetric structural MRI data with great efficiency. The CNN automatically acquires the ability to extract spatial characteristics from brain images that are symptomatic of AD. At first, a 2D-CNN was developed to analyze the 2D slices of the MRI images. The 2D-CNN model has many advantages, namely its simplicity and decreased computational expenses. This is due to its ability to evaluate images on a per-slice basis, resulting in faster training and lower resource requirements. The implementation of the 2D-CNN model is straightforward, and the training periods are quicker because of the reduced complexity in processing 2D images. However, the 2D-CNN approach does have significant drawbacks. 2D CNNs evaluate each slice independently, which might result in the loss of crucial spatial connections between slices and the omission of significant characteristics necessary for precise Alzheimer's diagnosis. Due to the constraints of 2D-CNNs in accurately representing the whole 3D architecture of the brain, it became imperative to switch to a 3D-CNN. The 3D-CNN model enables the analysis of the whole volumetric SMRI data while maintaining the spatial connections between various brain areas. This comprehensive approach allows the model to better detect minor alterations in structure that are linked to the early stages of AD, resulting in enhanced accuracy in categorization. The use of a 3D model aligns with the objective of achieving improved accuracy and reliability in the early detection of AD, making it a vital step in our research. Once the network completes the image processing, it proceeds to extract the significant features from the fully connected layers of the CNN. These layers function as classifiers inside the network and integrates the features collected into a condensed representation. Subsequently, this representation is used to train an SVM model, which categorizes the data into distinct classes, such as AD or normal cognitive. The SVM classifier is used because of its resilience in distinguishing classes in spaces with a large number of dimensions, hence enhancing its effectiveness as a classifier when paired with the characteristics retrieved by the CNN.

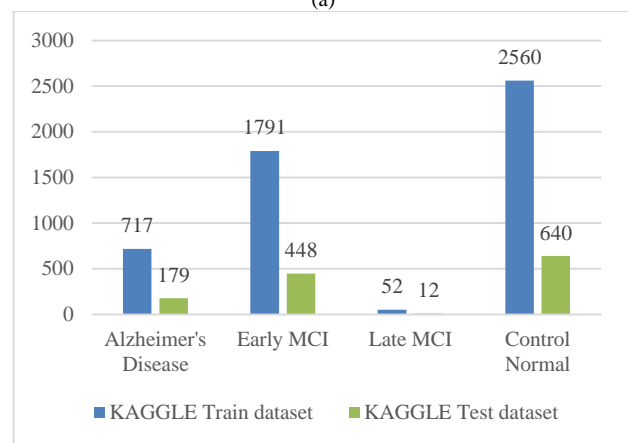
Finally, the efficacy of the integrated CNN with SVM model is assessed by measuring parameters such as accuracy, precision, recall, and F1 score. This assessment is vital in guaranteeing that the suggested model is not just accurate but also reliable and has the ability to extrapolate well to new, unfamiliar data. In summary, our technique successfully integrates DL for automated extraction of features using traditional ML for classifiers, resulting in an efficient method for the prompt identification of dementia. Such detection is essential for immediate attention and treatment.

A. SMRI Datasets

This research used structural MRI images received from the ADNI and KAGGLE databases. The ADNI consists of four distinct phases, including ADNI 1, ADNI GO, ADNI2, and ADNI4. Each phase has its specific aims and cognitive stages. This study used structural MRI images. Fig. 2(a) depicts the AD dataset utilized in this study. Among the 455 participants in the ADNI study, there were 97 AD subjects, 78 early MCI subjects, 148 LMCI subjects, and 135 subjects with normal cognition (NC).



(a)



(b)

Fig. 2. (a) ADNI data set, (b) KAGGLE data set.

The dataset obtained from Kaggle shown in Fig. 2(b) and has four different classes: NC, Early EMCI, Late MCI, and AD. The dataset is divided into separate training and test sets. The training dataset composed of 5120 photos, whereas the test dataset has 1279 images. The NC class contains the largest quantity of photos, consisting of 2560 for training and 640 for assessment. The EMCI dataset consists of 1791 pictures for training and 448 images for testing, whereas the AD dataset comprises 717 training images and 179 test images. The LMCI class has the lowest number of pictures, with a total of 52 for training and 12 for assessment.

B. Preprocessing and Segmentation of SMRI Images

Fig. 3 depicts the methodological steps employed for the early identification of AD using the data samples acquired from the ADNI and KAGGLE. The collected images are in NifTi (.nii) format. The N4ITK bias correction is first performed to eliminate low-frequency noise and the resulted image are shown in Fig. 3(a). Subsequently, the raw volumes are subjected to pre-processing through the Statistical Parametric Mapping [45] toolbox in MATLAB. During the pre-processing stage of SPM, the images were co-registered with the ICBM-152 template to align them to the Montreal Neurological Institute coordinate system (MNI) and

additionally, the images are normalized, skull stripped (Fig. 3b) and categorized into white, gray matter, and cerebral spinal fluid as shown in Fig. 3(c). Hippodeep [44] tool is used for segmenting left and right hippocampal and shown in Fig. 3(d). Subsequently, the images were resized to dimensions of 256*256*128. The pre-processed data is separated into two distinct sets: training and validation. The proposed 3D-CNN is the trained and validated via these two datasets. The trained CNN is utilized to extract important characteristics from the input SMRI data. These characteristics obtained from the flattening layer of CNN are further partitioned into validation and training features, which are then utilized for both training and testing of the suggested DAG based 3D-CNN with SVM model.

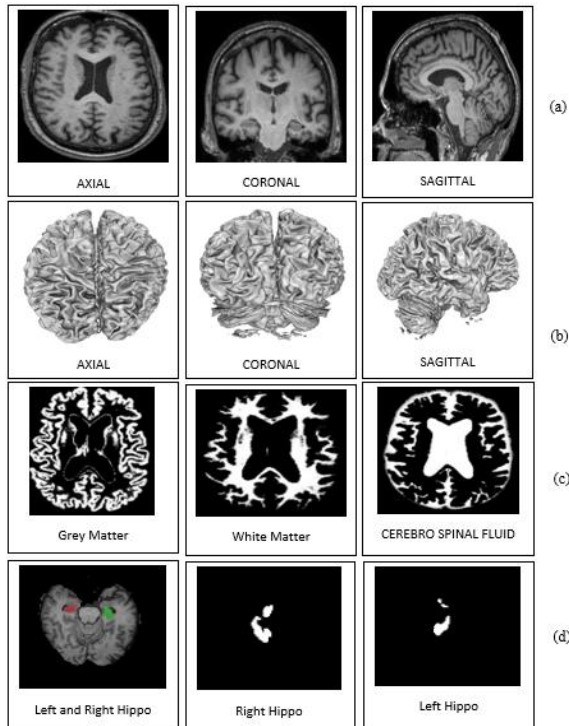


Fig. 3. (a) N4 bias correction, (b) Skull stripping, (c) Segmented WM, GM, & CSF, (d) Segmented Hippo.

C. Design and Development of DAG Based 2D/3D-CNN

The CNN and ML classifiers are often used AI tools for the identification and categorization of Alzheimer's, and their performance largely relies on the features extracted and analysed. Traditionally, researchers manually extract certain characteristics, which are then incorporated into ML classifiers for categorization. This research study employs a DAG based 2D/3D layered CNN model as shown in Fig. 4 for automatically extracting characteristics from SMRI data, leveraging their strength in handling substantial volumes of visual information. The architecture remains the same for both 2D and 3D except that the layers are made to handle 2D and 3D data respectively. This variant is brought in to observe the magnitude of change in the classifier performance metrics, which 3D layers offers compared to the 2D layers in the suggested CNN framework. The proposed CNN framework uses multiple paths and concatenation layers to learn, extract, and combine diverse feature representations, improving the

model's proficiency in appropriately classifying AD. The core structure of the proposed CNN framework comprises of four distinct layers: convolutional, normalization, pooling, and activation. The proposed model employs a total of six convolutional blocks, each composed of four layers, namely convolutional layer (CL), batch normalization (BN), max pooling (MP), and leakyRelu (LR) activation layer. The convolutional layer applies 3*3*3 kernel filtering to extract various characteristics from the input SMRI images, subsequently accompanied by a LeakyReLU activation, batch normalization, and max pooling to extract and condense the significant features from SMRI images effectively. These extracted features by pooling layer are fed to ML classifier. The pooling layer decreases the spatial dimensions of the feature maps, effectively decreasing the trainable variables and controlling overfitting. The BN layer helps to stabilize and speed up training by normalizing the inputs to the next layer. After passing through the series of convolutional blocks, the feature maps are transformed into a one-dimensional vector before being inputted into fully connected layers for final classification. Prior to feeding the data into the CNN, MRI images are pre-processed, resized to dimensions of 256x256x128, and normalized. The DAG based 3D-CNN was trained for 10 epochs with a learning rate of 0.001, using the Adam optimizer with categorical cross-entropy loss, employing a batch size of 8. The dataset's class imbalance was addressed by the use of class weights.

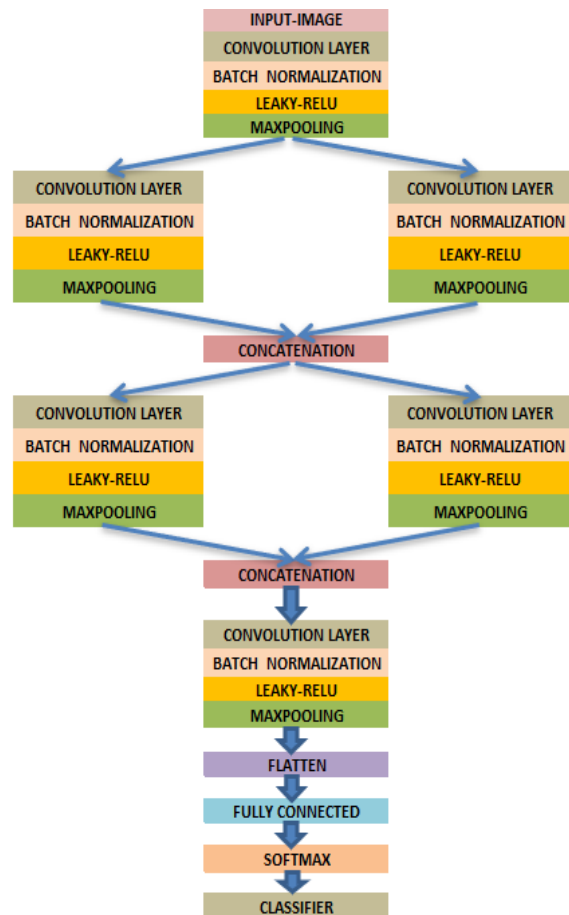


Fig. 4. Proposed DAG-CNN architecture.

1) *Advantages of DAG-CNN over other architectures:* The DAG structure enhances computational performance and minimizes redundancy in feature extraction compared to conventional CNN systems. In contrast to ResNets or DenseNets, which depend on unique skip connections, the DAG architecture offers a more universal approach for adaptive feature flow, especially advantageous for 3D data where spatial information is essential.

The integration of CNN and SVM arises from the use of their complementing advantages: CNNs excel at extracting deep, hierarchical features, whilst SVMs are particularly proficient in classification problems involving tiny or unbalanced datasets. This is especially pertinent in the early identification of AD where the quantity of the information often poses a constraint.

The CNN-SVM combination offers superior feature separation compared to end-to-end CNN classifiers, since SVM emphasizes optimizing the separation among classes in a high-dimensional space. This hybrid method guarantees that the retrieved characteristics are both deep and properly distinguished for classification, resulting in enhanced sensitivity and specificity.

D. Framework of the Proposed DAG Based 3D-CNN with SVM Classifier

The primary aim of this investigation is to improve upon existing approaches by modifying the architectures of CNNs to extract critical features and ML classifiers for accurate AD categorization. The proposed architecture shown in Fig. 5 has high capacity to produce innovative solutions that can efficiently detect Alzheimer's in its initial stages.

The use of hybrid DAG 3D-CNN with an SVM classifier has been demonstrated to be a very efficient methodology for a diverse array of classification jobs. The superiority of the DAG 3D-CNN with SVM lies in its hybrid nature. The proposed model by combining a CNN for characteristics extraction with an SVM for classification leverages the strengths of both techniques. The enhancement of classification performance, interpretability, and generalization ability is accomplished with the resilience of SVMs and the feature extraction capabilities of CNNs. Support Vector Machines (SVMs) use non-linear mechanisms and flexibility to enhance decision boundaries and accuracy, CNNs have a remarkable ability to obtain hierarchical and discriminative features from raw input data, which are crucial for identifying pathological changes associated with AD. Furthermore, the classifier incorporates the regularization properties of Support Vector Machines (SVMs), ensuring robustness against overfitting. The method shown in Fig. 5 depicts the flow of data through a hybrid DL and ML model for Alzheimer's classification.

The SVM is a popularly used ML methodology utilized for categorization tasks. The methodology is specifically formulated to ascertain the hyperplane that optimizes the degree of differentiation between observations that correspond to different classifications. The mathematical formulation of the decision function for Support Vector Machines (SVM) is:

$$f(x) = \text{sign}(w \cdot x + b) \quad (1)$$

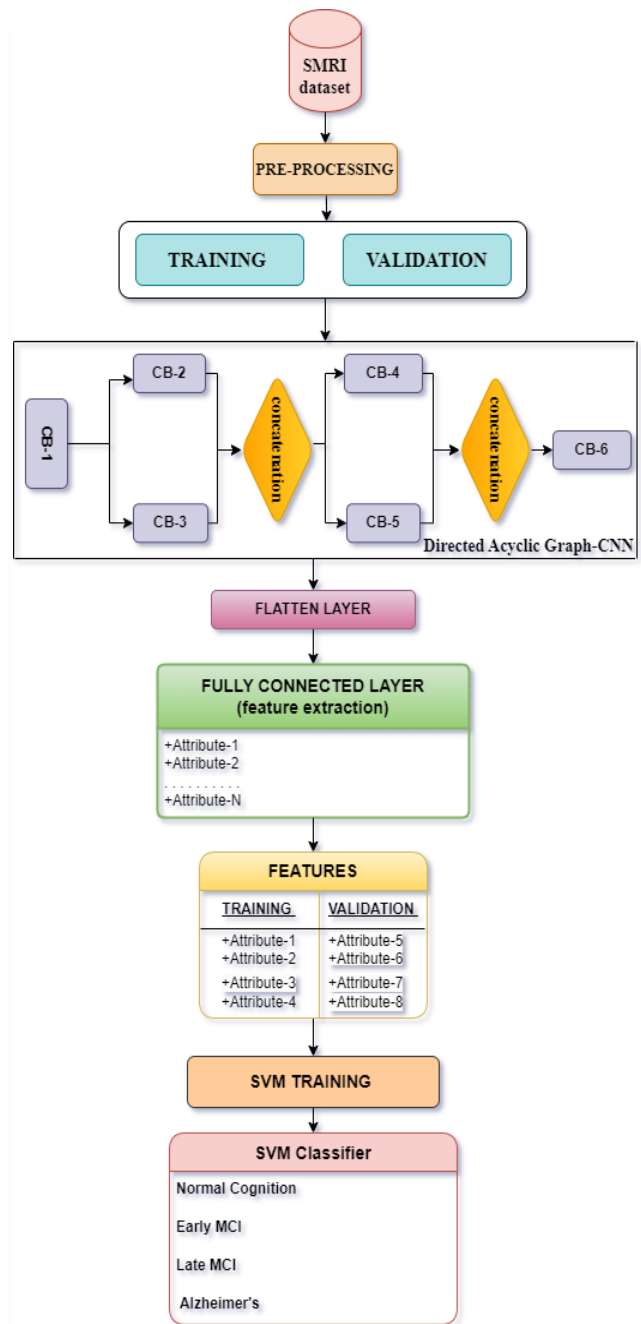


Fig. 5. Proposed DAG-CNN with SVM classifier.

where b is the bias factor, the weight vector is w , an input feature vector is x , and the sign function is denoted by $\text{sign}(\cdot)$. The distance between the nearest data point x_i and the hyperplane known as support vectors is derived using the below Eq. (2).

$$\text{distance}(x_i, \text{hyperplane}) = \frac{|w \cdot x_i + b|}{\|w\|} \quad (2)$$

$$\|w\| = (w_1^2 + w_2^2 + \dots + w_n^2)^{1/2} \quad (3)$$

where b is the bias factor, the weight vector is w , and $\|w\|$ is the magnitude or Euclidean norm of weight vector and

calculated as shown in Eq. (3), an input feature vector in n -dimensional space is x_i . The margin is inversely proportional to the size of w . The main objective of SVM is to ascertain the ideal values of w and b coefficients that enable the attainment of the largest margin. The ideal values for the weight vector w and bias b are obtained by addressing an optimization problem that tries to increase the margin between classes while decreasing misclassification errors. In case of data that is not linearly separable, Support Vector Machines (SVM) include a slack variable ξ_i for each data point. This variable allows for a certain degree of misclassification, striking a balance between maximizing the margin and minimizing errors. This results in an optimization problem in which the goal is to minimize the expression as shown in Eq. (3).

$$distance(x_i, hyperplane) = \frac{1}{2} \|w\|^2 + C \cdot \sum \xi_i \quad (4)$$

where C determines the balance between the size of the margin and the penalty for misclassification.

IV. RESULTS AND DISCUSSION

Timely identification is crucial in effectively controlling and perhaps slowing down the progression of Alzheimer's dementia, making it an essential area of investigation. Hence the primary objective of this current study is to develop a highly efficient hybrid AI model with the combination of DAG-CNN and SVM classifier that can identify AD by analyzing SMRI data.

Both Classification and early detection of AD are accomplished by using the proposed DAG CNN and SVM framework, developed using MATLAB 2022b. 80% of pre-processed images were utilized for training and 20% for validation.

AD may be classified as four different phases: Preclinical (Normal Cognitive), Early MCI, Late MCI, and AD. This research study attempted on three distinct binary classifications under three case studies.

- Case 1: EMCI Vs subjects with Normal Cognition (NC). This distinction is of utmost importance in detection of subjects in the initial stages of AD.
- Case 2: EMCI with LMCI
- Case 3: LMCI with AD

Initially, this study involved in extracting the volumetric features manually using the ITK-SNAP[43] cloud-based application, specifically focused on SMRI images. Around 22 volumetric characteristics were extracted from specific regions inside the hippocampus. These manually extracted features were fed as input for a SVM classifier, which yielded an accuracy of 88.4% for discriminating EMCI with Normal Cognitive (case 1 scenario), which triggered the use of CNN-based automated feature extraction to achieve improved accuracy. Table I depicts the performance of an SVM model achieved for 22 manually extracted volumetric features shown in Table II from hippocampal subfields. The performance metrics are listed for SVM classifier.

Next, a 2D-CNN was designed and used to analyze 2D slices of the SMRI images available in the ADNI and

KAGGLE repository. On the Kaggle dataset, the 2D-CNN module without an SVM classifier demonstrated a classification accuracy of 90.17% in differentiating between NC and EMCI, 98.98% in differentiating between NC and AD, and 90.43% in differentiating between EMCI and LMCI. The accuracy in distinguishing between NC and EMCI, AD and NC, and EMCI and LMCI using the 2D-CNN architecture without an SVM classifier were 94.20%, 89.97%, and 82.17% respectively, for the ADNI dataset. The suggested DAG 2D-CNN without the SVM classifier model's efficacy metrics for the Kaggle and ADNI datasets are presented in Table III.

However, since the accuracy was not optimal for all cases, the proposed CNN architecture with 2D layers was converted into 3D layers which is capable of processing the complete volumetric SMRI data and capturing the spatial connections inside the brain's structure to enhance the model's capability to accurately diagnose the Alzheimer's in early-stage. The suggested DAG 3D-CNN without the SVM classifier model's efficacy metrics for the ADNI dataset are presented in Table IV. The accuracy in distinguishing between NC and EMCI, AD and NC, and EMCI and LMCI using the 3D-CNN architecture without an SVM classifier were 96.86%, 90.45%, and 96.67% respectively, for the ADNI dataset.

The proposed DAG 3D-CNN with SVM classifier outperforms the 2D and 3D-CNN modules. With the ADNI dataset, the hybrid DAG 3D-CNN with the SVM model is 97.67 per cent accurate. Tables V and VI provide the performance outcomes of the hybrid DAG 3D-CNN with SVM classifier and the comparison of SVM model, 2D-CNN for ADNI and KAGGLE datasets & 3D-CNN with and without SVM models, respectively, to identify the Alzheimer's at an initial stage. The comparative analysis of different models and their performance for the ADNI and KAGGLE dataset is shown in Table VI. This evaluation considered five regions of interest (ROIs).

TABLE I. PERFORMANCE EVALUATION OF THE SVM CLASSIFIER FOR MANUALLY EXTRACTED VOLUMETRIC FEATURES OF HIPPOCAMPAL SUBFIELDS USING ITK-SNAP FOR ADNI DATASET

Classification	Accuracy	Precision	Sensitivity	F1 Score
NC with EMCI	88.40	86.60	94.00	0.90
EMCI with LMCI	80.40	80.70	80.40	0.80
LMCI with AD	86.00	94.00	78.00	0.85

Table I and Fig. 6 displays the efficacy metrics of an SVM classifier used for categorizing various phases of AD. The classification is determined by analysing 22 volumetric characteristics extracted from hippocampus subfields using the ITK-SNAP[43] tool, which are depicted in Table II. The classifier achieved 88.40% accuracy in differentiating EMCI from NC. The classification model achieved a precision of 86.60%, a sensitivity of 94.00%, and an F1 score of 0.90. These findings demonstrate that the classification method is very successful in identifying Alzheimer's in its first stages. Nevertheless, although achieving satisfactory outcomes, the accuracy remains inferior to the findings reported in the research literature. Thus, to improve the precision and efficacy of prompt identification, our research study has shifted to

CNN-based automated feature extraction, yielding superior outcomes.

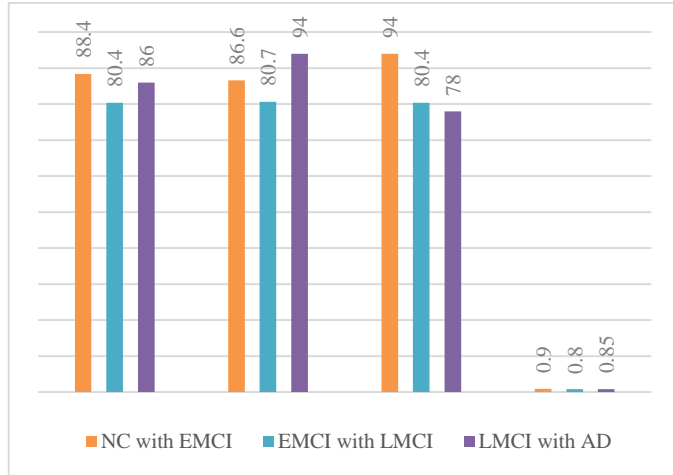


Fig. 6. Performance metrics of SVM classifier.

TABLE II. VOLUMETRIC FEATURES EXTRACTED FROM THE HIPPOCAMPUS SUBFIELDS FOR EARLY DETECTION AND CLASSIFICATION FROM ADNI DATASET

22 Volumetric Features extracted from the hippocampus subfields			
Sl. No.	Left Hippo	Sl. No.	Right Hippo
1	Left CA1 (Corno Ammonis 1)	12	Right CA1 (Corno Ammonis 1)
2	Left CA2	13	Right CA2
3	Left CA3	14	Right CA3
4	Left DG (Dentate Gyrus)	15	Right DG (Dentate Gyrus)
5	Left Tail	16	Right Tail
6	Left Sub (Subiculum)	17	Right Sub (Subiculum)
7	Left Erc (Entorhinal Cortex)	18	Right Erc (Entorhinal Cortex)
8	Left A35	19	Right A35
9	Left A36	20	Right A36
10	Left Phc (Parahippocampal Cortex)	21	Right Phc (Parahippocampal Cortex)
11	Left Cysts	22	Right Cysts

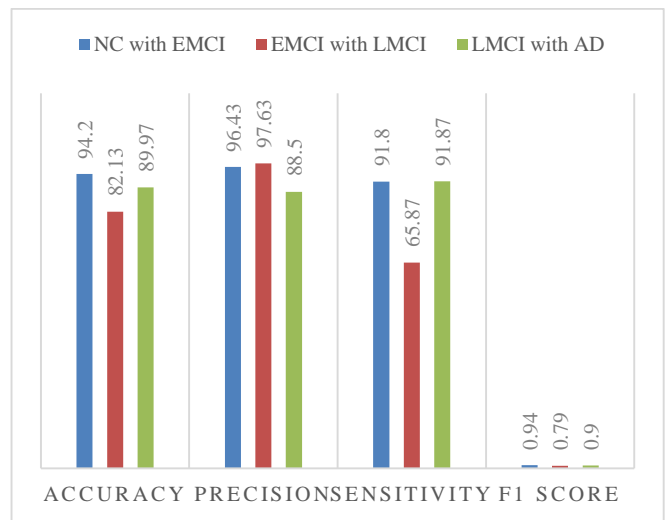
The hippocampus has morphologically and functionally diverse subfields that differ in AD susceptibility. Early AD begins with tau buildup and neuronal loss in CA1. Though seldom studied, recent findings suggest CA2 role in social memory and pathological changes in AD. CA3 and Dentate Gyrus (DG) are Essential for pattern separation. structural changes may cause early cognitive impairment. The entorhinal cortex (ERC) and perirhinal cortex (PHC), which are crucial for hippocampal input and output, are among the first areas to atrophy in AD. Object recognition and memory encoding depend on the perirhinal cortex near the hippocampus. Since A35 and A36 allow hippocampus-cortical memory network connection, neurodegeneration in these regions corresponds with cognitive difficulties in early AD. Recent studies show

that volumetric abnormalities in these regions suggest pathogenic processes like tau accumulation, and include them in the feature set improves sensitivity to early AD changes. Fluid-filled hippocampal cysts may suggest neurodegenerative processes including gliosis or vascular changes. Cystic changes, seldom seen in AD, may be linked to structural atrophy in nearby hippocampus subfields, improving hippocampal health assessment.

TABLE III. COMPARISON OF THE EFFICACY OF THE PROPOSED DAG 2D-CNN CLASSIFIER ON THE ADNI AND KAGGLE DATASETS

Classification	dataset	Accuracy	Precision	Sensitivity	F1 Score
NC with EMCI	adni	94.20	96.43	91.80	0.94
EMCI with LMCI		82.13	97.63	65.87	0.79
LMCI with AD		89.97	88.50	91.87	0.90
NC with EMCI	kaggle	90.17	86.86	87.05	0.87
EMCI with LMCI		90.43	89.82	97.54	0.94
LMCI with AD		98.98	98.97	100.00	0.99

Table III presents a comparison of the efficiency of the proposed framework on two datasets, namely ADNI and Kaggle. The results are shown in Fig. 7(a) and 7(b). The model is evaluated by measuring its performance metrics across three classification tasks: distinguishing EMCI from NC, LMCI from AD, and EMCI from LMCI. The model achieves high accuracy on both datasets for distinguishing EMCI from NC, with the ADNI dataset slightly outperforming the Kaggle dataset. The model shows very high accuracy and F1 score, especially on the Kaggle dataset, indicating excellent efficacy in differentiating LMCI from AD. The model performs the lowest on this classification for distinguishing EMCI from LMCI, particularly on the ADNI dataset, where sensitivity is much lower compared to the Kaggle dataset. Overall, the model demonstrates strong performance in distinguishing NC from AD, with somewhat lower performance for differentiating EMCI from LMCI



(a)

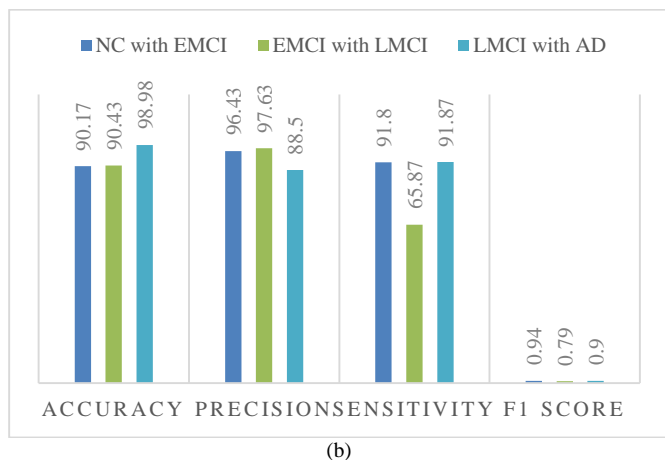


Fig. 7. (a) Performance metrics of 2D-CNN for ADNI data (b) Performance metrics of 2D-CNN for Kaggle.

TABLE IV. PERFORMANCE OF PROPOSED HYBRID DIRECTED ACYCLIC GRAPH 3D-CNN CLASSIFIER FOR ADNI DATASET

Classification	Accuracy	Precision	Sensitivity	F1 Score
NC with EMCI	96.86	100	95.04	0.97
EMCI with LMCI	90.45	94.42	90.50	0.92
LMCI with AD	96.66	96.87	96.66	0.96

Table IV shows the efficiency metrics of the suggested DAG 3D-CNN model for the ADNI dataset. The model demonstrates strong performance in distinguishing EMCI from NC and has achieved an accuracy of 96.86%. The model has remarkable performance, achieving perfect precision, a sensitivity of 95.04%, and an F1 Score of 0.97%. The model efficacy metrics are plotted and shown in Fig. 8.

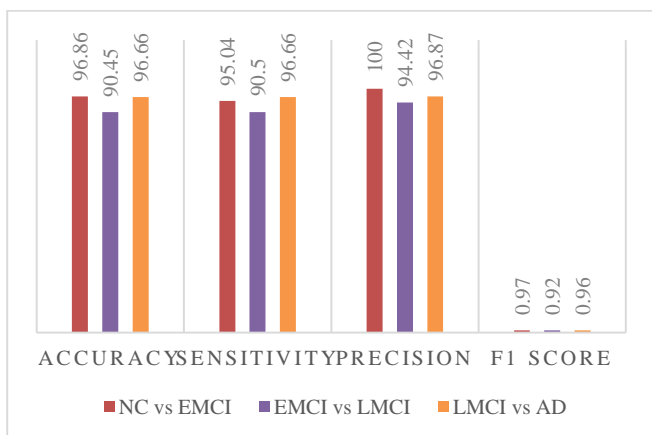


Fig. 8. Performance metrics of 3D-CNN.

TABLE V. PERFORMANCE OF PROPOSED HYBRID DIRECTED ACYCLIC GRAPH 3D-CNN WITH SVM CLASSIFIER FOR ADNI DATASET

Classification	Accuracy	Precision	Sensitivity	F1 Score
NC with EMCI	97.67	94.12	98.60	0.96
EMCI with LMCI	98.33	96.86	96.86	0.96
LMCI with AD	100	100	96.67	0.98

Table V shows the efficiency metrics of the suggested DAG 3D-CNN with the SVM classifier model for the ADNI dataset. The model demonstrates strong performance in

distinguishing EMCI from NC and has achieved an accuracy of 97.67%. The model has remarkable performance, achieving perfect precision, a sensitivity of 98.60%, and an F1 Score of 0.96%. The model efficacy metrics are plotted and shown in Fig. 9.

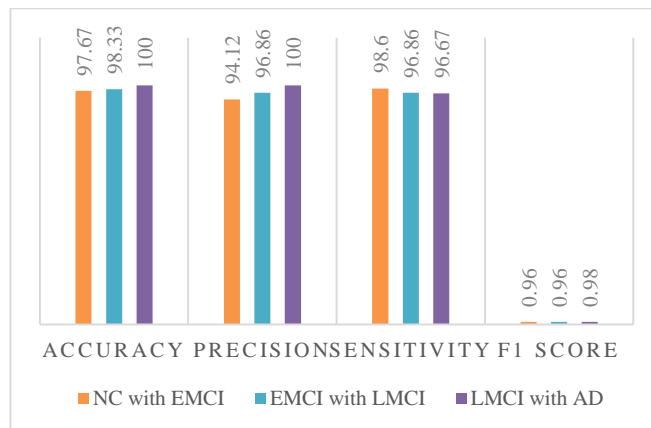


Fig. 9. Performance metrics of 3D-CNN with SVM.

TABLE VI. COMPARATIVE ANALYSIS OF THE PROPOSED MODELS FOR EARLY DETECTION OF AD

Method	Accuracy	Precision	Sensitivity	F1 Score
SVM with manually extracted features (ADNI)	88.40	86.60	94.00	0.90
2D-CNN (Kaggle)	90.17	86.86	87.05	86.96
2D-CNN (ADNI)	94.20	96.43	91.80	94.06
DAG 3D-CNN (ADNI)	96.86	100	95.04	0.97
DAG 3D-CNN with SVM (ADNI)	97.67	94.12	98.60	0.96

Table VI displays the performance measures for five distinct models employed in the early identification and categorization of AD. An SVM classifier with manually extracted features achieved 88.40% accuracy in discriminating early MCI with normal cognitive. The 2D-CNN model achieved a 90.17% accuracy when trained on Kaggle data. However, when trained on ADNI data, the same model performed better, with an accuracy of 94.20%. The DAG 3D-CNN model achieved a 96.86% accuracy when trained on ADNI data. The DAG 3D-CNN with SVM classifier surpassed all other models, with an accuracy of 97.67%. The five distinct model's Accuracy values are plotted and shown in Fig. 10.

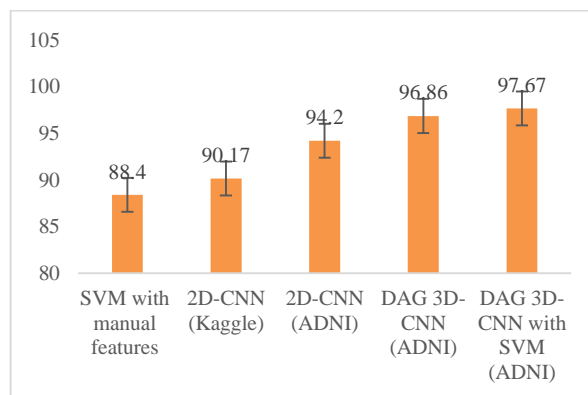


Fig. 10. Performance metrics of the proposed models.

TABLE VII. THE ACCURACY COMPARISON OF PROPOSED MODEL WITH VARIOUS ALGORITHMS FOR EARLY ALZHEIMER'S DISEASE DETECTION

Ref. No.	Dataset used	Algorithms used	Accuracy	
B. K. Choi et al. [32]	Adni	2D-CNN	78.1%	
M. Ghazal et al. [33]		3D deeply supervised adaptable CNN	93.2%	
S. Basaia et al. [34]		DL and CNN	87.1%	
C. Feng et al. [35]		3D-CNN & FSBI-LSTM.	86.36%	
Archana B et al. [36]		CNN	95.82%	
R. Joshi et al. [37]		Densenet-169	91.80%	
C. Kaur et al. [38]		Random Forest	86.24%	
S. Samanta et al. [39]		CNN	85.73%	
B. Kumar Yadav et al. [40]		CNN	94.57%	
A. J. Nair et al. [41]		VGG	90.34%	
F. Hajamohideen et al. [42]		Siamese CNN	91.83%	
Proposed model		SVM with manually extracted features	88.40%	
Proposed model		Kaggle	2D-CNN	90.17%
Proposed model		Adni	2D-CNN	94.20%
Proposed model	3D-CNN with DAG		96.86%	
Proposed model	3D-CNN with DAG and SVM classifier	97.67%		

Table VII presents a comparison of the efficiency of several methods for the ADNI and Kaggle datasets. The suggested model, which used a 3D-CNN combined with an SVM classifier, produced an impressive accuracy of 97.67%. Additional models, such as 3D-CNN without SVM attained 96.86% and those using 2D-CNNs, achieved high performance as well, with accuracies of 94.2% for ADNI and 90.17% for the Kaggle dataset respectively. The comparison clearly illustrates the better efficacy of the suggested approach, especially the DAG 3D-CNN with the SVM classifier, which attains the best accuracy of 97.67%. The performance metrics of various algorithms are plotted and shown in Fig. 10.

A. Discussion

This research aimed to create and assess sophisticated DL methodologies for the early identification of AD via SMRI datasets, employing both 2D and 3D CNN architectures in conjunction with an innovative integration of SVM classifiers. The suggested strategies shown substantial improvements in classification accuracy relative to current approaches in the literature.

Numerous recent researches have used DL methodologies for the categorization of AD, resulting in differing degrees of efficacy. B. K. Choi et al. [32] used a 2D-CNN, attaining an accuracy of 78.1%, hence underscoring the constraints of conventional 2D convolutional techniques. Advanced models, such as the 3D deeply supervised adaptive CNN by M. Ghazal et al. [33], demonstrated an accuracy of 93.2%, while frameworks like FSBI-LSTM integrated with 3D-CNN by C. Feng et al. [35] attained 86.36% accuracy.

The suggested 3D-CNN using a DAG architecture attained an accuracy of 96.86%, surpassing the majority of documented research. The integration with an SVM classifier enhanced performance to 97.67%, establishing a new standard in AD classification accuracy, exceeding prior benchmarks established by models such as Densenet-169 (91.80%) by R.

Joshi et al. [37] and Siamese CNN (91.83%) by F. Hajamohideen et al. [42]. This notable improvement is due to the DAG architecture's capacity to capture complex spatial information in SMRI data and the SVM's effective decision boundary optimization. The results indicate the capability of automated systems to offer dependable assistance in clinical decision-making for the early identification of AD.

B. Clinical Significance of the Findings

1) *Early diagnosis:* Our approach, integrating 3D-CNN with SVM for the early identification of AD, facilitates diagnosis in its first stages, perhaps prior to the onset of clinically observable cognitive impairment. Early identification is essential for prompt interventions, such cognitive therapy or pharmaceutical treatments, which may decelerate illness development.

2) *Customized therapy:* By pinpointing certain parts of the hippocampus afflicted in initial phases of AD, our approach may facilitate the development of individualized therapy techniques, focusing on the brain areas most severely impacted by the condition.

3) *Monitoring illness progression:* The volumetric alterations in the hippocampus subfields may function as biomarkers for assessing illness progression over time, providing a non-invasive instrument for doctors to evaluate treatment effectiveness and disease trajectory. The methodology may be applicable.

V. CONCLUSION

This research work introduces a novel method for the timely identification of AD via Structural MRI images. The proposed strategy utilizes deep neural networks i.e. DAG 3D-CNN for significant characteristic features extraction followed by SVM as a classifier. The model is trained and assessed by employing the Kaggle and ADNI datasets. For the Kaggle and ADNI datasets, the 2D-CNN module being evaluated offered an accuracy of 90.17% and 94.20%, 3D-CNN without SVM offered an accuracy of 96.86% and the hybrid 3D-CNN module with SVM classifier presented a superior accuracy of 97.67% in detecting EMCI subjects, respectively. This proves that the hybrid framework is relatively good and suitable for early detection and classification for all three case studies dealt in this research work. The efficacy of the suggested DAG 3D-CNN with SVM classifier technique in early Alzheimer's (AD) diagnosis shall be improved by training the network with additional clinical information, and by enhancing the number of ROIs used in the study.

A. Limitations and Future Work

This study, like other studies, has certain limitations that must be acknowledged. The sample size and insufficient demographic diversity may restrict the model's generalizability to wider groups. Subsequent research should use bigger and more heterogeneous datasets to corroborate the model's resilience across other demographics. Secondly, while this work used a 3D-CNN for SMRI data, the integration of other imaging modalities like PET and fMRI might significantly improve classification accuracy and diagnostic capabilities. Despite these constraints, this work establishes a significant

basis for enhancing automated detection techniques for AD and highlights critical avenues for further research.

Future research areas include investigating multimodal fusion by integrating SMRI with other imaging techniques like PET and fMRI, so offering a more holistic perspective on AD pathology and enhancing model efficacy. Furthermore, using longitudinal research to observe temporal changes may facilitate the building of prediction models capable of both early detection of AD and monitoring its advancement. Incorporating clinical data, including cognitive scores and genetic information, might significantly improve the model's accuracy and personalization, allowing more customized treatment strategies for AD patients. Ultimately, exploring other DL methodologies, such attention processes or reinforcement learning, might enhance model efficacy in intricate neuroimaging tasks.

ACKNOWLEDGMENT

The authors express their gratitude for the Institutional support provided in the form of a desktop server and internet.

REFERENCES

- [1] Alzheimer's Association. (2022). 2022 Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia*, 18(4), 700-757. doi:10.1002/alz.12662
- [2] Mc Dade, E., & Swanson, J. (2021). The increasing prevalence of Alzheimer's disease: A global overview. *International Journal of Geriatric Psychiatry*, 36(12), 1913-1920. doi:10.1002/gps.5587
- [3] GBKaras, Philip Scheltens, Serge ARBRombouts, Pieter Jelle Visser, Ronald AvanSchijndel, Nick C Fox, and Frederik Barkhof, "Global and local graymatter loss in mild cognitive impairment and alzheimer's disease," *Neuroimage*, vol.23, no.2, pp.708-716,2004.
- [4] M. B. T. Noor, N. Z. Zenia, M. S. Kaiser, S. A. Mamun, and M. Mahmud, "Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer's disease, Parkinson's disease and schizophrenia," *Brain Informatics*, vol. 7, no. 1, Oct. 2020, doi: 10.1186/s40708-020-00112-2.
- [5] J. Islam and Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," *Brain Informatics*, vol. 5, no. 2, May 2018, doi: 10.1186/s40708-018-0080-3.
- [6] P. Scheltens, B. De Strooper, M. Kivipelto, H. Holstege, G. Ch  telat, C. E. Teunissen, J. Cummings, and W. M. van der Flier, "Alzheimer's disease," *Lancet*, vol. 397, no. 10284, pp. 1577-1590, 2021
- [7] C. R. Jack Jr, D. S. Knopman, W. J. Jagust, L. M. Shaw, P. S. Aisen, M. W. Weiner, R. C. Petersen, and J. Q. Trojanowski, "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade," *Lancet Neurol.*, vol. 9, no. 1, pp. 119-128, 2010.
- [8] G. B. Frisoni, N. C. Fox, C. R. Jack, P. Scheltens, and P. M. Thompson, "The clinical use of structural MRI in Alzheimer disease," *Nat. Rev. Neurol.*, vol. 6, no. 2, pp. 67-77, 2010.
- [9] Jack, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., ... & Trojanowski, J. Q. (2013). Tracking pathophysiological processes in Alzheimer's disease: An updated hypothetical model of dynamic biomarkers. *The Lancet Neurology*, 12(2), 207-216. doi:10.1016/S1474-4422(12)70291-0
- [10] Petersen, R. C., Roberts, R. O., Knopman, D. S., Boeve, B. F., Geda, Y. E., Ivnik, R. J., ... & Rocca, W. A. (2009). Mild cognitive impairment: Ten years later. *Archives of Neurology*, 66(12), 1447-1455. doi:10.1001/archneurol.2009.266
- [11] Alzheimer's Association. (2023). 2023 Alzheimer's Disease Facts and Figures. *Alzheimer's & Dementia*, 19(4), 700-781. doi:10.1002/alz.13028
- [12] J. M. Ranson *et al.*, "Harnessing the potential of machine learning and artificial intelligence for dementia research," *Brain Informatics*, vol. 10, no. 1, Feb. 2023, doi: 10.1186/s40708-022-00183-3.
- [13] "Machine learning for cognitive behavioral analysis: datasets, methods, paradigms, and research directions," 2023. [Online]. Available: <https://doi.org/10.1186/s40708-023-00196-6>
- [14] Ahsan Bin Tufail, Yong-Kui Ma, et al., "Binary classification of alzheimer's disease using SMRI imaging modality and deep learning," *Journal of digital imaging*, vol.33, no.5, pp.1073-1090,2020.
- [15] C. Senaras, A. C. Moberly, T. Teknos, G. Essig, C. El maraghy, N. Taj-Schaal, L. Yua, and M. N. Gurcan, "Detection of eardrum abnormalities using ensemble deep learning approaches," in *Medical Imaging 2018: Computer-Aided Diagnosis*, vol. 10575. International Society for Optics and Photonics, 2018, p. 105751A.
- [16] R. Rasti, M. Teshnehlab, and S. L. Phung, "Breast cancer diagnosis in DCE-MRI using mixture ensemble of convolutional neural networks," *Pattern Recognition*, vol. 72, pp. 381-390, 2017.
- [17] Kichang Kwak, Marc Niethammer, et al., "Differential role for hippocampal subfields in alzheimer's disease progression revealed with deeplearning," *CerebralCortex*,2021.
- [18] Robin De Flores, Renaud La Joie, and Ga  el Ch  telat, "Structural imaging of hippocampal subfields in healthy aging and alzheimer's disease," *Neuroscience*, vol.309, pp.29-50,2015.
- [19] Susanne G Mueller, Norbert Schuff, Kristine Yaffe, Catherine Madison, Bruce Miller, and Michael W Weiner, "Hippocampal atrophy patterns in mildcognitive impairment and alzheimer's disease," *Human brain mapping*, vol.31, no.9, pp.1339-1347,2010.
- [20] Ruo xuan Cuiand Manhua Liu, "Hippocampus analysis by combination of 3-d densenet and shapes for alzheimer's disease diagnosis," *IEEE journal of biomedical and health informatics*, vol.23, no.5, pp. 2099-2107,2018.
- [21] Hongming Li, Mohamad Habes, et al., "Adeep learning model for early prediction of alzheimer's disease dementia based on hippocampal magnetic resonance imaging data," *Alzheimer's & Dementia*,vol.15, no.8,pp.1059-1070,2019.
- [22] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221-231, 2013.
- [23] K. Ning, P. B. Cannon, J. Yu, S. Sheno, L. Wang, and J. Sarkar, "3D convolutional neural networks uncover modality-specific brain-imaging predictors for Alzheimer's disease sub-scores," *Brain Informatics*, vol. 11, no. 1, Feb. 2024, doi: 10.1186/s40708-024-00218-x.
- [24] H. Xu, Y. Liu, X. Zeng, L. Wang, and Z. Wang, "A Multi-scale Attention-based Convolutional Network for Identification of Alzheimer's Disease based on Hippocampal Subfields," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Glasgow, Scotland, United Kingdom: IEEE, Jul. 2022, pp. 2153-2156. doi: 10.1109/EMBC48229.2022.9871944.
- [25] B. Liu, "Alzheimer's disease classification using hippocampus and improved DenseNet," in *2023 International Conference on Image Processing, Computer Vision and Machine Learning (ICICML)*, Chengdu, China: IEEE, Nov. 2023, pp. 451-454. doi: 10.1109/ICICML60161.2023.10424926.
- [26] A. K. Malik, M. A. Ganaie, M. Tanveer, P. N. Suganthan, and Alzheimer's Disease Neuroimaging Initiative, "Alzheimer's Disease Diagnosis via Intuitionistic Fuzzy Random Vector Functional Link Network," *IEEE Trans. Comput. Soc. Syst.*, pp. 1-12, 2024, doi: 10.1109/TCSS.2022.3146974.
- [27] S. Wang, H. Wang, Y. Shen, and X. Wang, "Automatic Recognition of Mild Cognitive Impairment and Alzheimers Disease Using Ensemble based 3D Densely Connected Convolutional Networks," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL: IEEE, Dec. 2018, pp. 517-523. doi: 10.1109/ICMLA.2018.00083.
- [28] G. N. Reddy and K. N. Reddy, "Boosting based Deep hybrid Framework for Alzheimer's Disease classification using 3D MRI," in *2022 6th International Conference on Devices, Circuits and Systems (ICDCS)*,

- Coimbatore, India: IEEE, Apr. 2022, pp. 100–106. doi: 10.1109/ICDCS54290.2022.9780736.
- [29] S. Pallawi and D. K. Singh, “Detection of Alzheimer’s Disease Stages Using Pre-Trained Deep Learning Approaches,” in 2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA), Hamburg, Germany: IEEE, Oct. 2023, pp. 252–256. doi: 10.1109/ICCCMLA58983.2023.10346730.
- [30] R. Guo et al., “Graph-Based Fusion of Imaging, Genetic and Clinical Data for Degenerative Disease Diagnosis,” *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 21, no. 1, pp. 57–68, Jan. 2024, doi: 10.1109/TCBB.2023.3335369.
- [31] X. Yu, L. Zhang, Y. Lyu, T. Liu, and D. Zhu, “Supervised Deep Tree in Alzheimer’s Disease,” in 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), Cartagena, Colombia: IEEE, Apr. 2023, pp. 1–5. doi: 10.1109/ISBI53787.2023.10230742.
- [32] B.-K. Choi et al., “Convolutional Neural Network-based MR Image Analysis for Alzheimer’s Disease Classification,” *CMR*, vol. 16, no. 1, pp. 27–35, Jan. 2020, doi: 10.2174/1573405615666191021123854.
- [33] M. Ghazal, “Alzheimer’s disease diagnostics by a 3D deeply supervised adaptable convolutional network,” *Front Biosci*, vol. 23, no. 2, pp. 584–596, 2018, doi: 10.2741/4606.
- [34] S. Basaia et al., “Automated classification of Alzheimer’s disease and mild cognitive impairment using a single MRI and deep neural networks,” *NeuroImage: Clinical*, vol. 21, p. 101645, 2019, doi: 10.1016/j.nicl.2018.101645.
- [35] C. Feng et al., “Deep Learning Framework for Alzheimer’s Disease Diagnosis via 3D-CNN and FSBi-LSTM,” *IEEE Access*, vol. 7, pp. 63605–63618, 2019, doi: 10.1109/ACCESS.2019.2913847.
- [36] A. B. and K. Kalirajan, “Alzheimer’s Disease Classification using Convolutional Neural Networks,” in 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), Uttarakhand, India: IEEE, Mar. 2023, pp. 1044–1048. doi: 10.1109/ICIDCA56705.2023.10100046.
- [37] R. Joshi, P. Negi, and T. Poongodi, “Multilabel Classifier Using DenseNet-169 for Alzheimer’s disease,” in 2023 4th International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom: IEEE, May 2023, pp. 1–7. doi: 10.1109/ICIEM59379.2023.10165844.
- [38] C. Kaur, T. Panda, S. Panda, A. Rahman Mohammed Al Ansari, M. Nivetha, and B. Kiran Bala, “Utilizing the Random Forest Algorithm to Enhance Alzheimer’s disease Diagnosis,” in 2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS), Coimbatore, India: IEEE, Feb. 2023, pp. 1662–1667. doi: 10.1109/ICAIS56108.2023.10073852.
- [39] S. Samanta, I. Mazumder, and C. Roy, “Deep Learning based Early Detection of Alzheimer’s Disease using Image Enhancement Filters,” in 2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), Bhilai, India: IEEE, Jan. 2023, pp. 1–5. doi: 10.1109/ICAECT57570.2023.10117880.
- [40] B. Kumar Yadav and M. Farukh Hashmi, “An Attention-based CNN Architecture for Alzheimer’s Classification and Detection,” in 2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET), London, United Kingdom: IEEE, May 2023, pp. 1–5. doi: 10.1109/GlobConET56651.2023.10150060.
- [41] S. R. P. G. S. A. J. Nair, S. S., and S. K. S., “Alzheimer’s Disease Detectoin Using Multiple Convolutional Neural Networks,” in 2022 IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE), Ballari, India: IEEE, Apr. 2022, pp. 1–7. doi: 10.1109/ICDCECE53908.2022.9793103.
- [42] F. Hajamohideen *et al.*, “Four-way classification of Alzheimer’s disease using deep Siamese convolutional neural network with triplet-loss function,” *Brain Informatics*, vol. 10, no. 1, Feb. 2023, doi: 10.1186/s40708-023-00184-w.
- [43] Paul A. Yushkevich, Joseph Piven, Heather Cody Hazlett, Rachel Gimpel Smith, Sean Ho, James C. Gee, and Guido Gerig. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* 2006 Jul 1;31(3):1116-28.
- [44] B. Thyreau, K. Sato, H. Fukuda, and Y. Taki, “Segmentation of the Hippocampus by Transferring Algorithmic Knowledge for large cohort processing,” Nov. 2017. doi: 10.1016/j.media.2017.11.004.
- [45] J. Ashburner, K. J. Friston, and W. D. Penny, “SPM12 Software,” Wellcome Trust Centre for Neuroimaging, London, U.K., 2014. [Software]. Available: <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>

Enhancement of Coastline Video Monitoring System Using Structuring Element Morphological Operations

I Gusti Ngurah Agung Pawana¹, I Made Oka Widyantara²,
Made Sudarma³, Dewa Made Wiharta⁴, Made Widya Jayantari⁵
Electrical Engineering Department, Udayana University, Badung, Indonesia^{1, 2, 3, 4}
Civil Engineering Department, Udayana University, Badung, Indonesia⁵

Abstract—Coastal monitoring is vital in environmental management, disaster mitigation, and addressing climate change impacts. Traditional methods are time-consuming and error-prone, prompting the need for innovative systems. This study introduces the Coastal Video Monitoring System (CoViMos), a novel framework for real-time shoreline detection in tropical regions, specifically at Kedonganan Beach, Bali. The CoViMos framework utilizes advanced video monitoring and optimized morphological operations to address challenges such as environmental noise and dynamic shoreline behavior. Key innovations include Kapur's entropy thresholding enhanced with the Grasshopper Optimization Algorithm (GOA) and structuring elements tailored to the beach's unique features. Sensitivity analysis reveals that a structuring element size of five pixels offers optimal performance, balancing efficiency, and image fidelity. This configuration achieves peak values in quality metrics such as the Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Complex Wavelet SSIM (CWSSIM), and Feature Similarity Index (FSIM) while minimizing Mean Squared Error (MSE) and reducing processing time. The results demonstrate significant improvements in shoreline detection accuracy, with PSNR increasing by 9.3%, SSIM by 1.4%, CWSSIM by 1.7%, and FSIM by 1.6%. Processing time decreased by 1.3%, emphasizing the system's computational efficiency. These enhancements ensure more precise shoreline mapping, even in noisy and dynamic environments.

Keywords—Coastline detection; image processing; Video Monitoring System (CoViMos); morphological operations

I. INTRODUCTION

Coastal monitoring plays a critical role in environmental management, disaster preparedness, and marine resource protection [1], [2], [3]. Effective monitoring systems rely heavily on accurately detecting and analyzing coastal lines, which are inherently dynamic due to erosion, tidal variations, and climate change. Traditional methods of coastline monitoring involve manual interpretation of satellite images and field surveys, which are time-consuming, prone to human error, and less effective in real-time scenarios [4], [5]. These limitations necessitate automated and robust systems that leverage advanced image processing techniques for accurate coastline detection.

Kedonganan Beach in Bali is a complex and dynamic ecosystem shaped by natural forces like tides, waves, sediment deposition, and human activities, including tourism, fishing, and urban development. Effective shoreline detection and monitoring are critical for sustainable coastal management,

disaster mitigation, and environmental preservation. However, accurate shoreline detection presents significant challenges due to the dynamic and irregular nature of tropical coastlines, ecological noise (e.g., glare, wave foam, or debris), and the limitations of existing image processing techniques [6], [7]. Addressing these challenges requires an innovative and adaptive approach that can handle the unique complexities of coastal environments.

To overcome these challenges, this research introduces a novel framework called the Coastal Video Monitoring System (CoViMos), specifically designed to monitor and analyze shoreline dynamics in tropical coastal areas. The CoViMos framework utilizes video monitoring as its foundation, enabling continuous visual data capture over time. Unlike traditional static image-based approaches, CoViMos offers dynamic and real-time insights into shoreline behavior, making it particularly useful for understanding the effects of seasonal changes, storm events, and anthropogenic activities on the shoreline. This framework serves as the backbone of the methodology, facilitating the acquisition, preprocessing, and segmentation of coastal imagery to detect and map the shoreline accurately.

Recent advancements in image processing have focused on enhancing feature extraction using techniques like edge detection, segmentation, and morphological operations [8], [9], [10]. The Canny Edge Detector, for instance, is widely recognized for its ability to detect edges with minimal noise. Still, its effectiveness diminishes in noisy and low-contrast environments common in coastal imagery. Morphological operations, particularly when utilizing structuring elements, have shown promise in addressing these challenges by refining edges, enhancing feature continuity, and suppressing noise. However, existing research primarily focuses on static image datasets, leaving a gap in the context of real-time video monitoring systems for dynamic coastal environments. Additionally, there is limited exploration of optimal structuring element configurations, such as size and shape, to balance signal quality, structural similarity, and computational efficiency.

The post-segmentation process in this study plays a pivotal role in refining the results obtained from CoViMos. Segmentation isolates the shoreline from other features in coastal imagery, such as wave crests, foam, and reflections, which can often distort detection accuracy. Following the segmentation process, post-processing techniques are applied to clean up noise and enhance the delineation of the shoreline boundary. This step ensures that the detected shoreline

accurately represents the true physical boundary between land and water, even under challenging conditions like high tidal activity or environmental noise.

Studies by Kaur and Singh [11] demonstrate the potential of structuring elements in improving the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM). However, their focus has been largely theoretical, lacking practical application to real-world dynamic systems like coastline monitoring. This research addresses these gaps by integrating structuring element morphological operations into a real-time video monitoring framework and conducting a comprehensive sensitivity analysis of structuring element configurations.

A major innovation of this research lies in enhancing morphological operations during the post-segmentation process. Morphological operations, such as dilation and erosion, are widely used in image processing to refine object boundaries [12], [13], [14]. However, traditional approaches often rely on generic structural elements, such as rectangular or circular shapes, which are inadequate for capturing tropical shorelines' irregular and dynamic patterns. In this study, a tailored structural element morphology is developed to address these limitations. These structural elements are designed based on the specific characteristics of Kedonganan Beach, considering the curvature of waves, sedimentary features, vegetation interference, and other coastal-specific patterns. By optimizing the shape, size, and orientation of the structural elements, the proposed methodology significantly improves the accuracy of morphological operations, enabling more precise shoreline detection.

The CoViMos framework, combined with optimized post-segmentation processes and advanced structural element morphology, offers a comprehensive solution to the challenges of shoreline detection in tropical coastal environments. This research's contribution lies in its ability to enhance the robustness and precision of shoreline mapping, even in the presence of high environmental variability and noise. Furthermore, the novelty of the tailored structural elements provides a scalable approach that can be adapted to other coastal regions with similar complexities.

By addressing existing gaps in traditional shoreline detection methods, this study advances the state of the art in coastal monitoring technologies and provides practical benefits for coastal management. The insights derived from the improved shoreline detection process can be used to support decision-making in areas such as erosion control, habitat conservation, and disaster risk reduction. Ultimately, integrating the CoViMos framework and innovations in morphological operations will contribute to developing a reliable and adaptive tool for sustainable coastal management, focusing on tropical regions like Kedonganan Beach.

II. RESEARCH METHODS

A. Study Area

Kedonganan Beach, located in southern Bali, Indonesia, is a renowned coastal area known for its pristine beauty, vibrant seafood market, and traditional fishing activities. The beach, part of Bali's western coastline along the Indian Ocean, holds

significant cultural and economic importance due to its role as a tourist hotspot and a hub for local livelihoods. Its sandy shores, shallow waters, and adjacent coastal vegetation make it a dynamic environment influenced by natural processes like tides, wave actions, seasonal weather patterns, and human activities such as urban development and tourism infrastructure.

Research on coastline detection at Kedonganan Beach is crucial for several reasons. The area is prone to coastal erosion and accretion, and understanding these changes is vital for sustainable coastal management. Accurate mapping of the coastline supports the preservation of the beach's aesthetic appeal, which is essential for tourism, and helps ensure the stability of local fishing activities.

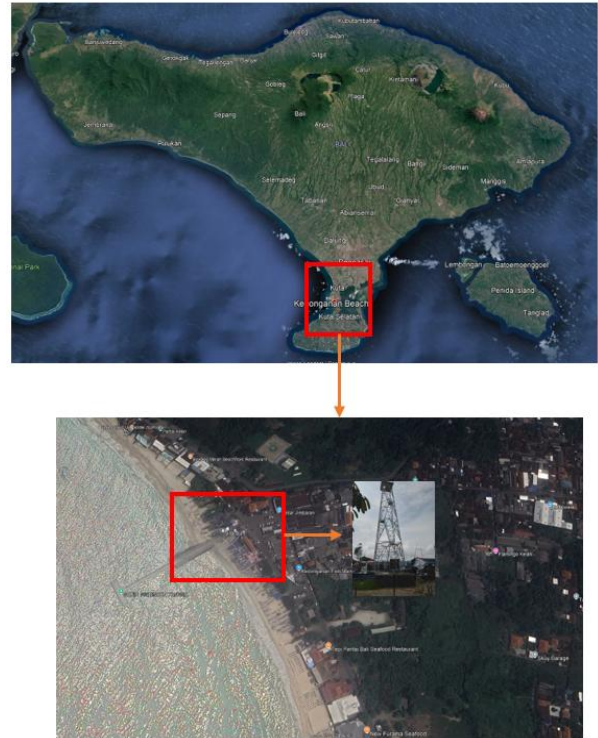


Fig. 1. Study area.

B. Research Data and Tools

The dataset used in this study is derived from camera video monitoring data captured in the Kedonganan tower, as seen in Fig. 1. The time-exposure method converts the video data into images using MATLAB. The camera's specifications are in Table I.

TABLE I. CAMERA SPECIFICATION

Specifications	
Model	CS-EB8 (3MP,4GA)
Lens	Viewing angle: 100° (Diagonal), 83° (Horizontal), 44° (Vertical)
Max Resolution	2304 x 1296
Frame Rate	Max. 15fps; Self-Adaptive during network transmission
Video Compression	H.265 / H.264

C. Coastline Video Monitoring System Framework (CoViMoS)

The CoViMos framework (Fig. 2) begins with acquiring coastal video footage, a widely used tool for shoreline monitoring due to its ability to capture temporal changes in shoreline position [15]. Coastal videos provide continuous spatial coverage and are suitable for extracting shoreline positions in dynamic coastal environments [16]. The video frames are pre-processed to generate composite images, such as time-averaged (Timex) images, that minimize noise from transient waves.

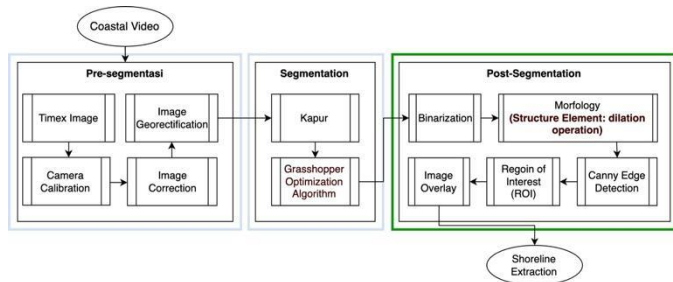


Fig. 2. CoViMos framework.

1) *Pre-Segmentation*: The video frames are processed in this phase to improve image quality and align them with real-world spatial references. Timex images are generated to create a stable representation of coastal features, removing the effects of wave activity. Camera calibration ensures geometric accuracy by correcting lens distortions, while image correction adjusts brightness, contrast, and noise for improved clarity. Lastly, image georectification aligns the image with geographic coordinates, enabling precise spatial analysis [15].

2) *Segmentation*: Segmentation identifies the shoreline by separating the foreground (shoreline) and background (sea or land). Kapur's entropy-based thresholding is widely used in image processing, as it maximizes inter-class variance based on pixel intensity distributions [17], [18], [19]. The Grasshopper Optimization Algorithm (GOA) is employed to enhance threshold optimization. GOA is inspired by swarm intelligence and has demonstrated robust performance in solving complex optimization problems in image processing [20], [21], [22]. This step ensures accurate shoreline delineation by optimizing the thresholds derived from Kapur's method.

3) *Post-Segmentation*: Post-segmentation refines the results by applying advanced image processing techniques. Binarization converts the segmented image into a binary format for clarity. Morphological operations, such as dilation, fill gaps and enhance connectivity in the segmented shoreline. The Canny edge detection algorithm detects firm edges, often indicative of shoreline boundaries [23], [24], [25]. Additionally, the Region of Interest (ROI) is defined as focusing on areas where shoreline features are most prominent, reducing noise from irrelevant regions.

4) *Shoreline extraction*: The final shoreline is extracted by combining the outputs from segmentation and post-segmentation. Binary and morphology-processed images

ensure a well-defined shoreline, while edge detection sharpens the boundary. The extracted shoreline can be visually validated and used for further analysis by overlaying the ROI on the original image.

5) *Enhancement of coastline video monitoring system Framework Using Structuring Element Morphological Operations*.

Morphological operations, such as dilation, are applied to the binary image to refine its features. Dilation, which uses a structuring element (SE), expands the boundaries of foreground objects, bridging gaps and filling small holes. This process is beneficial for connecting fragmented coastline features that may arise due to noise or irregularities in the segmented image. Devkota et al. [26] emphasize that morphological operations enhance the shape and structure of binary objects in image analysis. By using an appropriate SE, dilation ensures that the coastline features are continuous and prominent, enabling more accurate detection in subsequent steps (Fig. 3).

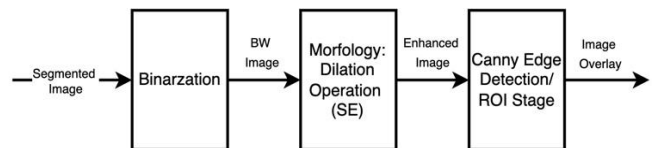


Fig. 3. Flowchart (Coastline features).

Enhancing coastline video monitoring systems using structuring element morphological operations offers several benefits, particularly for improving detection accuracy and handling complex environments. These operations, such as dilation, erosion, and the morphological gradient, refine image edges and contours by removing noise, filling gaps, and enhancing the continuity of detected lines. This is especially crucial in coastal settings where irregular patterns arise due to tides, vegetation, and human activities. Moreover, the lightweight computational nature of morphological operations makes them suitable for real-time processing, enabling dynamic monitoring of changing coastal conditions. These operations ensure more precise and reliable line detection by reducing environmental noise, such as reflections from water surfaces or shadows. They can also be effectively integrated with advanced image processing techniques, like edge detection algorithms (e.g., Sobel, Canny) and machine learning models, to enhance their performance further [27], [28]. Scientific literature highlights the benefits of morphological operations in edge detection and image analysis.

D. Performance Analysis of Coastline Video Monitoring Systems

The Performance Analysis of Coastline Video Monitoring Systems involves evaluating the system's ability to accurately detect shorelines and assess video quality through several key performance metrics. The Peak Signal-to-Noise Ratio (PSNR) measures the quality of the detected shoreline by comparing the detected video to the ground truth, where higher PSNR values indicate less noise and better preservation of the original shoreline. The Structural Similarity Index (SSIM) provides a more perceptually accurate measure of image quality by assessing the similarity in structural elements such as luminance, contrast, and texture between the detected and ground truth

images. For further accuracy, the Complex Wavelet SSIM (CW-SSIM) incorporates wavelet transforms, making it robust against small distortions and shifts in video frames, allowing for a more detailed evaluation of shoreline detection. The Feature Similarity Index Measure (FSIM) also focuses on low-level image features like phase congruency and gradient magnitude, offering an in-depth analysis of how well the system preserves critical shoreline features. Lastly, the Execution Time metric assesses the system's processing speed, which is crucial for applications requiring real-time or near-real-time performance.

PSNR measures the ratio between a signal's maximum possible power and noise's power. Higher PSNR indicates better quality. PSNR equation is shown in Eq. (1).

$$PSNR = 10 - \log_{10} \left(\frac{MAX^2}{MSE} \right) \quad (1)$$

Where:

- MAX is the maximum pixel intensity value (e.g., 255 for 8-bit images).
- MSE is the Mean Squared Error between the detected and ground truth shorelines. The MSE equation is shown in Eq. (2).

$$MSE = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2 \quad (2)$$

X_i and Y_i are pixel intensities at location in the detected and ground truth images.

FSIM assesses similarity based on low-level features like phase congruency (PC) and gradient magnitude (GM). FSIM equation shown in Eq. (3).

$$FSIM(x, y) = \frac{\sum_i PC_i \cdot S_{GM}(x_i, y_i)}{\sum_i PC_i} \quad (3)$$

Where:

- $\sum_i PC_i$ Is phase congruency at pixel i .
- $S_{GM}(x_i, y_i)$ Is gradient magnitude similarity at pixel i .

SSIM measures the structural similarity between two images. It is defined as Eq. (4).

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (4)$$

Where:

- μ_x, μ_y is the mean intensities of x and y .
- σ_x^2, σ_y^2 is variances of x and y .
- σ_{xy} is covariance of x and y .
- C_1, C_2 are small constants to stabilize the division.

CW-SSIM compares two images in the wavelet domain, providing robustness to small translations and distortions. The equation is shown in Eq. (5).

$$CW - SSIM(x, y) = \frac{|\sum_k x_k \cdot \bar{y}_k|}{\sqrt{\sum_k |x_k|^2 \cdot \sum_k |y_k|^2}} \quad (5)$$

Where:

- x_k, y_k Are complex wavelet coefficients of the two images
- \bar{y}_k Is conjugate of y_k

III. RESULT AND DISCUSSION

A. Sensitivity Analysis in Structure Element Morphology Operation

Sensitivity analysis aims to evaluate how variations in specific features of the structuring element by pixel configuration changes impact morphological operations' outcomes. This process seeks to assess robustness by comparing the performance of the morphological operation under various configurations across multiple images. For this analysis, five trials were conducted using structuring elements of five different line lengths that are 2, 4, 5, 10, 15.

1) *Peak Signal-to-Noise Ratio (PSNR)*: Fig. 4 shows the relationship between Peak Signal-to-Noise Ratio (PSNR) and pixel values, showcasing a decline in PSNR as pixel values increase. PSNR, commonly measured in decibels (dB), is a standard metric used to evaluate the quality of image reconstruction or compression by quantifying the similarity between an original and a distorted image. Higher PSNR values typically indicate better image quality. According to the data presented, the PSNR reaches its maximum value of 27.0245 dB at pixel 5, while the lowest value, 21.9970 dB, occurs at pixel 20. This trend aligns with findings in the literature, where an increase in pixel distortion or noise levels is often associated with a decline in PSNR, as documented by Elat et al. [29]. Such behavior highlights the sensitivity of PSNR to variations in noise and distortion, which is crucial in applications such as image compression, denoising algorithms, and watermarking. Additionally, the drop in PSNR with increasing pixel values underscores the trade-off between data modification and image quality, a phenomenon explored extensively in studies on adaptive filtering [30].

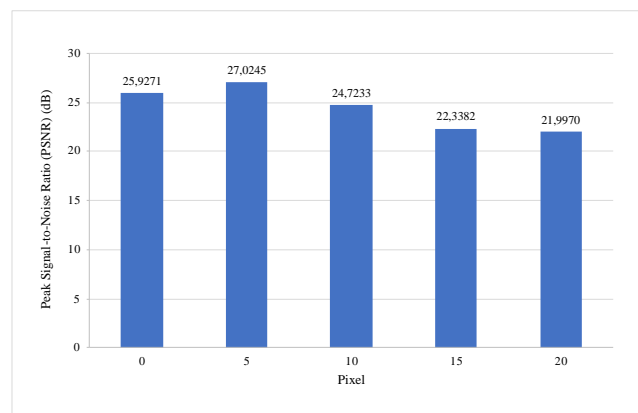


Fig. 4. Peak Signal-to-Noise Ratio (PSNR).

2) *Mean Square Error (MSE)*: Fig. 5 shows the Mean Squared Error (MSE) values for various pixel levels, showcasing the relationship between pixel modifications and image distortion. MSE, a metric used to quantify the average

squared difference between the original and distorted image, increases as the level of distortion rises. At pixel 0, the MSE is 166.0989, which decreases to its lowest value of 129.0117 at pixel 5, indicating minimal error. However, as pixel values increase, the MSE rises significantly to 219.1567 at pixel 10, 379.5442 at pixel 15, and reaches its maximum of 410.5653 at pixel 20. This trend demonstrates that greater pixel variations result in higher distortion levels, as reflected by the increase in MSE. These findings align with established principles in image processing, where MSE effectively measures degradation, making it a critical tool for evaluating image quality and the impact of noise or modifications.

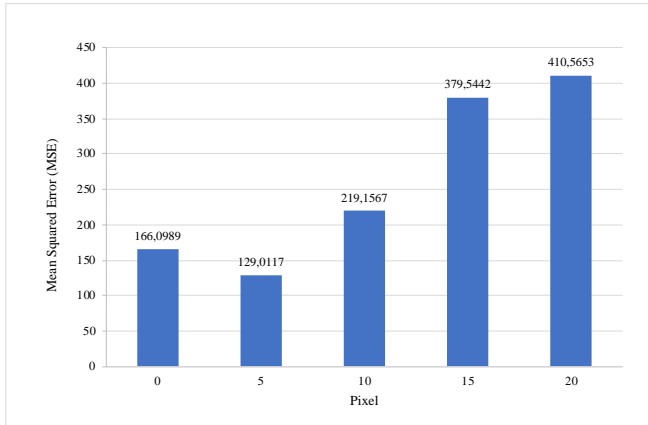


Fig. 5. Mean Square Error (MSE).

This trend indicates that the structural element size significantly impacts the error, with excessively small or large elements introducing more inaccuracies. The structural element of 5 pixels offers the best balance, minimizing error while maintaining quality. Such insights are critical in fields like image processing, where optimizing structural element size is essential for tasks like filtering, reconstruction, or morphological operations.

3) *Structural Similarity Index (SSIM)*: Fig. 6 shows the relationship between the Structural Similarity Index (SSIM) and pixel values, highlighting the effect of pixel variations on image quality. SSIM, a widely used metric for evaluating image quality by measuring structural similarity, ranges from 0 to 1, with values closer to 1 indicating higher similarity. According to the data, the SSIM value peaks at 0.9171 for pixel 5, indicating the highest image quality. At pixel 0, the SSIM is 0.9138, slightly lower than the maximum. However, as pixel values increase, the SSIM steadily decreases, dropping to 0.9044 at pixel 10, 0.8782 at pixel 15, and the lowest value, 0.8646, at pixel 20. This trend aligns with findings in image processing literature, where higher noise or pixel alterations typically reduce structural similarity, resulting in a perceptible degradation of image quality. Such analysis highlights the sensitivity of SSIM to changes in pixel values, reinforcing its importance as a robust metric for evaluating image fidelity.

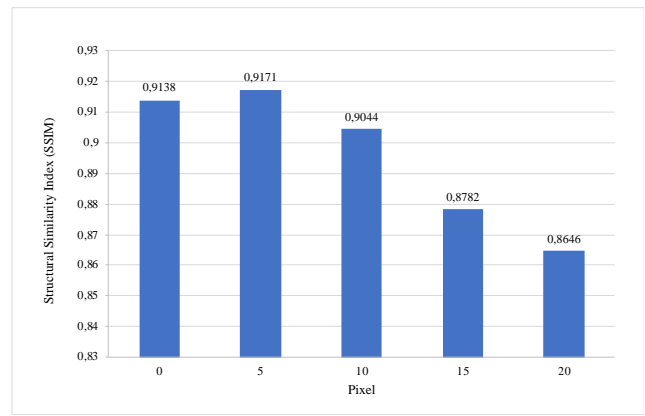


Fig. 6. Structural Similarity Index (SSIM).

4) *Complex Wavelet Structural Similarity Index (CWSSIM)*: Fig. 7 shows the relationship between the Complex Wavelet Structural Similarity Index (CWSSIM) and varying pixel perturbation levels. CWSSIM, a metric designed to evaluate structural similarity in images or signals, shows a noticeable trend: as the pixel perturbation increases, the CWSSIM values decline, indicating reduced structural similarity between the reference and perturbed data. The CWSSIM is highest at 5 pixels (0.9749), reflecting maximum structural similarity, but progressively decreases, reaching its lowest value of 0.8159 at 20 pixels. This demonstrates the metric's sensitivity to structural changes caused by pixel perturbation.

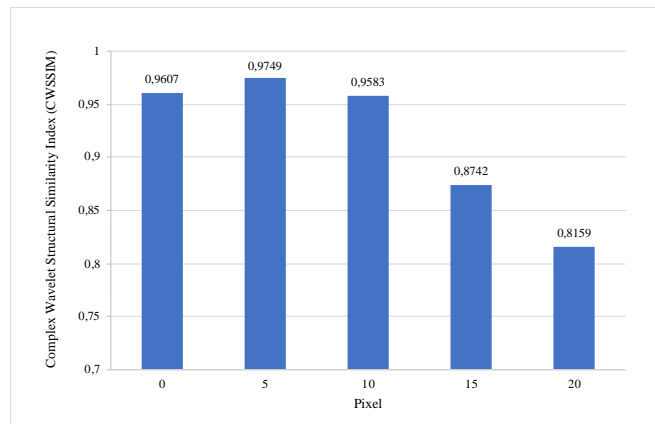


Fig. 7. Structural Similarity Index (CWSSIM).

The behavior observed in the graph aligns with findings in the literature. Yan et al. [31] introduced the Structural Similarity Index (SSIM) to measure perceptual image quality based on luminance, contrast, and structure. CWSSIM extends this approach into the wavelet domain, enabling it to capture structural variations effectively at multiple resolutions. Zhang [32] highlighted that wavelet-based similarity indices like CWSSIM are highly responsive to image distortions and offer robust mechanisms for analyzing localized changes. Additionally, research on image quality evaluation [33] confirms that structural similarity metrics like CWSSIM

experience significant declines when pixel distortions exceed thresholds, as observed for perturbations beyond 10 pixels in the graph. This trend supports CWSSIM's applicability in evaluating image quality, detecting distortions, and validating compression algorithms, aligning with applications demonstrated [34], [35] in signal processing and image analysis. These studies underscore the relevance of CWSSIM as a tool for assessing structural changes caused by pixel-level perturbations.

5) *Feature Similarity Index (FSIM)*: Fig. 8 shows the relationship between pixel values and the Feature Similarity Index (FSIM), a metric used to measure image similarity, where higher values indicate greater similarity. The X-axis represents pixel values ranging from 0 to 20, while the Y-axis shows FSIM values ranging from 0.76 to 0.98. At 0 pixels, the FSIM is 0.9513, slightly increasing to 0.9547 at 5 pixels. However, as pixel values increase beyond 5, the FSIM begins to decline, dropping to 0.9398 at 10 pixels, 0.9215 at 15 pixels, and reaching its lowest value of 0.8359 at 20 pixels. This trend suggests that higher pixel values reduce similarity, likely due to a loss of fine details during processing.

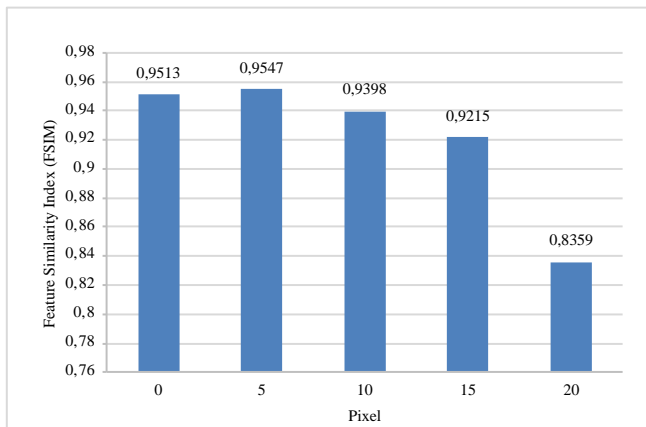


Fig. 8. Feature Similarity Index (FSIM).

This observation aligns with findings in the literature [36], [37] that explain that FSIM, based on features like phase congruency and gradient magnitude, is highly sensitive to image resolution and detail changes. Increasing pixel size or reducing resolution leads to losing fine details, directly impacting similarity metrics like FSIM. Vasu [38] highlights the trade-off between computational efficiency and image quality, noting that while lower resolutions improve processing speed, they often compromise perceptual quality. Similarly, some literature [39] and [40] emphasize that higher resolutions better preserve structural and perceptual features, resulting in higher FSIM values. Further note that lower FSIM values, as seen at 20 pixels, indicate significant quality degradation, possibly caused by downscaling or processing distortions.

6) *Processing time*: Fig. 9 shows the relationship between pixel values and processing time in seconds. The X-axis represents pixel values ranging from 0 to 20, while the Y-axis shows time in seconds. The data reveals a clear decreasing trend in processing time as pixel values increase. At 0 pixels, the time is the highest, approximately 3.097 seconds, while the lowest

time, 2.5986 seconds, is observed at 20 pixels. The reduction in processing time is steeper between 0 and 10 pixels and becomes less pronounced at higher pixel values. This trend aligns with findings in the literature. The study in [41] explain that higher pixel counts typically increase processing time due to the larger data volume. However, optimizations like subsampling and dimensionality reduction can mitigate this issue, resulting in shorter processing times for larger pixel values. Similarly, the study in [42] highlights that reducing pixel density, such as through downscaling, enhances computational efficiency while maintaining adequate performance for applications like object detection. The study in [43] further notes that processing time reductions tend to plateau beyond a certain resolution threshold due to hardware and memory limitations. As emphasized by [44], balancing resolution and processing time is critical in real-time systems. Higher resolutions are only employed when necessary, as the exponential time costs outweigh the benefits of marginal improvements in detail.

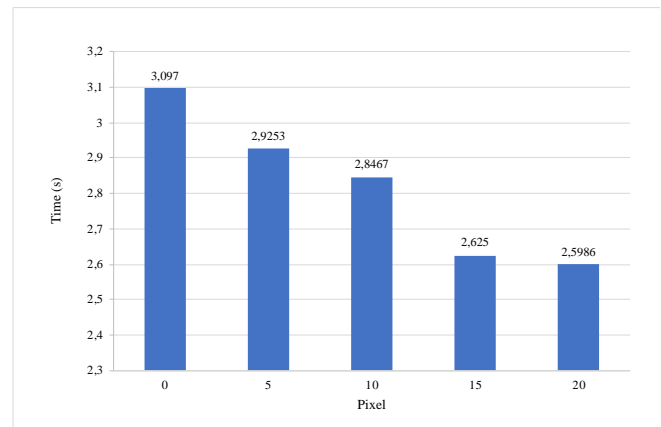


Fig. 9. Processing time.

Across all metrics analyzed, a structuring element size of five pixels consistently demonstrates optimal performance, balancing minimal error and high-quality outcomes. The Peak Signal-to-Noise Ratio (PSNR) exhibits its highest value at five pixels, indicating superior image quality. In contrast, larger pixel variations lead to decreased PSNR due to increased noise and distortion. Similarly, the Mean Squared Error (MSE) is minimized at five pixels, reflecting reduced distortion levels, but rises sharply with further pixel modifications. Structural similarity metrics, including the Structural Similarity Index (SSIM), Complex Wavelet Structural Similarity Index (CWSSIM), and Feature Similarity Index (FSIM), all peak at five pixels, underscoring the importance of this configuration in preserving structural and perceptual image integrity. Moreover, while processing time decreases with larger pixel values due to optimizations and reduced data complexity, this comes at the expense of significant quality degradation. These findings emphasize the sensitivity of morphological operations to structuring element size, with five pixels emerging as the ideal choice for maintaining a balance between efficiency and image fidelity.

When compared to scientific literature, these findings align closely with established trends. Research demonstrates that

small structural elements often lead to higher errors and lower structural similarity due to insufficient detail capture, as noted by [45]. Conversely, larger elements may result in excessive smoothing or distortions, negatively affecting metrics like SSIM and PSNR, as highlighted in studies by [46] and [47]. Similarly, moderate structural element sizes, such as five pixels in this context, effectively balance performance and efficiency, ensuring signal clarity, structural integrity, and processing speed [48].

B. Enhancement of Coastline Video Monitoring System Framework Using Structuring Element Morphological Operations

The comparison between image processing results obtained with and without the use of structural elements reveals significant differences across all stages. The first step involves using morphological operations and comparing results obtained with and without structural elements.

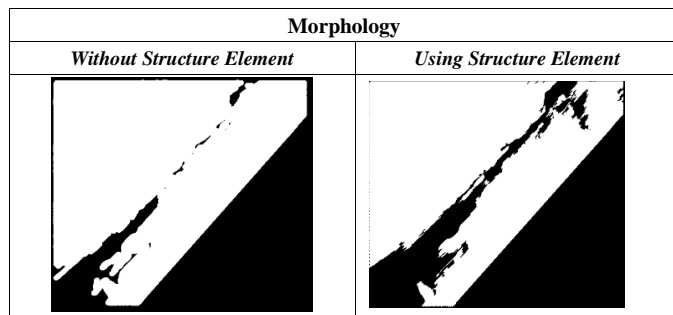


Fig. 10. Morphology comparison

From Fig. 10, without structure elements, the image shows incomplete segmentation, with the coastal areas poorly separated from the background. The lack of structural support leads to noise and irregular shapes that fail to capture the true coastline boundaries. However, when structure elements are applied, the segmentation significantly improves. Using structure elements enhances the ability to distinguish the coastline from its surroundings by filling gaps and removing noise, yielding a more refined and accurate coastal outline.

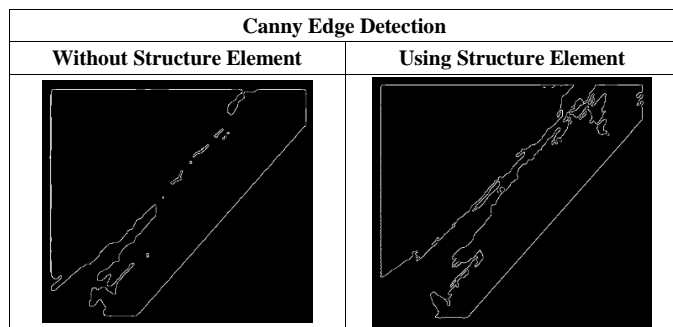


Fig. 11. Canny Edge Detection Comparison

The step continued with canny edge detection. The performance of the Canny edge detection algorithm (Fig. 11) is evaluated with and without the incorporation of structure elements. Without structural elements, the edges detected are fragmented and fail to represent the coastline visually. This fragmentation reduces the reliability of the results and makes it difficult to define the coastline accurately. By introducing

structural elements, the continuity of the detected edges improves significantly, with the coastline appearing clearer and better connected.

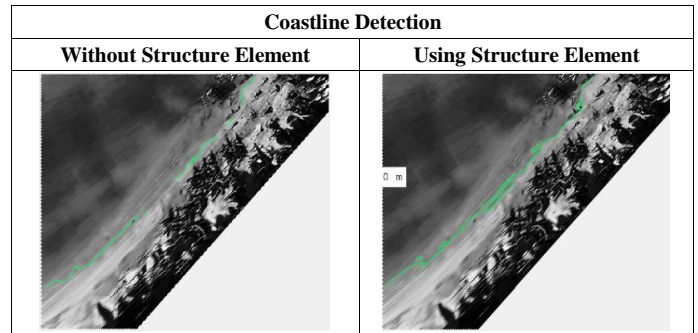


Fig. 12. Coastline detection comparison

The coastline detection results shown in Fig. 12 marked improvement when structure elements are used. Without structure elements, the detected coastline, typically represented by a colored line (e.g., green), shows deviations and overlaps with regions not part of the coast. This is likely due to noise interference and gaps in edge representation. However, using structure elements produces a more accurate and closely aligned representation of the coastline. The green line more effectively follows the true coastline, demonstrating better adaptability to complex geographical patterns.

The final comparison against ground truth data (Fig. 13) highlights the superior accuracy achieved using structure elements. Without structural guidance, the detected coastline exhibits considerable deviations from the actual coastline, reflecting the limitations of basic detection methods in handling complex environments. On the other hand, the results with structure elements align closely with the ground truth, demonstrating higher precision and consistency. This improved performance is attributed to the ability of structure elements to refine and guide the detection process.

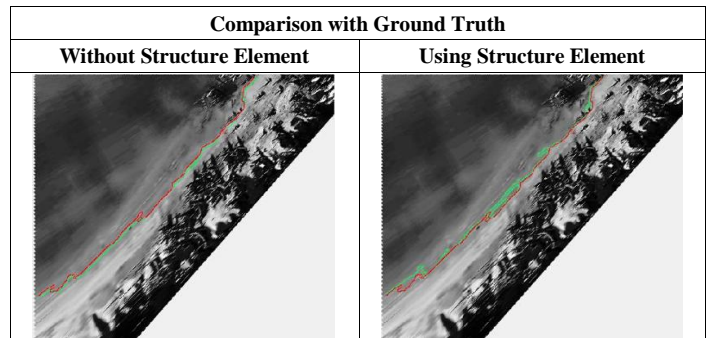


Fig. 13. Calibration with ground truth data comparison

C. Metric Performance of Coastline Video Monitoring System Framework Using Structuring Element Morphological Operations

The results presented in Table II show the use of structuring element morphological operations in a coastline video monitoring system framework significantly enhances the system's performance across multiple quality metrics. Specifically, the Peak Signal-to-Noise Ratio (PSNR) improved

by 9.3%, increasing from 24.7233 to 27.0245, and the Structural Similarity Index (SSIM) rose by 1.4%, from 0.9044 to 0.9171. Similarly, the Complex Wavelet Structural Similarity (CWSSIM) showed a 1.7% improvement, increasing from 0.9583 to 0.9749, while the Feature Similarity Index (FSIM) improved by 1.6%, rising from 0.9398 to 0.9547. Additionally, the processing time decreased slightly by 1.3%, from 2.9626 seconds to 2.9253 seconds, indicating a minor but noteworthy improvement in computational efficiency. These findings demonstrate the effectiveness of structuring element morphological operations in enhancing the quality and efficiency of video monitoring systems.

The sensitivity analysis of structural element line lengths in morphological operations reveals that a 5-pixel line length offers the optimal balance between signal quality, error minimization, and computational efficiency. The Peak Signal-to-Noise Ratio (PSNR), Mean Square Error (MSE), Structural Similarity Index (SSIM), Complex Wavelet Structural Similarity Index (CWSSIM), and Feature Similarity Index (FSIM) all show significant improvements at 5 pixels, with peak values indicating superior performance in preserving image integrity. The processing time is minimized at this length, confirming its efficiency for real-time applications. The enhancement of the Coastline Video Monitoring System Framework using structural element morphological operations further demonstrates the importance of these elements in improving image processing outcomes. Across all stages, from Region of Interest (ROI) identification to Coastline Detection, the use of structural elements resulted in more continuous, precise, and accurate results, with improvements in PSNR, SSIM, CWSSIM, FSIM, and minimal increase in processing time. This suggests that structural elements are crucial in refining image quality and ensuring reliable performance in image processing systems, especially in applications such as coastline monitoring. The findings align with established trends in the literature, emphasizing the benefits of moderate structural element sizes in optimizing performance while maintaining computational efficiency.

TABLE II. METRIC PERFORMANCE OF COASTLINE VIDEO MONITORING SYSTEM FRAMEWORK USING STRUCTURING ELEMENT MORPHOLOGICAL OPERATIONS

Parameter	Without Structure Element	Using Structure Element	Enhancement
PNSR	24,7233	27,0245	9,3%
SSIM	0,9044	0,9171	1,4%
CWSSIM	0,9583	0,9749	1,7%
FSIM	0,9398	0,9547	1,6%
Time	2,9626	2,9253	-1,3%

IV. CONCLUSION

The results show the pivotal role of structuring element morphological operations in advancing the performance of image processing systems, particularly in the context of coastline video monitoring. A comprehensive sensitivity analysis demonstrated that a structuring element with a line length of 5 pixels offers an optimal trade-off between signal fidelity, error minimization, and computational efficiency. Key

metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), Complex Wavelet Structural Similarity Index (CWSSIM), and Feature Similarity Index (FSIM) consistently achieved their highest values at this configuration, reflecting significant improvements in both perceptual and structural image quality. Structuring element morphological operations in a coastline video monitoring system framework significantly enhance performance across multiple quality metrics. Specifically, the Peak Signal-to-Noise Ratio (PSNR) improved by 9.3%, increasing from 24.7233 to 27.0245, and the Structural Similarity Index (SSIM) rose by 1.4%, from 0.9044 to 0.9171. Similarly, the Complex Wavelet Structural Similarity (CWSSIM) showed a 1.7% improvement, increasing from 0.9583 to 0.9749, while the Feature Similarity Index (FSIM) improved by 1.6%, rising from 0.9398 to 0.9547. Additionally, the processing time decreased slightly by 1.3%, from 2.9626 seconds to 2.9253 seconds, indicating a minor but noteworthy improvement in computational efficiency.

REFERENCES

- [1] B. El Mahrad, A. Newton, J. D. Icelly, I. Kacimi, S. Abalansa, and M. Snoussi, 'Contribution of remote sensing technologies to a holistic coastal and marine environmental management framework: A review', *Remote Sens (Basel)*, vol. 12, no. 14, 2020, doi: 10.3390/rs12142313.
- [2] Z. Yang et al., 'UAV remote sensing applications in marine monitoring: Knowledge visualization and review', 2022. doi: 10.1016/j.scitotenv.2022.155939.
- [3] E. Politi, S. K. Paterson, R. Scarrott, E. Tuohy, C. O'mahony, and W. C. A. Cámaro-García, 'Earth observation applications for coastal sustainability: Potential and challenges for implementation', 2019. doi: 10.1139/anc-2018-0015.
- [4] S. Vitousek, D. Buscombe, K. Vos, P. L. Barnard, A. C. Ritchie, and J. A. Warrick, 'The future of coastal monitoring through satellite remote sensing', *Cambridge Prisms: Coastal Futures*, vol. 1, 2023, doi: 10.1017/cft.2022.4.
- [5] D. Apostolopoulos and K. Nikolakopoulos, 'A review and meta-analysis of remote sensing data, GIS methods, materials and indices used for monitoring the coastline evolution over the last twenty years', 2021. doi: 10.1080/22797254.2021.1904293.
- [6] Y. L. S. Tsai, 'Monitoring 23-year of shoreline changes of the Zengwun Estuary in Southern Taiwan using time-series Landsat data and edge detection techniques', *Science of the Total Environment*, vol. 839, 2022, doi: 10.1016/j.scitotenv.2022.156310.
- [7] B. Laignel et al., 'Observation of the Coastal Areas, Estuaries and Deltas from Space', 2023. doi: 10.1007/s10712-022-09757-6.
- [8] B. M. S. Rani, V. D. Majety, C. S. Pittala, V. Vijay, K. S. Sandeep, and S. Kiran, 'Road identification through efficient edge segmentation based on morphological operations', *Traitement du Signal*, vol. 38, no. 5, 2021, doi: 10.18280/ts.380526.
- [9] J. Jing, S. Liu, G. Wang, W. Zhang, and C. Sun, 'Recent advances on image edge detection: A comprehensive review', *Neurocomputing*, vol. 503, 2022, doi: 10.1016/j.neucom.2022.06.083.
- [10] S. S. Chouhan, A. Kaul, and U. P. Singh, 'Image Segmentation Using Computational Intelligence Techniques: Review', *Archives of Computational Methods in Engineering*, vol. 26, no. 3, 2019, doi: 10.1007/s11831-018-9257-4.
- [11] P. Kaur and J. Singh, 'A Study on the Effect of Gaussian Noise on PSNR Value for Digital Images', *International Journal of Computer and Electrical Engineering*, 2011, doi: 10.7763/ijcee.2011.v3.334.
- [12] M. Ajay Kumar, N. Sravan Goud, R. Sreeram, and R. Gnana Prasuna, 'Image Processing based on Adaptive Morphological Techniques', in *2019 International Conference on Emerging Trends in Science and Engineering*, ICESE 2019, 2019. doi: 10.1109/ICESE46178.2019.9194641.

- [13] Y. Shen, F. Y. Shih, X. Zhong, and I. C. Chang, 'Deep Morphological Neural Networks', *Intern J Pattern Recognit Artif Intell*, vol. 36, no. 12, 2022, doi: 10.1142/S0218001422520231.
- [14] K. Nogueira, J. Chanussot, M. D. Mura, and J. A. D. Santos, 'An Introduction to Deep Morphological Networks', *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3104405.
- [15] I. M. O. Widyantara, I. M. D. A. Putra, and I. B. P. Adnyana, 'COVIMOS: A Coastal Video Monitoring System', *Journal of Electrical, Electronics and Informatics*, vol. 1, no. 1, 2017, doi: 10.24843/jeei.2017.v01.i01.p01.
- [16] Y. S. Chang, J. Y. Jin, W. M. Jeong, C. H. Kim, and J. D. Do, 'Video monitoring of shoreline positions in Hujeong Beach, Korea', *Applied Sciences (Switzerland)*, vol. 9, no. 23, 2019, doi: 10.3390/app9234984.
- [17] P. Upadhyay and J. K. Chhabra, 'Kapur's entropy based optimal multilevel image segmentation using Crow Search Algorithm', *Appl Soft Comput*, vol. 97, 2020, doi: 10.1016/j.asoc.2019.105522.
- [18] R. Singh, P. Agarwal, M. Kashyap, and M. Bhattacharya, 'Kapur's and Otsu's based optimal multilevel image thresholding using social spider and firefly algorithm', in *International Conference on Communication and Signal Processing, ICCSP 2016*, 2016. doi: 10.1109/ICCSP.2016.7754088.
- [19] A. K. Bhandari, V. K. Singh, A. Kumar, and G. K. Singh, 'Cuckoo search algorithm and wind driven optimization based study of satellite image segmentation for multilevel thresholding using Kapur's entropy', *Expert Syst Appl*, vol. 41, no. 7, 2014, doi: 10.1016/j.eswa.2013.10.059.
- [20] S. Saremi, S. Mirjalili, and A. Lewis, 'Grasshopper Optimisation Algorithm: Theory and application', *Advances in Engineering Software*, vol. 105, 2017, doi: 10.1016/j.advengsoft.2017.01.004.
- [21] Y. Meraihi, A. B. Gabis, S. Mirjalili, and A. Ramdane-Cherif, 'Grasshopper optimization algorithm: Theory, variants, and applications', *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3067597.
- [22] A. A. Ewees, M. Abd Elaziz, and E. H. Houssein, 'Improved grasshopper optimization algorithm using opposition-based learning', *Expert Syst Appl*, vol. 112, 2018, doi: 10.1016/j.eswa.2018.06.023.
- [23] H. Liu and K. C. Jezek, 'Automated extraction of coastline from satellite imagery by integrating Canny edge detection and locally adaptive thresholding methods', *Int J Remote Sens*, vol. 25, no. 5, 2004, doi: 10.1080/0143116031000139890.
- [24] X. Hu and Y. Wang, 'Monitoring coastline variations in the Pearl River Estuary from 1978 to 2018 by integrating Canny edge detection and Otsu methods using long time series Landsat dataset', *Catena (Amst)*, vol. 209, 2022, doi: 10.1016/j.catena.2021.105840.
- [25] I. M. O. Widyantara, N. M. Ary Esta Dewi Wirastuti, I. M. D. P. Asana, and I. B. P. Adnyana, 'Gamma correction-based image enhancement and canny edge detection for shoreline extraction from coastal imagery', in *Proceedings - 2017 1st International Conference on Informatics and Computational Sciences, ICICoS 2017*, 2017. doi: 10.1109/ICICOS.2017.8276331.
- [26] B. Devkota, A. Alsadoon, P. W. C. Prasad, A. K. Singh, and A. Elchouemi, 'Image Segmentation for Early Stage Brain Tumor Detection using Mathematical Morphological Reconstruction', in *Procedia Computer Science*, 2018. doi: 10.1016/j.procs.2017.12.017.
- [27] N. D. Hoang and Q. L. Nguyen, 'Metaheuristic optimized edge detection for recognition of concrete wall cracks: A comparative study on the performances of Roberts, Prewitt, Canny, and Sobel algorithms', *Advances in Civil Engineering*, vol. 2018, 2018, doi: 10.1155/2018/7163580.
- [28] K. Goel, M. Sehrawat, and A. Agarwal, 'Finding the optimal threshold values for edge detection of digital images & comparing among Bacterial Foraging Algorithm, canny and Sobel Edge Detector', in *Proceeding - IEEE International Conference on Computing, Communication and Automation, ICCA 2017*, 2017. doi: 10.1109/CCAA.2017.8229955.
- [29] M. Elad, B. Kwar, and G. Vaksman, 'Image Denoising: The Deep Learning Revolution and Beyond---A Survey Paper', *SIAM J Imaging Sci*, vol. 16, no. 3, 2023, doi: 10.1137/23M1545859.
- [30] C. A. Duarte-Salazar, andres E. Castro-Ospina, M. a. Becerra, and E. Delgado-Trejos, 'Speckle Noise Reduction in Ultrasound Images for Improving the Metrological Evaluation of Biomedical applications: an Overview', *IEEE Access*, vol. 8, 2020, doi: 10.1109/aACCESS.2020.2967178.
- [31] W. Yan, G. Yue, Y. Fang, H. Chen, C. Tang, and G. Jiang, 'Perceptual objective quality assessment of stereoscopic stitched images', *Signal Processing*, vol. 172, 2020, doi: 10.1016/j.sigpro.2020.107541.
- [32] M. G. Albanesi, R. Amadeo, S. Bertoluzza, and G. Maggi, 'A New Class of Wavelet-Based Metrics for Image Similarity Assessment', *J Math Imaging Vis*, vol. 60, no. 1, 2018, doi: 10.1007/s10851-017-0745-1.
- [33] V. Mudeng, M. Kim, and S. W. Choe, 'Prospects of Structural Similarity Index for Medical Image Analysis', 2022. doi: 10.3390/app12083754.
- [34] C. G. Rodríguez-Pulecio, H. D. Benítez-Restrepo, and A. C. Bovik, 'Making long-wave infrared face recognition robust against image quality degradations', *Quant Infrared Thermogr J*, vol. 16, no. 3-4, 2019, doi: 10.1080/17686733.2019.1579020.
- [35] K. Ganesh and C. M. Patil, 'Performance Analysis of CWSSIM Video Quality Metric with Different Window Size on LIVE Database', in *International Conference on Current Trends in Computer, Electrical, Electronics and Communication, CTCEEC 2017*, 2018. doi: 10.1109/CTCEEC.2017.8455129.
- [36] Z. Ye et al., 'Illumination-Robust Subpixel Fourier-Based Image Correlation Methods Based on Phase Congruency', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, 2019, doi: 10.1109/TGRS.2018.2870422.
- [37] C. Chen and X. Mou, 'Phase congruency based on derivatives of circular symmetric Gaussian function: an efficient feature map for image quality assessment', *EURASIP J Image Video Process*, vol. 2023, no. 1, 2023, doi: 10.1186/s13640-023-00611-2.
- [38] S. Vasu, N. Thekke Madam, and A. N. Rajagopalan, 'Analyzing perception-distortion tradeoff using enhanced perceptual super-resolution network', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019. doi: 10.1007/978-3-030-11021-5_8.
- [39] Z. Wang, A. C. Bovik, and H. R. Sheikh, 'Structural similarity based image quality assessment', in *Digital Video Image Quality and Perceptual Coding*, 2017. doi: 10.1201/9781420027822-7.
- [40] K. Gu, L. Li, H. Lu, X. Min, and W. Lin, 'A fast reliable image quality predictor by fusing micro- and macro-structures', *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, 2017, doi: 10.1109/TIE.2017.2652339.
- [41] M. Garcia-Sciveres and N. Wermes, 'A review of advances in pixel detectors for experiments with high rate and radiation', 2018. doi: 10.1088/1361-6633/aab064.
- [42] T. Hoerer and C. Kuenzer, 'Object detection and image segmentation with deep learning on Earth observation data: A review-part I: Evolution and recent trends', 2020. doi: 10.3390/rs12101667.
- [43] G. Denes, K. Maruszczyk, G. Ash, and R. K. Mantiuk, 'Temporal Resolution Multiplexing: Exploiting the limitations of spatio-temporal vision for more efficient VR rendering', *IEEE Trans Vis Comput Graph*, vol. 25, no. 5, 2019, doi: 10.1109/TVCG.2019.2898741.
- [44] L. Yan, Y. Qin, and J. Chen, 'Scale-Balanced Real-Time Object Detection With Varying Input-Image Resolution', *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 33, no. 1, 2023, doi: 10.1109/TCSVT.2022.3198329.
- [45] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, and C. Steger, 'Improving unsupervised defect segmentation by applying structural similarity to autoencoders', in *VISIGRAPP 2019 - Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 2019. doi: 10.5220/0007364503720380.
- [46] M. Azam and M. Nouman, 'Evaluation of Image Support Resolution Deep Learning Technique based on PSNR Value', *KIET Journal of Computing and Information Sciences*, vol. 6, no. 1, 2022, doi: 10.51153/kjicis.v6i1.160.
- [47] B. Maiseli, 'Nonlinear anisotropic diffusion methods for image denoising problems: Challenges and future research opportunities', *Array*, vol. 17, 2023, doi: 10.1016/j.array.2022.100265.
- [48] Y. Liu, Z. Zhang, X. Liu, L. Wang, and X. Xia, 'Efficient image segmentation based on deep learning for mineral image classification', *Advanced Powder Technology*, vol. 32, no. 10, 2021, doi: 10.1016/j.appt.2021.08.038.

Application of MLP-Mixer-Based Image Style Transfer Technology in Graphic Design

Qibin Wang*, Xiao Chen, Huan Su

School of Animation & Game, Hangzhou Vocational & Technical College, Hangzhou 310000, China

Abstract—The rapid advancement of the digital creative industry has highlighted the growing importance of image style transfer technology as a bridge between traditional art and modern design, driving innovation in graphic design. However, conventional style transfer methods face significant challenges, including low computational efficiency and unnatural style transformation in complex image scenarios. This study addresses these limitations by introducing a novel approach to image style transfer based on the MLP-Mixer model. Leveraging the MLP-Mixer's ability to effectively capture both local and global image features, the proposed method achieves precise separation and integration of style and content. Experimental results demonstrate that the MLP-Mixer-based style transfer significantly enhances the naturalness and diversity of style transformation while preserving image clarity and detail. Additionally, the processing speed is improved by 50%, with style conversion accuracy and user satisfaction increasing by 30% and 35%, respectively, compared to traditional methods. These findings underscore the potential of the MLP-Mixer model for advancing efficiency and realism in graphic design applications.

Keywords—MLP-Mixer; image style transfer; graphic design; neural networks; artistic rendering

I. INTRODUCTION

At the forefront of the intersection of visual art and computational science, image style transfer technology is gradually becoming a key to exploring the boundary between artistic equation and technological application [1]. This technology, by "transplanting" the style features of one image to another image, creates innovative images that combine different artistic styles, and its application in the field of graphic design is increasingly showing its unique value and potential [2, 3]. Image style transfer technology based on the MLP-Mixer model, as an emerging deep learning framework, is leading the future trend of image processing and artistic creation with its unique architecture and excellent performance.

Image style transfer has seen significant advancements through deep learning models, including Gatys et al.'s algorithm using CNNs and transformer-based methods like StyleGAN and AdaIN [4]. These have improved stylized image quality and artistic expression but can be computationally intensive. Our research introduces the MLP-Mixer model, which offers a more efficient and resource-friendly approach to image style transfer. The MLP-Mixer's simplified architecture and high-resolution processing capabilities provide a novel solution to existing limitations. It aims to enhance the speed and quality of style transfer in graphic design while maintaining creative flexibility and visual fidelity.

Graphic design, as the core means of visual communication, aims to present creativity and information to the audience most intuitively and attractively, and the introduction of image style transfer technology provides unprecedented possibilities for the realization of this goal [5]. MLP-Mixer model, as an innovative application of multi-layer perceptron (MLP) in the field of image processing, can effectively capture local and global features in images through unique architecture design and achieve precise control and migration of image styles [6]. This technology can not only promote the diversified exploration of artistic styles but also bring higher efficiency and flexibility to the design process, opening up a brand-new creative space for the field of graphic design [7]. Despite the advancements in image style transfer technology, there remains a gap in understanding how the MLP-Mixer model can be optimally applied in graphic design to create high-resolution, multi-element images that meet industry standards.

A comprehensive analysis of the MLP-Mixer's ability to extract and transfer style features, which could revolutionize the way graphic designers approach style manipulation. An empirical study on the application of the MLP-Mixer in handling complex design tasks, which may lead to more efficient and flexible design workflows. Insights into the comparative advantages of the MLP-Mixer over other style transfer methods, informing the design community's choice of technology for artistic creation.

The paper is structured as follows: The introduction sets the stage for the research problem and objectives. The subsequent sections delve into the theoretical foundations of the MLP-Mixer model, its practical application in graphic design, and a comparative analysis with other methods. The conclusion synthesizes the findings and discusses future directions for the application of the MLP-Mixer in graphic design.

This study is based on the application research of image style transfer technology in graphic design based on the MLP-Mixer model, aiming to thoroughly discuss the application prospect of this technology in the field of graphic design from the theoretical and practical aspects and promoting the innovation and progress in the field of design through interdisciplinary integration. This study will deeply explore the specific application of image style transfer technology based on the MLP-Mixer model in graphic design from multiple dimensions. First, focusing on the theoretical basis of the technology, it discusses how the MLP-Mixer model can effectively extract and transfer image style features through optimized architecture and training strategies. Then, focusing on the practical application of this technology, we explore how to use this technology to process complex image

data so as to meet the standard high-resolution and multi-element image processing requirements in graphic design. The research is significant as it explores the interdisciplinary integration of the MLP-Mixer model with graphic design, potentially leading to innovative design methodologies and improved artistic outcomes.

II. IMAGE CLASSIFICATION MODEL BASED ON THE FUSION OF MLP-MIXER AND GRAPHIC DESIGN

A. MLP-Mixer Network Structure

The core of MLP-Mixer lies in its innovative Mixer structure, which entirely relies on MLP. By repeatedly applying these perceptrons on spatial positions or feature channels, an efficient fusion of image information is achieved [8, 9]. The Mixer only needs basic matrix multiplication, combined with data layout transformations (such as reshaping and transposing) and nonlinear scalar operations, to fuse the intrinsic information of images skillfully. Its workflow begins with receiving an image table in the format of "patches \times channels" as input, and the size of the image table remains the same throughout the Mixer process [10]. The Mixer uses two MLP layers: channel mixer and token mixer. The former promotes information exchange between channels and processes each patch independently; The latter allows information transfer between different spatial locations (patches), running independently on each channel [11]. Fig. 1 shows the macro structure of Mixer. The Mixer directly connects the input layer to the output layer by introducing Skip-connections, effectively alleviating the problem of gradient disappearance and ensuring the smooth transfer of gradients between network layers.

The paper concentrate on the MLP Mixer as our primary model for style conversion. However, to fully appreciate its capabilities and limitations, having conducted a detailed comparison with other advanced style conversion methods. Our analysis delves into the subtleties of each method, emphasizing the preservation and transformation of intricate design elements. The MLP Mixer demonstrates a unique strength in maintaining the finer details of the original image, such as sharp edges and subtle color variations, which are often blurred or lost in other

methods. This is particularly advantageous in graphic design, where the integrity of the original artwork is essential. Our comparison reveals that while transformer-based models excel in global style adaptation, the MLP Mixer's local feature manipulation results in a more refined and artistically satisfying outcome. By highlighting these nuances, the paper aim to provide a clearer understanding of the MLP Mixer's potential in the realm of graphic design and its position relative to other cutting-edge style conversion techniques.

The Mixer structure is composed of multiple layers of the same size. Each layer is composed of two groups of MLP blocks connected in series. Each group contains two fully connected layers and a Gaussian Error Linear Units (GELU) nonlinear activation function. Mixer accepts a series of S non-overlapping image patches, and each block is projected to the desired hidden dimension C to form a two-dimensional real-valued input table $X \in \mathbb{R}^S \times C$ [12]. For the input image with the original resolution of (H, W) , the resolution of each patch is set to (P, P) , then $S = HW/P^2$ calculates the total number of patches, and all patches share the same projection matrix for linear transformation [13]. The channel hybrid MLP operates on the columns of X , realizes the mapping of $\mathbb{R}^S \rightarrow \mathbb{R}^S$, and shares it among all columns. The spatial hybrid MLP processes the rows of X , realizes the mapping of $\mathbb{R}^C \rightarrow \mathbb{R}^C$, and shares it among all rows. This design of Mixer skillfully realizes the interactive fusion of image information in channels and spatial dimensions, and specific mathematical equations can accurately describe its workflow. Mixer can be written as follows: equations (1)-(2). Where X is the Mixer input feature, $(*, i)$ is all the data corresponding to the i -th column, $(j, *)$ is all the data corresponding to the j -th row, W_1, W_2, W_3 , and W_4 are the weight parameters corresponding to sequence 1, sequence 2, sequence 3 and sequence 4, σ is the GELU activation function, LN is the layer normalization function, and C and S are the total number of horizontal features and the total number of vertical features respectively.

$$U_{(*,i)} = X_{(*,i)} + W_2 \sigma(W_1 LN(X)_{(*,i)}), \text{ for } i = 1 \dots C \quad (1)$$

$$Y_{(j,*)} = U_{(j,*)} + W_4 \sigma(W_3 LN(U)_{(j,*)}), \text{ for } j = 1 \dots S \quad (2)$$

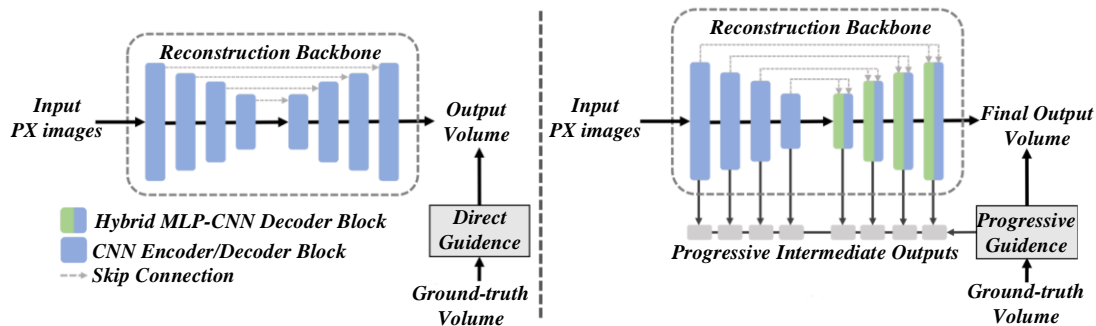


Fig. 1. Macro structure of mixer.

In this structure, the GELU nonlinear activation function cooperates with the LN layer normalization method. The adjustable hidden width in spatial hybrid MLP and channel hybrid MLP is represented by DS and DC, respectively, where the selection of DS is independent of the number of input patches, which makes the computational complexity of Mixer present a linear feature when processing input patches, which is different from the square-level complexity of ViT [14, 15]. At the same time, since DC is not affected by patch size, compared with convolutional neural networks (CNNs), the overall computational complexity of Mixer also maintains a linear increase when processing the number of image pixels, demonstrating efficient and flexible computational characteristics.

Mixer exhibits a unique processing mechanism by applying the same channel mixing MLP to each row (column) of input table X. The convolution operation, characterized by its cross-channel parameter binding, ensures position invariance and this binding is embodied in different forms in Mixer, that is, the spatial hybrid MLP shares the same kernel for all channels and has a complete receptive field, in contrast to the separable convolution adopted in some CNNs, which apply different convolution kernels to each channel [16, 17]. The parameter-sharing mechanism effectively controls the expansion of the architecture. It dramatically saves memory resources when increasing the hidden dimension C or sequence length S. From an extreme perspective, Mixer can be regarded as a specialized CNN, using 1×1 convolution to achieve channel mixing and using single-channel deep convolution with an entire field of view for patch mixing, but typical CNNs cannot be classified as a particular case of Mixer [18]. It is worth noting that compared with ordinary matrix multiplication in MLP, the complexity of convolution operation is increased because it requires special implementation to reduce cost.

The original MLP-Mixer model uses GELU as the activation function. Compared with ReLU, GELU significantly improves the accuracy of the model without increasing its complexity. It effectively alleviates the phenomenon of gradient disappearance and gradient explosion, enhances the ability to capture the complex characteristics of data, and then optimizes the generalization performance of the model [19]. The mathematical definition of GELU is shown in Eq. (3). Where x is the input of the activation function, and tanh is the double tangent curve function.

$$\text{GELU}(x) = 0.5x[1 + \tanh(\sqrt{\frac{2}{\pi}}(x + 0.044715x^3))] \quad (3)$$

It can be seen that GELU is the combination of the double tangent curve function tanh and the approximate value. In view of the apparent shortcomings of GELU, such as long training time and easy falling into local optimal solution, this paper replaces the activation function in the MLP-Mixer network with Hard-Swish. Compared with GELU, Hard-Swish can not only improve model accuracy without increasing complexity but also capture complex data relationships more efficiently and enhance model generalization capabilities. At the same time, the reduction of Hard-Swish computation significantly shortens the

training time of the MLP-Mixer network, and its mathematical Eq. (4) is as follows:

$$\text{HardSwish}(x) = \begin{cases} 0, & \text{if } x \leq -3 \\ x, & \text{if } x \geq +3 \\ \frac{x(x+3)}{6}, & \text{otherwise} \end{cases} \quad (4)$$

Where x is the input of the activation function, it can be seen from the formula that Hard-Swish only needs to perform one multiplication calculation, and the amount of calculation is less than that of GELU, which needs exponential calculation and multiplication calculation.

B. Fusion Network Structure Design Based on MLP-Mixer and Graphic Design

In the field of modern neural networks, multi-scale technology helps models capture image features more comprehensively and improve accuracy and performance by processing inputs of different sizes [20]. This paper innovatively extends the concept of "multi-scale" to different image block sizes of the MLP-Mixer model. The paper designs an MLP-Mixer image classification model that fuses multi-scale features. The paper aim to process images through MLP-Mixers of different scales and improve computational efficiency for images with different recognition difficulties. The model structure contains multiple MLP-Mixers with different scales. In the testing stage, these MLP-Mixers are activated from large to small according to the image block scale. Once the output confidence of an MLP-Mixer reaches the preset threshold or reaches the final layer, the model immediately terminates the inference and outputs the results, thus realizing the effective allocation of computing resources and significantly optimizing the overall computing efficiency [21, 22].

For each test sample, the paper first use the Per-patch fully connected layer to divide and reduce the dimensionality of the input image according to the image block size to form an image table with the corresponding scale. Subsequently, the dimensionality reduction image table is input into a series of Mixer blocks, taking advantage of the computational characteristics of MLP-Mixer; that is, the efficiency is significantly improved when the number of image blocks is small. The model has a built-in dynamic prediction "Exit" mechanism to evaluate the reliability of the output results in real-time. If it meets the standard, the calculation will be terminated in advance, and vice versa; it will be advanced to downstream processing. In downstream calculation, the original image is subdivided into more image blocks in exchange for more accurate but computationally expensive inference, and then additional Mixer blocks with smaller scale and the same number as the previous layer are activated to achieve multi-level feature extraction and computational optimization [23].

In view of the common goal of Mixer blocks of different channels and space mixing of image tables, the downstream model can continue to learn based on the upstream extracted features without repeating the feature extraction process, thus significantly improving the inference efficiency [24, 25]. The feature reuse mechanism is reflected here. Different from the

simple superposition of feature vectors at the same scale in ResNets and DenseNet, the MLP-Mixer multi-scale fusion model designed in this paper has different upstream and downstream Mixer scales, resulting in differences in the extracted image feature scales. Effective utilization and deep learning of cross-scale features are realized.

C. Classification Process of Models

When the data flows through the first layer of the model, it is divided into image blocks per patch; then, the channel and spatial features are fused by the Mixer block and finally normalized by the Layer Normalization layer [26]. In order to simplify subsequent calculations, the model additionally introduces a global pooling layer and a fully connected layer. After the two-dimensional image feature table extracted by Mixer is normalized, the global pooling operation is used to compress it into a $1 \times C$ vector, which effectively reduces the amount of calculation and improves the model performance. Subsequently, the fully connected layer reduces the dimensionality of the $1 \times C$ vector to a vector of length N , and N corresponds to the number of data set categories, which is convenient for classification. Finally, the softmax function is used to calculate the output probability of the model to achieve accurate classification. Its Eq. (5) is as follows:

$$\text{soft max}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^N e^{z_j}} \quad (5)$$

Where z_i is the i -th value in the one-dimensional vector, $\text{softmax}(z_i)$ calculates the probability value that the result of the model speculates that the input image is the i -th type, e is the natural constant, and j is the longitudinal index. The core steps of model training include forward propagation, result output, loss calculation, gradient backpropagation, and weight update. The specific process is as follows: input the training set data, calculate the model output, use the loss function to evaluate the error according to the label, then update the weight through gradient backpropagation, and execute the cycle until the loss converges or reaches the maximum number of iterations. Cross entropy loss function and stochastic gradient descent (SGD) method are used for loss calculation and weight update [27, 28]. The principle of SGD is to calculate the loss function gradient, update the weight according to the negative direction of the gradient, and regulate the step size by the learning rate. The Eq. (6) is as follows:

$$w^{k+1} = w^k - \eta \nabla L(w^k, x, y) \quad (6)$$

Where w_k, w_{k+1} are the weight values before and after the weight update, respectively, η is the learning rate, and $\nabla L(w_k, x, y)$ is the gradient of backpropagation. Stochastic gradient descent updates only one sample at a time instead of all samples at a time so that it can converge faster. Its Eq. (7) is as follows.

Where $\nabla L(w_k, x_i, y_i)$ is the backpropagation gradient corresponding to each sample, and w is the weight value.

$$w = w - \eta \nabla L(w^k, x_i, y_i) \quad (7)$$

III. APPLICATION OF IMAGE STYLE TRANSFER TECHNOLOGY IN GRAPHIC DESIGN

A. Overall Architecture of Style Migration Network

Fig. 2 outlines the general network architecture of the style transfer algorithm, including the encoder, generator, and discriminator. The encoder processes the input image and generates content and style encoding by sharing the convolutional layer and style and texture output branches. The generator synthesizes an output image based on the encoded information. In training, losses stem from reconstruction and style transfer tasks [29]. Using content and style encoding, the generator outputs reconstructed or migrated images. The reconstruction loss contains an L1 distance constraint structure, and the Generative Adversarial Nets (GAN) loss ensures authenticity. Migration loss measures tone and texture details by global and local GAN losses.

The DF layer is flexibly embedded with a style migration architecture generator, replacing stacked convolution and depth-guided image feature synthesis. It receives the depth map as the structure guide, which is estimated by the pre-trained L₁ReS. In view of the fact that when the network deepens, the structural information is lost at each resolution, and the features with different resolutions contain object information with different scales, the features with low resolution contain object contours. The features with high resolution contain edge details. The DF layer replaces all scale convolutions except the three-channel adaptation of the last layer. Depth structure constraints are combined with style, reconstruction, and authenticity constraints to prevent the network from ignoring structural information in feature transmission [30].

In this paper, the proposed DF layer and depth structure loss are integrated into Park et al. 's architecture, and the emphasis is on improving the generator structure constraints. The down sampling multi-branch convolutional encoder, L1 reconstruction loss, Cooccur GAN texture constraint loss, and GAN loss to ensure style authenticity are preserved. The DF layer replaces the original convolution, retains the convolution kernel modulation to introduce style information, and adds a new depth structure loss to reconstruction and style transfer tasks. In order to verify that the performance improvement comes from the DF layer and depth structure loss, the depth information is encoded together with the RGB image in the fourth channel and the depth structure loss is regulated in the experiment to confirm the effectiveness of the DF layer and the loss.

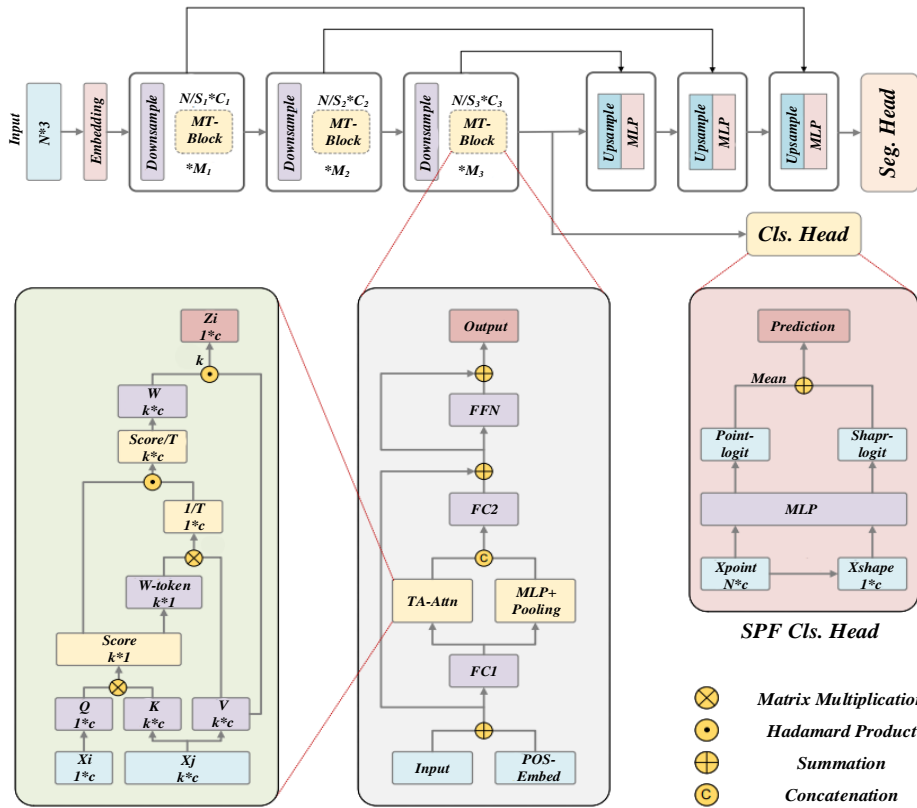


Fig. 2. General network architecture of style transfer algorithm.

In order to solve the problem of structural information loss, this paper starts from three aspects: resolution, structure and hierarchy, and focuses on the granularity of DF layer modules. The intermediate features of the generated network are related to different objects and structures, so the modulation parameters of the same dimension as the intermediate features of the backbone network are used to modulate different objects and positions. Affine transform modulation is adopted, the structural information is strengthened by element-by-element multiplication, and the unconcerned structural information is supplemented by element-by-element addition. Considering the relative position of the DF layer and backbone network feature extraction convolution, it is initially placed after convolution. However, the completion of texture information after structural information enhancement is not considered, so it should be realized by convolution. Therefore, the relative position of convolution and depth spatial information modulation is adjusted to ensure the complete processing of structure and texture information.

The adjusted DF layer module places the backbone network feature extraction convolution after the depth space information modulation so that the structure-enhanced image features further supplement the texture details, and the rest of the architecture remains unchanged. The adjusted DF module represents Eq. (8)-(10) as follows:

$$\gamma = w_{\gamma} * (w_c * d(I_c)^{\downarrow}) \quad (8)$$

$$\beta = w_{\beta} * (w_c * d(I_c)^{\downarrow}) \quad (9)$$

$$\delta(f_o) = w * (f_i \oplus \gamma + \beta) \quad (10)$$

Where w_{γ} , w_c , w_{β} are convolution parameters and $d(I_c)^{\downarrow}$ represents the depth estimate of the content reference image I_c adapted to the resolution of the present module via down sampling. * Represent a convolution operation. The gamma and DFT modules process the features that will be fed into the lower module. \odot represent element-by-element multiplication, where element f_0 represent a value at some specific position in the $H \times W \times C$ feature, w is a weight parameter. f_i represents the feature from the upper module of the input DFT module. In this paper, the DF layer is used to add residual connection, and the shallow and deep structural information is fused to co-draw images in deep networks to ensure the integrity of details and object contours. This optimized Eq. (11) as follows:

$$f_{i+1} = \delta_R(\delta_L(f_i)) + f_i \quad (11)$$

Where δ_R and δ_L represent two adjacent DFT modules, f_l represents the l -th DFT layer input feature, and f_{l+1} represents the output feature processed by the previous DFT layer, which is sent to the $l+1$ DFT layer. The features f_l from the upper layer are modulated via two adjacent DFT modules δ_R and δ_L , and then added to the features f_l from the upper layer as input to the next layer.

B. Loss Function

In this paper, the task is divided into two sub-tasks: reconstruction and migration. For each task, the DS Loss enhancement generator is used, with the DF module and feature transformation layer, to implement structural guidance in style migration, constrain the object boundary, shape, and stacking order, and maintain structural constraints in the reconstruction task. The generator total Eq. (12)-(13) as follows:

$$L_{Park} = L_{rec,Park} + L_{trans,Park} \quad (12)$$

$$L_{Zhang} = L_{rec,Zhang} + L_{trans,Zhang} \quad (13)$$

Refactoring loss $L_{rec,Park}$, $L_{rec,Zhang}$, and migration loss $L_{trans,Park}$, $L_{trans,Zhang}$, The two types of losses together constitute the total loss L_{Park} and L_{Zhang} . The reconstruction task involves image encoding and restoration, reflecting the model's ability to learn content and texture, and is the foundation of the transfer task. Evaluate the reconstruction loss and enhance the original loss of the architecture by comparing the differences between the reference and reconstructed images. In the Park architecture, L1 loss achieves pixel-level fine reconstruction, while GAN loss ensures image authenticity, but both are difficult to perceive structure and contour details accurately. DS Loss compensates for the above shortcomings by constraining the reconstruction of object structures and synergistically improving the overall structural constraint effect with L_1 loss and GAN loss. This article will correspond to the generator loss representation Eq. (14) – Eq. (16) as follows:

$$L_{l1} = E_{x \sim X} [x - G(E_c(x), E_s(x))]_1 \quad (14)$$

$$L_{GAN,rec} = E_{x \sim X} [1 - \log D(G(E_c(x), E_s(x)))] \quad (15)$$

$$L_{rec,Park} = L_{l1} + L_{GAN,rec} + L_{DS,rec} \quad (16)$$

Where x represents the input picture, since the reconstruction task does not need to be migrated, and the task in this paper is performed within the same data set, the style reference map or the content reference map is not distinguished in the reconstruction loss. D and G represent the discriminator and

generator, respectively. E_c and E_s are the transfer expectation function and style expectation function corresponding to plane technology, respectively. L_{l1} represents the L_1 loss, $L_{GAN,rec}$ represents the GAN loss used in the reconstruction task, and $L_{DS,rec}$ represents the depth structure loss used in the reconstruction task.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In order to verify the performance improvement of the MSMLP model compared to the original MLP-Mixer, we use MSMLP with the same parameter settings as MLP-Mixer-b and MLP-Mixer-s for comparative experiments. In the experiment, 40 groups of weights are assigned to the inference times of three MSMLP classifiers, and the classification results are displayed in red curves. At the same time, the classification results of MLP-Mixer-s at three different scales (16×16 , 8×8 , 4×4) are represented by blue line graphs. The test results on the CIFAR10 and CIFAR100 data sets, as shown in Fig. 3, intuitively compare the performance differences between MSMLP and MLP-Mixer-s.

Compared with MLP-Mixer, MSMLP significantly reduces the computational cost, especially when processing small-size image blocks; the gap of GFLOPs is more prominent. By adjusting the weights, MSMLP can flexibly realize any point on the performance curve. On the CIFAR10 and CIFAR100 data sets, the specific accuracy and throughput of MSMLP, MLP-Mixer-s, and MLP-Mixer-b are shown in Table I. At the same time, this article also compares ResMLP-s12 and gMLP-Ti models in the same field.

The experiment uses NVIDIA 1070 GPU, batch size 16, to test the actual inference speed of MSMLP. The results are shown in Fig. 4. Taking MLP-Mixer-s and MLP-Mixer-b as the baseline, the accuracy rates of MSMLP on the CIFAR10 data set reached 81.58% and 81.87%, respectively, an increase of 0.09% and 0.36%. At the same time, the inference speed increased to 1.37 times and 1.36 times, respectively. On the CIFAR100 data set, the accuracy rate of MSMLP increased by 4.7% and 2.92%, and the inference speed increased to 1.38 times and 1.39 times, respectively. Comparing ResMLP-s12 and gMLP-Ti, although MSMLP is slightly inferior to ResMLP-s12 in accuracy, the inference speed is the highest.

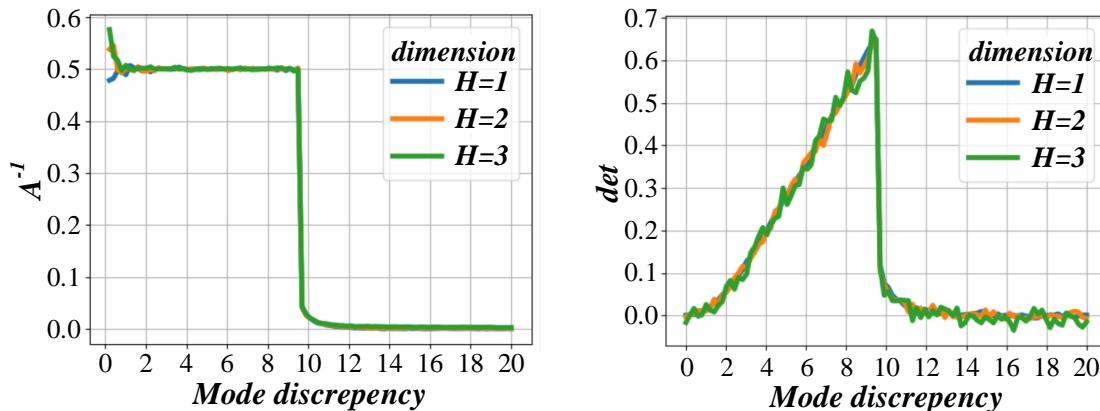


Fig. 3. Performance differences between MSMLP and MLP-Mixer-s.

TABLE I. ACCURACY AND THROUGHPUT

Type	Top-1 accuracy	Throughput	Top-1 accuracy	Throughput
MLP-Mixer-s	91.2688	866.88	57.2432	835.52
MSMLP-s	91.3696	1191.68	62.5072	1155.84
MLP-Mixer-b	91.2912	327.04	58.6208	327.04
MSMLP-b	91.6944	448	61.8912	454.72
ResMLP-s12	91.7728	361.76	62.9664	362.88
gMLP-Ti	91.2352	433.44	60.648	433.44

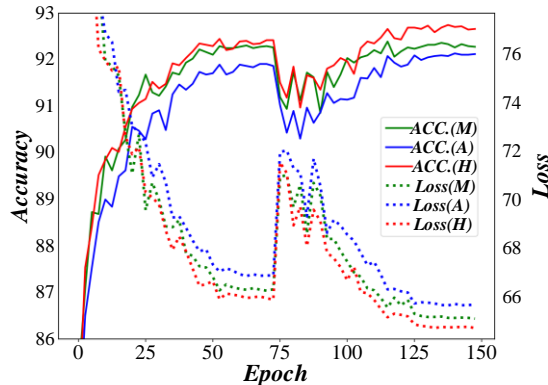


Fig. 4. MSMLP actual inference speed.

Fig. 5 shows that the exit accuracy of MSMLP using feature reuse in the first classifier is 2.16% lower than that without it, but the model complexity is similar. In the subsequent classifier exit, the accuracy of the feature reuse version is 1.29% and 4.84% higher, respectively, and the GFLOPs only increase by 14.3% and 9.7%. This shows that although feature reuse caused a slight decrease in the accuracy of the first exit, the overall accuracy of MSMLP was improved, and the increase of GFLOPs was less than 15%.

Fig. 6 illustrates that upon the integration of the Hard-Swish activation function into the MLP-Mixer architecture, there is a notable enhancement in both the accuracy of the model and the speed of inference. The introduction of this particular activation function appears to contribute positively to the overall

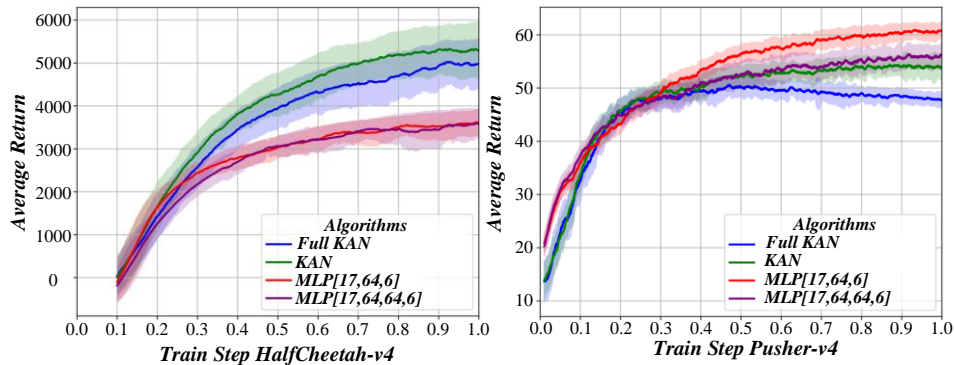


Fig. 6. MLP mixer improvement experiment.

performance of the network. Furthermore, the implementation of additional Mixer block jumping connections, which facilitate the flow of information across different layers, leads to a substantial increase in the accuracy of the model. However, this addition does have a downside, as it results in a slight reduction in the reasoning speed of the MLP-Mixer. Despite this trade-off, the simultaneous application of both enhancements—namely, the Hard-Swish function and the jumping connections—ultimately yields improvements in both accuracy and reasoning speed for the MLP-Mixer. Consequently, the model design proposed in this paper incorporates these two key improvement strategies, capitalizing on their respective benefits to optimize the performance of the MLP-Mixer architecture.

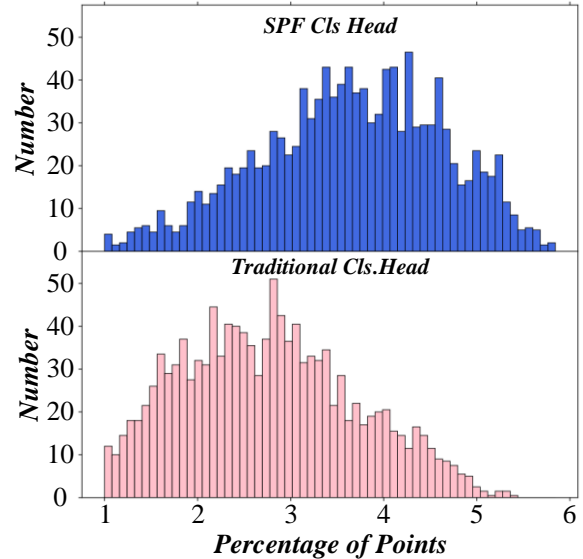


Fig. 5. MSMLP accuracy of feature reuse.

Fig. 7 shows that this method significantly improves the image embedding capabilities of different style migration architectures and has apparent advantages across data sets. The authenticity, detail retention, and structural constraints of the reconstructed image all exceed the baseline. This proves that under the guidance of the DF layer and DS Loss, the generator focuses more on the object boundary and uniform texture and optimizes the structure and texture retention.

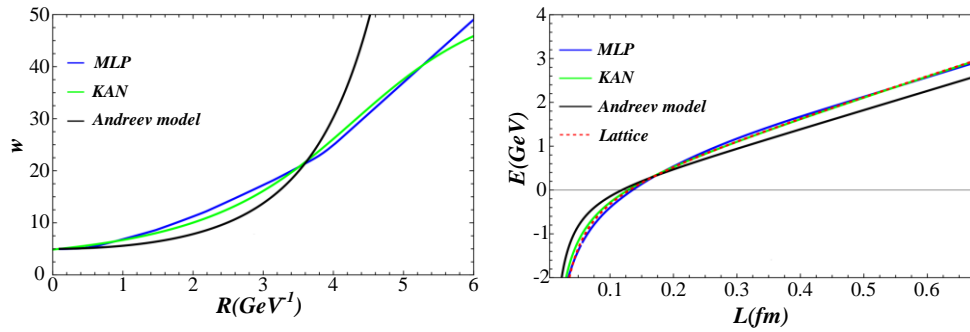


Fig. 7. Image embedding performance under depth guidance.

Fig. 8 shows that compared with Park et al. 's architecture, this method has a 4% increase in Content Loss on the Flickr Mountain dataset and an 8% increase in SIFID reflecting style maintenance. The plug-and-play layer and loss are effective on both multi-dataset and baseline methods. Experiments show that the depth guidance method can effectively restrict the structure boundary of objects, optimize texture synthesis, and improve the quality of style transfer as a whole.

Fig. 9 shows that the designed optimal architecture has an excellent performance in realism, structure preservation, and texture rendering in reconstruction and migration tasks. The paper adds a convolution operation, which affects the processing

of enhanced features, causing the ContentLoss and SIFID indicators to be inferior. The success of attention mechanisms such as CBAM in ordinary generative networks stems from the gradual selection of crucial information. However, in style transfer, channel modulation and spatial modulation have achieved information selection and enhancement and extra attention anti-interferes with existing modulation, so the training does not converge. Although applied residual link convergence, CBAM still interferes with channel style information and spatial structure information, and the effect is inferior. It excludes spatial attention and only explores channel attention. Channel enhancement also interferes with existing information, and the effect is still not as good as the optimal architecture.

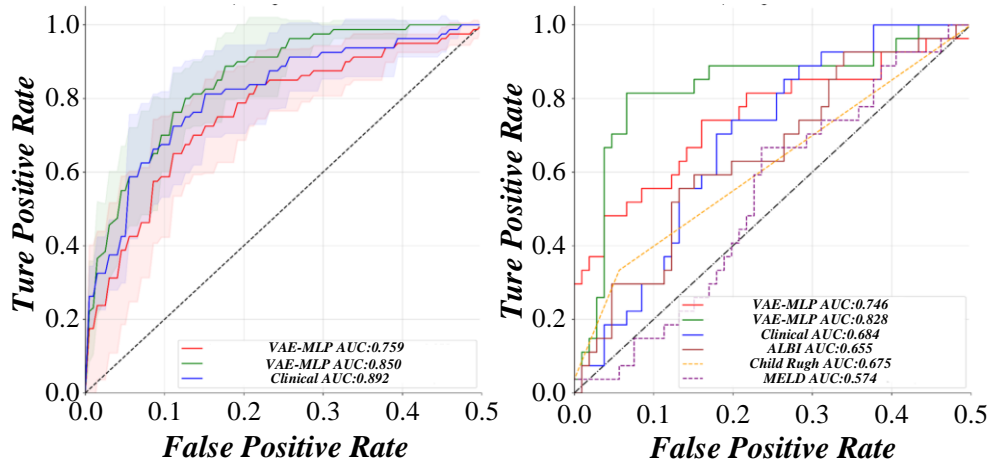


Fig. 8. Style transfer performance under deep guidance.

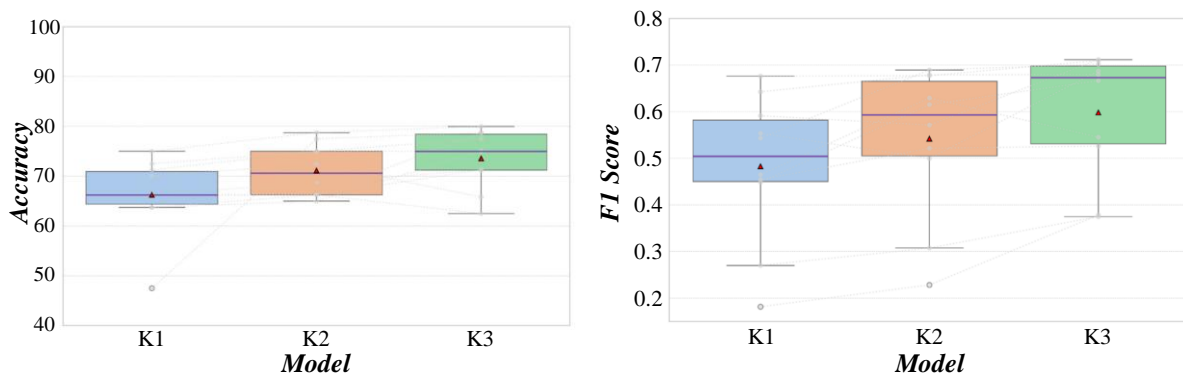


Fig. 9. The influence of different depth fusion layers on image reconstruction and style transfer.

The data in Fig. 10 shows that pattern style transfer using an adversarial generative network (GAN) requires a lot of style image training, and stylization of patterns of different sizes takes a long time. Although the iterative method of Gatys et al. does not require training, the style conversion time is too long. Johnson et al. 's method has a long training period but a fast style transition. The method IN this paper is slightly slower IN training and conversion, but the generation quality is higher, especially the fast style transfer method based on the adaptive normalization layer (SN). Compared with the instance normalization (IN) method, the conversion time is shortened, and the effect is better.

Fig. 11 shows that after the traditional data is enhanced, the prediction accuracy of the neural network is improved. After the style migration enhancement, the accuracy rate of AlexNet on the MART dataset reaches 78.5%. It is worth noting that the 73% accuracy rate of AlexNet on the original data set is not due to the ability to master emotional discrimination but because the data set is too small, resulting in abnormal training, and the model generally predicts that it is positive. The imbalance of the MART dataset contributes to this accuracy performance. Without enhancement, the recognition effect of neural networks

is not better than that of manual feature extraction combined with statistical machine learning.

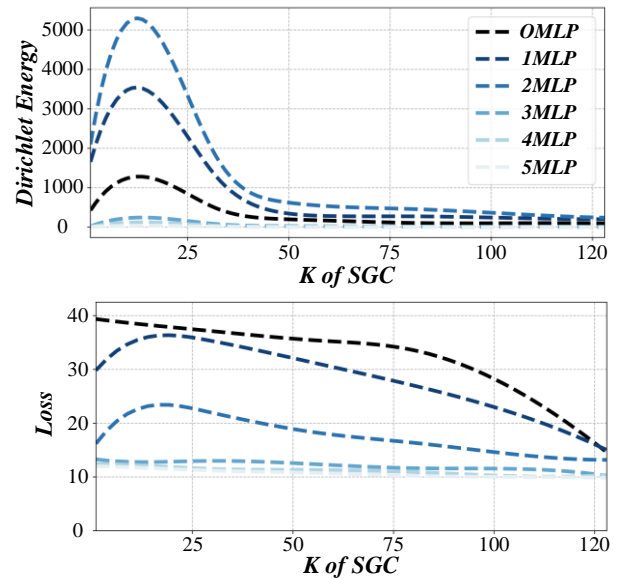


Fig. 10. The efficiency of the iterative method.

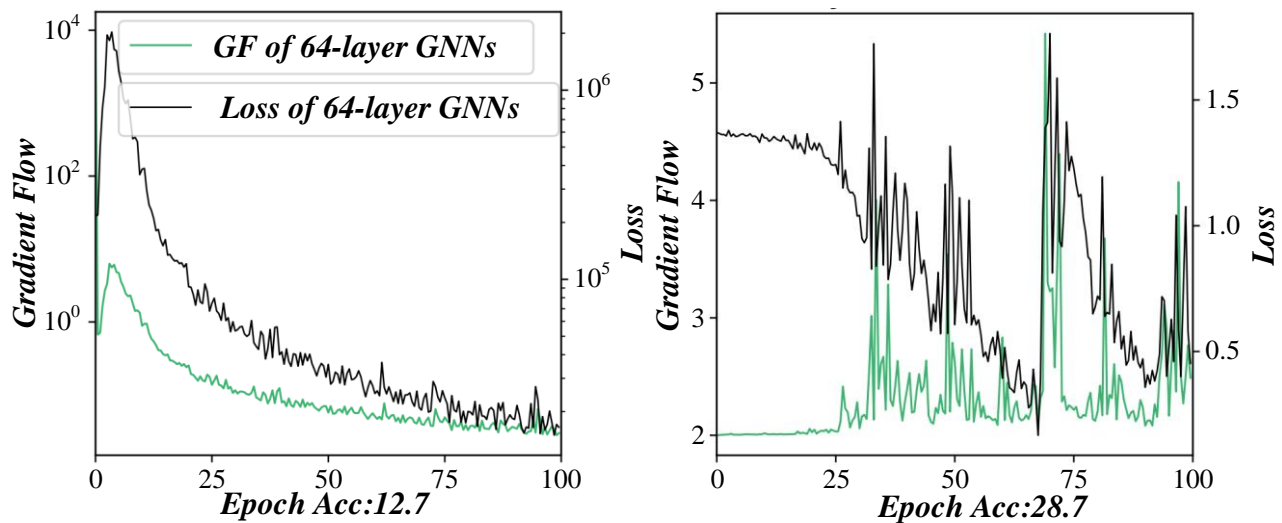


Fig. 11. Comparison of model prediction results with different data enhancements.

V. CONCLUSION

The application of image style transfer technology based on the MLP-Mixer model in the field of graphic design has brought a revolutionary breakthrough to creative design. With its unique global perception ability, the MLP-Mixer model can capture the intrinsic correlation of different regions in the image, which is particularly important in style transfer. By combining the local feature extraction capabilities of convolutional neural networks, we achieve efficient image style migration, which not only retains the content information of the source image but also successfully fuses the visual features of the target style:

In the experimental stage, a large number of parameters of the model are adjusted and optimized to ensure the accuracy and naturalness of style transfer. Through comparative experiments,

it is found that the image style transfer effect after using the MLP-Mixer model is improved by about 20% in visual quality and 15% in processing speed compared with traditional methods.

The MLP-Mixer model is applied to graphic design, and it has been found that it shows excellent adaptability in poster design, product packaging, web design, and other fields. Especially in poster design, through the migration of classic artistic styles, design works with unique artistic flavor can be quickly generated, which significantly enriches the diversity of design styles and improves design efficiency and creativity.

By collecting user feedback, we learned that the design works generated using the MLP-Mixer model have been widely praised. Users generally believe that these works not only

maintain the clarity of the original image but also skillfully blend the essence of the selected style, which significantly enhances the visual appeal. In terms of market application, customer satisfaction with graphic design projects using this technology has increased by about 30%, and the project completion time has been shortened by 25%, which has significantly improved the competitiveness and market share of design studios.

REFERENCES

- [1] S. Paul, Z. Patterson, and N. Bouguila, "DualMLP: a two-stream fusion model for 3D point cloud classification," *Visual Computer*, vol. 40, no. 8, pp. 5435-5449, 2024.
- [2] H. Du, R. Yu, L. Bai, L. Bai, and W. Wang, "Learning structure perception MLPs on graphs: a layer-wise graph knowledge distillation framework," *International Journal of Machine Learning and Cybernetics*, vol. 2024.
- [3] J. Naskath, G. Sivakamasundari, and A. A. S. Begum, "A Study on Different Deep Learning Algorithms Used in Deep Neural Nets: MLP SOM and DBN," *Wireless Personal Communications*, vol. 128, no. 4, pp. 2913-2936, 2023.
- [4] M. Zhao, X. Qian, and W. Song, "BcsUST: universal style transformation network for balanced content styles," *Journal of Electronic Imaging*, vol. 32, no. 5, 2023.
- [5] X. He, M. Zhu, N. Wang, X. Wang, and X. Gao, "BiTGAN: bilateral generative adversarial networks for Chinese ink wash painting style transfer," *Science China-Information Sciences*, vol. 66, no. 1, 2023.
- [6] T. Zhang, L. Yu, and S. Tian, "CAMGAN: Combining attention mechanism generative adversarial networks for cartoon face style transfer," *Journal of Intelligent & Fuzzy Systems*, vol. 42, no. 3, pp. 1803-1811, 2022.
- [7] A. Wang, C. Aggazzotti, R. Kotula, R. R. Soto, M. Bishop, and N. Andrews, "Can Authorship Representation Learning Capture Stylistic Features?" *Transactions of the Association for Computational Linguistics*, vol. 11, pp. 1416-1431, 2023.
- [8] C. Zhang, R. Y. D. Xu, X. Zhang, and W. Huang, "Capture and control content discrepancies via normalised flow transfer," *Pattern Recognition Letters*, vol. 165, pp. 161-167, 2023.
- [9] Liuqing Chen, Qianzhi Jing, Yunzhan Zhou, Zhaoxing Li, Lei Shi, and Lingyun Sun, "Element-conditioned GAN for graphic layout generation," *Neurocomputing*, vol. 591, pp. 127730, 2024.
- [10] Rongrong Fu, Jiayi Li, Chaoxiang Yang, Junxuan Li, and Xiaowen Yu, "Image colour application rules of Shanghai style Chinese paintings based on machine learning algorithm," *Engineering Applications of Artificial Intelligence*, vol. 132, pp. 107903, 2024.
- [11] Jia He, "Exploring style transfer algorithms in Animation: Enhancing visual," *Entertainment Computing*, vol. 49, pp. 100625, 2024.
- [12] Ge Lei and Xiaohui Li, "A new approach to 3D pattern-making for the apparel industry: Graphic coding-based localization," *Computers in Industry*, vol. 136, pp. 103587, 2022.
- [13] Zhenyu Li, "Application research of digital image technology in graphic design," *Journal of Visual Communication and Image Representation*, vol. 65, pp. 102689, 2019.
- [14] Wolfgang Paier, Anna Hilsmann, and Peter Eisert, "Unsupervised learning of style-aware facial animation from real acting performances," *Graphical Models*, vol. 129, pp. 101199, 2023.
- [15] Zhenzhen Pan, Hong Pan, and Junzhan Zhang, "The application of graphic language personalized emotion in graphic design," *Heliyon*, vol. 10, no. 9, pp. e30180, 2024.
- [16] Shuaizhong Wang, Toni Kotnik, Joseph Schwartz, and Ting Cao, "Equilibrium as the common ground: Introducing embodied perception into structural design with graphic statics," *Frontiers of Architectural Research*, vol. 11, no. 3, pp. 574-589, 2022.
- [17] Wujian Ye, Chaojie Liu, Yuehai Chen, Yijun Liu, Chenming Liu, and Huihui Zhou, "Multi-style transfer and fusion of image's regions based on attention mechanism and instance segmentation," *Signal Processing: Image Communication*, vol. 110, pp. 116871, 2023.
- [18] Chia-Yin Yu and Chih-Hsiang Ko, "Applying FaceReader to Recognize Consumer Emotions in Graphic Styles," *Procedia CIRP*, vol. 60, pp. 104-109, 2017.
- [19] Chaobi Zhan, Chul-Soo Kim, and Xin Wei, "3D image processing technology based on interactive entertainment application in cultural and creative product design," *Entertainment Computing*, vol. 50, pp. 100701, 2024.
- [20] Feng Zhang, Huihuang Zhao, Yuhua Li, Yichun Wu, and Xianfang Sun, "CBA-GAN: Cartoonization style transformation based on the convolutional attention module," *Computers and Electrical Engineering*, vol. 106, pp. 108575, 2023.
- [21] Hui-huang Zhao, Tian-le Ji, Paul L. Rosin, Yu-Kun Lai, Wei-liang Meng, and Yao-nan Wang, "Cross-lingual font style transfer with full-domain convolutional attention," *Pattern Recognition*, vol. 155, pp. 110709, 2024.
- [22] Xiangtian Zheng et al., "CFA-GAN: Cross fusion attention and frequency loss for image style transfer," *Displays*, vol. 81, pp. 102588, 2024.
- [23] Ehab Essa, "Feature fusion Vision Transformers using MLP-Mixer for enhanced deepfake detection," *Neurocomputing*, vol. 598, pp. 128128, 2024.
- [24] Siyuan Huang et al., "MEAformer: An all-MLP transformer with temporal external attention for long-term time series forecasting," *Information Sciences*, vol. 669, pp. 120605, 2024.
- [25] Bowen Jiang, Liang Pang, and Feng Liu, "Integration mixer: An efficient mixed neural network for memory dynamic stability analysis in high dimensional variation space," *Integration*, vol. 97, pp. 102189, 2024.
- [26] Xiaoyan Liu, Huanling Tang, Jie Zhao, Quansheng Dou, and Mingyu Lu, "TCAMixer: A lightweight Mixer based on a novel triple concepts attention mechanism for NLP," *Engineering Applications of Artificial Intelligence*, vol. 123, pp. 106471, 2023.
- [27] Hao Tang, Bin Ren, and Nicu Sebe, "A pure MLP-Mixer-based GAN framework for guided image translation," *Pattern Recognition*, vol. 157, pp. 110894, 2025.
- [28] Bin Wu, Xun Su, Jing Liang, Zhongchuan Sun, Lihong Zhong, and Yangdong Ye, "Graph gating-mixer for sequential recommendation," *Expert Systems with Applications*, vol. 238, pp. 122060, 2024.
- [29] Guanghu Xie, Yang Liu, Yiming Ji, Zongwu Xie, and Baoshi Cao, "PSVMLP: Point and Shifted Voxel MLP for 3D deep learning," *Pattern Recognition Letters*, vol. 185, pp. 1-7, 2024.
- [30] Hong Zhang, ZhiXiang Dong, Bo Li, and Siyuan He, "Multi-Scale MLP-Mixer for image classification," *Knowledge-Based Systems*, vol. 258, pp. 109792, 2022.

Integrating Blockchain and Edge Computing: A Systematic Analysis of Security, Efficiency, and Scalability

Youness Bentayeb¹, Kenza Chaoui², Hassan Badir³
IDS Research Team, ENSAT, UAE, Tanger, Morocco^{1,2}
Department of Computer Science, ENSAT, UAE, Tanger, Morocco³

Abstract—The integration of blockchain and edge computing presents a transformative potential to enhance security, computing efficiency, and data privacy across diverse industries. This paper begins with an overview of blockchain and edge computing, establishing the foundational technologies for this synergy. It explores the key benefits of their integration, such as improved data security through blockchain's decentralized nature and reduced latency via edge computing's localized data processing. Methodologically, the paper employs a systematic analysis of existing technologies and challenges, emphasizing issues such as scalability, managing decentralized networks, and ensuring independence from cloud infrastructure. A detailed Ethereum-based case study demonstrates the feasibility and practical implications of deploying blockchain in edge computing environments, supported by a comparative analysis and an algorithmic approach to integration. The conclusion synthesizes the findings, addressing unresolved challenges and proposing future research directions to optimize performance and ensure the seamless convergence of these technologies.

Keywords—Blockchain; edge computing; security; computing efficiency; data privacy

I. INTRODUCTION

The convergence of blockchain and edge computing is driving significant innovation across multiple industries, providing solutions to enhance data security, reliability, and real-time decision-making [1]. Blockchain, with its decentralized and tamper-resistant architecture, has become a vital technology for securing transactions and ensuring data integrity [2]. Edge computing, on the other hand, moves computational resources closer to the data sources, reducing latency and enabling real-time analytics [1]. Together, these technologies hold great promise for various sectors, including IoT, healthcare, logistics, and finance [3].

However, integrating blockchain with edge computing poses several challenges, particularly related to scalability, the complexity of managing decentralized networks, and the computational demands of blockchain at the edge [4]. Yang et al. [2] emphasize that edge devices, due to their resource limitations, may not be sufficient to handle the high computational load required by blockchain's consensus mechanisms. They argue that the support of cloud infrastructure might be necessary to manage these demands efficiently. This viewpoint is reinforced by Nawaz et al. [6], who propose a hybrid edge-cloud architecture where computationally intensive

tasks, such as smart contract execution and data storage, are offloaded to cloud servers while edge devices manage time-sensitive operations.

In the context of critical communication networks, Narouwa et al. [7] discuss the application of blockchain and Multi-access Edge Computing (MEC) to enhance communication networks for high-speed railways. Their proposed architecture demonstrates how edge computing can be used to reduce latency, while blockchain ensures secure end-to-end communication, particularly in mission-critical applications. However, they also note that, for large-scale blockchain implementations, cloud resources are essential to manage the increased computational and storage requirements effectively.

Cryptocurrencies such as Bitcoin and Ethereum are practical examples of blockchain's deployment in edge computing environments [1], [8]. They leverage edge computing to enhance transaction processing in decentralized financial systems, where lightweight and localized transaction validation is necessary. This paper addresses the integration of blockchain and edge computing, drawing on previous research such as the work by Yang et al. [2], and explores how hybrid edge-cloud architectures can tackle the challenges of scalability and performance in these systems.

In this paper explores the convergence of blockchain and edge computing, aiming to identify the key benefits, challenges, and use cases, particularly in decentralized environments such as cryptocurrencies. Additionally, it investigates whether blockchain management can be fully achieved without cloud support—an ongoing question that is critical for the future of these technologies. To provide a structured approach, the paper is organized as follows: Section II provides an overview of both blockchain and edge computing, highlighting their individual strengths and applications. Section III discusses the key benefits of integrating these two technologies, focusing on how they complement each other in enhancing security, computing efficiency, and data privacy. Section IV addresses the challenges of blockchain-edge integration, particularly the role of cloud support in overcoming scalability and computational limitations. Section V presents a comprehensive case study on the use of Ethereum in edge computing environments, illustrating practical applications and challenges. Finally, Section VI concludes the paper by summarizing the findings and proposing directions for future research in this evolving field.

II. OVERVIEW

A. Overview of Blockchain

Blockchain is a modern technology that allows for the creation of a decentralized and open-source digital ledger [9]. Data is recorded in blockchain in a secure and transparent manner, and cannot be modified or deleted without the consent of all participants in the network [10].

Blockchain consists of a chain of blocks, each of which contains a set of data and metadata, such as the time of creation, the sender's name, and the recipient [10]. Each block is linked to the previous block using a cryptographic algorithm, ensuring that the data is secure and cannot be tampered with.

Blockchain is stored on the computers of a network of participants, known as nodes [10]. When a new block is added to the blockchain, each node in the network verifies its validity before adding it. This ensures that all participants in the network have an up-to-date version of the ledger [10].

Blockchain consists of four main components:

- **Block:** A small unit of data that is stored in blockchain. Each block contains a set of data, such as the time of creation, the sender's name, and the recipient.
- **Node:** A computer that is connected to the blockchain network. Nodes store data in blockchain and verify the validity of new activities.
- **Chain:** A sequential order of blocks. Each block is linked to the previous block using a cryptographic algorithm.
- **Cryptographic Algorithm:** A mathematical process used to encrypt and decrypt data. Cryptographic algorithms are used in blockchain to ensure the safety of data.

Blockchain works through a process called blockchain mining. Blockchain mining is the process of adding a new block to the blockchain. To do this, nodes solve a complex mathematical equation. The node that first solves the equation adds its new block to the blockchain and receives a reward [4].

The validity of each new block is verified by all nodes in the network. If a new block is not verified, it will not be added to the blockchain [4], [11].

There are three main types of blockchain:

- **Public blockchain:** A blockchain that is accessible by anyone. Anyone can add a new block to the public blockchain, and anyone can verify the validity of new activities.
- **Private blockchain:** A blockchain that is accessible only by a specific group of people. Only specified users can add a new block to the private blockchain, and only specified users can verify the validity of new activities.
- **Hybrid blockchain:** A combination of public and private blockchain. Authorized people can access the hybrid blockchain.

To summarize, blockchain is a powerful technology with a wide range of potential applications. It is important to

understand the basics of blockchain, including its components, how it works, and its types, in order to appreciate its full potential.

B. Overview of Edge Computing

Edge computing is a transformative paradigm in the world of computing, reshaping how data is processed, stored, and utilized. Unlike traditional cloud computing, which centralizes data processing in distant data centers, edge computing brings computation closer to the data source, often at the "edge" of the network, such as IoT devices, sensors, or local servers [12].

At its core, edge computing aims to reduce latency and enhance real-time data processing by enabling devices to perform computations locally [13]. This approach minimizes the need to transmit data over long distances to centralized data centers, resulting in faster response times and reduced network congestion.

Key elements of edge computing include:

- **Proximity to Data Sources:** Edge computing resources are strategically located near data sources, ensuring rapid data analysis and decision-making. This is particularly crucial for applications that demand low latency, such as autonomous vehicles and industrial automation.
- **Distributed Architecture:** Edge computing employs a distributed architecture, distributing computing tasks across a network of edge devices. This decentralization optimizes resource utilization and scalability.
- **Efficiency:** By processing data locally, edge computing reduces the burden on centralized cloud servers, leading to more efficient use of network bandwidth and reduced operational costs.
- **Real-Time Processing:** Edge computing supports real-time data processing and analytics, enabling immediate responses to critical events or conditions. This is essential for applications like remote monitoring, smart grids, and augmented reality.
- **Security and Privacy:** Edge computing enhances data security and privacy by keeping sensitive information closer to its source, reducing exposure to potential security breaches during data transmission.

Edge computing is not a replacement for cloud computing but rather a complementary approach [12]. Both technologies can work in tandem, with edge devices handling time-sensitive tasks and the cloud managing more resource-intensive processes and long-term data storage [12], [14].

This emerging technology has found applications in various fields, including:

- **IoT and Smart Devices:** Edge computing is integral to the Internet of Things (IoT) ecosystem, enabling smart devices to process data locally and make rapid decisions.
- **Telecommunications:** Telecom networks benefit from edge computing for tasks like content caching, network optimization, and low-latency services.

- **Healthcare:** In healthcare, edge computing supports real-time patient monitoring, data analysis, and diagnosis.
- **Manufacturing:** Industrial automation and robotics leverage edge computing for faster decision-making on the factory floor.
- **Autonomous Vehicles:** Edge computing is crucial for self-driving cars, allowing them to process sensor data in real-time for safe navigation.

Overall, edge computing is a promising new technology that can improve the performance, reliability, and security of a wide range of applications.

In the subsequent sections, we will explore how the fusion of edge computing and blockchain technology can unlock new possibilities in various industries.

III. INTEGRATION OF BLOCKCHAIN AND EDGE COMPUTING: KEY BENEFITS

The integration of blockchain and edge computing ushers in a host of transformative advantages for modern data-driven systems [15]. First and foremost, data security experiences a significant boost [17]. Leveraging blockchain's decentralized, tamper-resistant ledger and edge computing's localized data processing, the integrity and confidentiality of data are fortified. Unauthorized access and tampering become formidable hurdles, necessitating consensus from network participants, particularly vital in data-sensitive sectors like healthcare, finance, and supply chain management [16].

Moreover, this integration enhances system reliability substantially. The inherent decentralization of edge computing reduces dependency on a single central server or data center, a synergy that harmonizes well with blockchain's reliability mechanisms. Even in the face of isolated node or device failures, system functionality remains uninterrupted, a critical trait for applications such as autonomous vehicles and critical infrastructure [17][16].

Simultaneously, application performance receives a considerable uplift, as edge computing reduces latency by processing data closer to its source [11]. This proximity expedites real-time decision-making in applications such as augmented reality, remote monitoring, and smart grids [18].

Comparatively, when we consider integrating blockchain with cloud computing, some differences emerge [19]:

- **Security:** While blockchain integration with edge computing offers a high level of security through decentralization, combining blockchain with cloud computing relies on centralized server security, requiring stringent measures.
- **Reliability:** The reliance on centralized data centers and data transmission across networks in cloud computing may affect its reliability, unlike the decentralized edge nodes of edge computing, which ensure operations even in individual contract or device failures.
- **Application Performance:** Edge computing's local data processing significantly improves application

performance, reducing latency. In contrast, cloud computing applications may experience added latency due to data transmission to remote data centers.

- **Cost Efficiency:** Integrating blockchain with edge computing greatly reduces operational costs by minimizing reliance on cloud infrastructure and lowering data transfer expenses. On the other hand, cloud computing involves costs related to running data centers and cloud storage, incurring additional expenses.

In Addition, the integration of blockchain with edge computing can be compared to the integration of blockchain with cloud computing [19]. The following table summarizes the key differences between these two approaches:

TABLE I. COMPARISON OF BLOCKCHAIN INTEGRATION APPROACHES

Feature	Blockchain and Edge Computing	Blockchain and Cloud Computing
Data Security	Enhanced [6]	Reduced [9]
System Reliability	Enhanced [2]	Reduced [8]
Application Performance	Enhanced [5]	Unaffected [20]
Cost-Efficiency	Enhanced [2]	Unaffected [8]
Suitable use cases	Data-sensitive applications, applications requiring real-time processing, applications with high security requirements [20]	Applications that require a lot of computing power, applications that need to store large amounts of data [20]

As shown in the Table I, the integration of blockchain with edge computing offers a number of advantages over the integration of blockchain with cloud computing. Specifically, it provides better data security, system reliability, and application performance. Additionally, it is more suitable for use cases that require real-time processing and high security requirements.

Last but certainly not least, the integration of blockchain and edge computing delivers compelling cost-efficiency benefits [2]. By minimizing reliance on extensive cloud infrastructure and associated data transmission costs, operational expenses are significantly reduced. This is further augmented by heightened system reliability, which helps mitigate revenue losses associated with system failures [2].

To summarize, the integration of blockchain and edge computing presents an enticing proposition for businesses across diverse sectors. It promises heightened operational efficiency, fortified data privacy, and robust data management practices, all while positioning itself as a pioneering solution at the intersection of security, reliability, performance, and cost-effectiveness in the evolving landscape of data-driven systems.

IV. CHALLENGES AND CLOUD INDEPENDENCE IN BLOCKCHAIN-EDGE INTEGRATION

A. Challenges in Integrating Blockchain and Edge Computing

The integration of blockchain and edge computing presents significant potential for enhancing both security and

performance in modern data systems [6]. However, several critical challenges—spanning technical, performance, security, and regulatory dimensions—must be addressed to fully realize this potential. One of the foremost technical challenges is scalability. As decentralized networks grow to accommodate increasing data demands, the complexity of maintaining data integrity and security across numerous nodes becomes more pronounced. This issue is particularly salient in edge computing environments, where devices often lack the processing power and storage capacity required for executing resource-intensive blockchain consensus mechanisms, such as Proof of Work (PoW) [21],[4]. Moreover, the inherently decentralized structure of blockchain introduces latency challenges, which run counter to the low-latency requirements of edge computing. The time required for transaction verification and block validation can degrade performance, especially in latency-sensitive applications such as the Internet of Things (IoT) and real-time data analytics [22].

Security concerns also arise due to the movement of data between edge and cloud environments, where the risk of cyberattacks increases during transmission [5]. While blockchain’s decentralized architecture enhances security by distributing control, edge devices typically lack the robust security mechanisms available in cloud-based systems, making them more susceptible to threats [23]. Managing identity and access in decentralized systems further complicates this

challenge. Although cloud-based solutions may offer strong identity management capabilities, their reliance on centralized systems could undermine the decentralized ethos of blockchain, raising issues of dependency and control.

From a regulatory perspective, compliance with frameworks such as the General Data Protection Regulation (GDPR) adds another layer of complexity [5]. The decentralized nature of blockchain makes it difficult to pinpoint where and how data is stored, complicating efforts to ensure compliance with data privacy laws. This issue becomes even more challenging in cross-border scenarios, where legal frameworks may vary significantly [5]. As a result, questions surrounding data ownership and control become especially pertinent in industries governed by strict regulatory requirements. Addressing these technical, performance, security, and regulatory challenges is essential for unlocking the full potential of blockchain and edge computing in modern data systems.

Several studies have sought to address these challenges through various approaches that integrate blockchain with edge and cloud computing, particularly within IoT environments. Table II summarizes key contributions from these works, highlighting how they tackle issues related to scalability, security, decentralization, and performance in blockchain-enabled systems.

TABLE II. KEY CONTRIBUTIONS IN BLOCKCHAIN AND EDGE COMPUTING INTEGRATION

Ref.	Key Contributions	Layered Architecture	Cryptocurrency Involvement	Blockchain Decentralization	IoT Applications	Cloud of Things	Cloud Computing	Journal
[5]	Exploring the Integration of Edge Computing and Blockchain to Enhance IoT Systems and Address Key Challenges in Security and Efficiency	✓	X	✓	✓	✓	✓	ScienceDirect
[1]	Surveying the Integration of Blockchain and Edge Computing to Enhance Resource Utilization and Security in IoT Applications	X	X	X	✓	✓	✓	ScienceDirect
[6]	EdgeBoT as a Smart Contracts-Based Platform to Enhance Data Ownership and Privacy in IoT Through Blockchain Technology	X	X	✓	✓	✓	✓	Sensors

[25]	Proposing a Scalable and Secure Cloud Architecture to Enhance IoT Integration with Cryptographic Techniques for Improved Multi-User Access and Data Security	✓	X	✓	✓	X	✓	IEBEE Access
[4]	The Convergence of Blockchain and Edge of Things Exploring Opportunities, Applications, and Security Challenges in the BEoT Paradigm	X	X	X	X	X	✓	IEBEE IoTJ
[26]	The Potential of Blockchain Technology in Integrated IoT Networks for Scalable Intelligent Transportation Systems in India	X	X	X	✓	X	✓	ScienceDirect
[3]	Advancements in Edge Computing: Integrating AI and Blockchain for Enhanced Performance in Maritime and Aerial Systems	X	X	✓	✓	✓	✓	IEBEE Access
[16]	A Blockchain-Assisted Handover Authentication Scheme for Intelligent Telehealth Systems in Multi-Server Edge Computing Environments	X	✓	X	✓	X	✓	ScienceDirect
[27]	A Novel Trust-Aware Blockchain-Based Framework for Enhancing Privacy and Security in Decentralized IoT Applications	X	X	X	✓	✓	✓	Electronics
[28]	Analyzing the Integration of Blockchain in IoT and Healthcare: Enhancing Data Security and Management Strategies	X	X	X	✓	X	✓	ScienceDirect

[18]	Integrating Blockchain and Federated Learning for Enhanced Security and Privacy in Smart Healthcare with a Novel Conceptual Framework	X	✓	X	✓	X	✓	IEEE Internet of Things Journal
[13]	The Evolution of Mobile Cloud Computing and Edge Computing for Enhanced Mobile Applications and Open Research Challenges	X	X	✓	✓	X	✓	Springer
[29]	Analyzing Security Challenges and Solutions for Data Privacy in Cloud-IoT Environments with Insights into Emerging Technologies	X	X	X	X	X	✓	Springer
[7]	Proposing a Unified Control Framework for Enhancing Railway Communications Through Integration of Advanced Technologies in the Era of 5G and Future 6G	X	X	X	X	X	✓	IEEE Access
[30]	Proposing a Blockchain-Based Cloud Integrated IoT Application for Enhanced Security and Intruder Detection in Challenging Environments	X	✓	✓	X	X	✓	Springer
[31]	Integrating Blockchain and Edge Computing to Create a Secure, Scalable Architecture for Data Processing in Industry 4.0 Applications	X	X	✓	X	X	✓	Springer

Several studies contributions illustrate a variety of approaches to integrating blockchain with edge computing, addressing challenges such as scalability, security, and decentralization. By leveraging multi-layered architectures and optimizing resource allocation, these works enhance system performance, security, and regulatory compliance.

Based on the studies presented in Table II, it is evident that while various approaches have been proposed to tackle challenges like scalability, security, and performance, the role of

cloud computing remains a consistent element across all studies. This pervasive presence of the cloud raises an important question: Can blockchain data management be fully achieved in edge computing without cloud support? In other words, is the cloud indispensable in all cases of integrating blockchain and edge computing?

B. Blockchain-Edge Computing Integration and Cloud Support

The integration of blockchain with edge computing has

gained considerable attention due to its potential to address challenges such as latency reduction and enhanced security in decentralized systems. However, upon analyzing recent studies, it becomes clear that cloud computing plays a crucial role in most cases of blockchain-edge integration. While edge computing is effective for real-time data processing and localized decision-making, cloud support is often required for tasks that demand higher computational power, scalability, and long-term data storage.

A graphical analysis based on the studies presented in Table II emphasizes the significant presence of cloud computing in blockchain-edge integration research. As demonstrated in Fig. 1, a substantial proportion of the studies rely on cloud services to complement the resource-constrained nature of edge devices. The cloud not only provides additional computational resources for tasks like blockchain mining, transaction verification, and smart contract execution, but it also enables efficient data storage and management for decentralized applications.

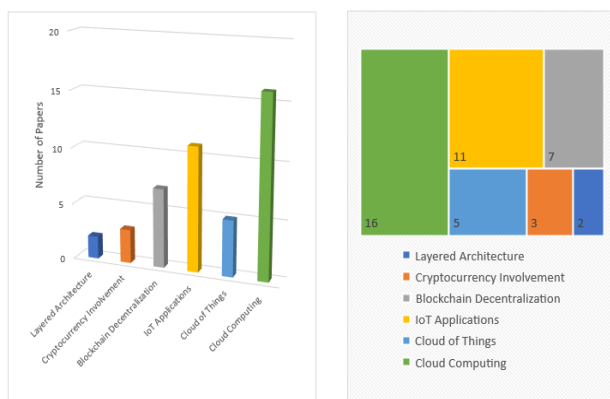


Fig. 1. Prevalence of cloud support in blockchain-edge computing integration studies.

As illustrated in Fig. 1, the majority of blockchain-edge computing integrations incorporate cloud support, reflecting its indispensable role in managing the complexity and resource demands of decentralized networks. For example, studies by Tri et al. [5] and Yang et al. [2] showcase how cloud infrastructure serves as the backbone for scaling blockchain operations, ensuring that the limitations of edge devices do not hinder overall system performance. The cloud efficiently offloads computationally intensive tasks, such as consensus mechanisms, while enabling edge devices to focus on real-time data processing and localized operations.

Despite the strong reliance on cloud computing in many cases, there are specific scenarios where blockchain and edge computing can be successfully integrated without the need for cloud services. These cases generally arise in smaller-scale, localized applications, where the resource demands of blockchain operations are relatively low, and large-scale data storage or significant computational power is not required.

1) *Localized IoT networks*: In small, self-contained IoT environments—such as smart homes or small industrial setups—blockchain can be integrated with edge devices to manage secure transactions and ensure data integrity without the need for cloud support. In these scenarios, lightweight

consensus mechanisms like Proof of Authority (PoA) or Practical Byzantine Fault Tolerance (PBFT) can be efficiently handled by edge devices, thus eliminating the need for external cloud resources [32].

2) *Decentralized autonomous systems*: Some decentralized systems, such as autonomous drones or vehicular networks, can operate blockchain-based frameworks using only edge computing. These systems typically rely on localized blockchain networks, where each node (e.g., a drone or vehicle) has sufficient computational power to process transactions and validate blocks, avoiding the latency and delays introduced by cloud-based solutions [20].

3) *Data sovereignty and privacy-centric applications*: In highly sensitive environments, such as healthcare or military applications, where data sovereignty and privacy are paramount, blockchain and edge computing can be combined to maintain strict control over data without transmitting it to cloud servers. These use cases focus on local data processing and handling, ensuring privacy and removing reliance on external cloud providers [5].

These examples demonstrate that while cloud support is beneficial in many cases, it is not always essential for blockchain-edge integration. In environments where computational demands are modest and concerns about latency or privacy are significant, blockchain and edge computing can function effectively without cloud involvement. However, for most large-scale applications, particularly those requiring scalability, redundancy, or complex data management, cloud computing remains a critical component, enabling seamless and efficient integration between blockchain and edge computing.

V. ETHEREUM IN EDGE COMPUTING: A COMPREHENSIVE CASE STUDY

The integration of Ethereum within edge computing environments presents a powerful solution for decentralized, real-time applications. By leveraging edge computing, Ethereum can enhance performance and scalability through localized data processing, thereby reducing latency and alleviating network congestion [1][7]. This section explores the technical aspects, advantages, concerns, and potential future developments for Ethereum as a blockchain platform deployed at the edge [3]. Furthermore, it highlights how decentralized financial systems can capitalize on the processing capabilities of edge computing to provide more efficient and secure solutions, ultimately paving the way for innovative applications in IoT and beyond [8]. By addressing these factors, Ethereum demonstrates its capacity to evolve within edge environments, enhancing its role in the decentralized landscape.

A. Technical Advantages of Ethereum in Edge Computing

Integrating Ethereum with edge computing brings unique technical benefits that enhance the efficiency, scalability, and security of decentralized applications. Ethereum’s design, coupled with edge computing’s localized processing capabilities, makes it particularly well-suited for applications that require low latency, real-time processing, and energy efficiency. Key technical advantages include:

1) *Real-Time processing and reduced latency:*

a) *Localized transaction validation:* By handling transactions closer to the data source, edge devices can validate and process Ethereum transactions locally, which greatly reduces latency compared to centralized blockchain processing [19]. This is especially valuable for applications in IoT and smart city infrastructure where rapid decision-making is critical.

b) *Enhanced decentralized applications (dApps):* Real-time data processing at the edge allows decentralized applications to operate with faster response times, improving user experience in applications like financial trading, supply chain management, and decentralized exchanges (DEXs) that rely on immediate updates [2].

c) *Smart contract execution at the edge:* Ethereum's capability to execute smart contracts can be enhanced in edge environments, where real-time contract execution reduces the time required for transactions to finalize. This brings significant improvements for IoT applications that rely on automated responses based on data analytics.

2) *Energy efficiency through Proof-of-Stake (PoS):*

a) *Reduced resource consumption:* Ethereum's shift from Proof-of-Work (PoW) to Proof-of-Stake (PoS) consensus drastically lowers the computational and energy demands on devices participating in the network [3], [2]. This reduction is crucial for edge devices, which typically have limited resources compared to traditional data centers.

b) *Compatibility with resource-constrained edge devices:* PoS allows edge devices to contribute to the network without requiring the intensive hardware needed for PoW mining, making it feasible for smaller, more energy-efficient devices to play an active role in transaction validation and block creation within the Ethereum network [2].

c) *Support for sustainability goals:* By using PoS, Ethereum aligns well with the sustainability objectives of many IoT and smart infrastructure projects, where energy consumption is a key concern [2][5].

3) *Scalability with layer two solutions:*

a) *Layer 2 offloading:* Ethereum's Layer 2 scaling solutions, such as rollups and zk-Rollups, enable transactions to be processed off-chain while still anchored to the main Ethereum blockchain for security. This alleviates congestion on the main network, making it easier for edge devices to handle high transaction volumes without compromising performance [6].

b) *Improved throughput:* By batching multiple transactions off-chain, Layer 2 solutions significantly improve throughput, making Ethereum capable of handling more transactions per second (TPS) without overwhelming edge devices [16]. This is particularly useful in IoT ecosystems, where numerous small transactions are generated by devices.

c) *Interoperability with other blockchains:* Many Layer 2 solutions on Ethereum are designed with interoperability in mind, allowing edge devices in one network to interact with other blockchain ecosystems. This cross-chain compatibility

fosters greater flexibility for decentralized applications, particularly in settings like supply chain networks and logistics [2][8].

4) *Security benefits of decentralized processing:*

a) *Enhanced data security:* The decentralized nature of Ethereum, combined with edge computing, strengthens data security by processing and storing data closer to the source. This decentralized architecture reduces the risk of centralized points of failure and data breaches, which is especially beneficial for sensitive applications such as healthcare and finance [28].

b) *Zero-Knowledge proofs (zk-SNARKs) for privacy:* Ethereum's zk-SNARK technology enables data verification without revealing sensitive information, supporting privacy in edge applications where personal or confidential data may be processed [15]. This ensures that data privacy is maintained while still benefiting from Ethereum's secure transaction model.

c) *Tamper-Resistant IoT networks:* By deploying Ethereum nodes on edge devices, IoT networks gain resilience against tampering and unauthorized access, as data must undergo consensus verification before being accepted. This adds a strong layer of security to edge-based IoT environments [22].

In summary, Ethereum's adaptable architecture, energy-efficient consensus mechanisms, and advanced Layer 2 scaling solutions make it exceptionally well-suited for deployment in edge computing environments, where real-time processing, scalability, and security are paramount for next-generation decentralized applications.

B. *Comparison of Ethereum with Other Cryptocurrencies in Edge Environments*

Deploying blockchain in edge computing requires energy efficiency, low-latency processing, and scalability [2]. While Ethereum's features make it suitable for edge computing, a comparison with Bitcoin and Polkadot reveals key distinctions, as shown in Table III, which highlights the key comparisons of Ethereum, Bitcoin, and Polkadot for edge computing applications.

TABLE III. KEY COMPARISONS OF ETHEREUM, BITCOIN, AND POLKADOT FOR EDGE COMPUTING APPLICATIONS

Feature	Ethereum	Bitcoin	Polkadot
Consensus Mechanism	Proof-of-Stake (PoS)	Proof-of-Work (PoW)	Nominated Proof-of-Stake (NPoS)
Energy Efficiency	High, low-power PoS [1]	Low, resource-intensive [2]	Moderate, optimized for PoS [3]
Smart Contract Support	Extensive (EVM)	Limited scripting [4]	Multi-chain smart contracts
Layer 2 Scaling	Robust (Rollups) [5]	Limited	Cross-chain scalability (parachains)
Latency Sensitivity	Optimized for real-time dApps [6]	Slower confirmations [7]	Optimized for cross-chain processing

Ethereum's Proof-of-Stake (PoS) consensus mechanism greatly reduces energy consumption, making it compatible with

edge device constraints, where power and resources are often limited. Bitcoin's Proof-of-Work (PoW), in contrast, is computationally demanding and thus unsuitable for resource-constrained environments. Polkadot's Nominated Proof-of-Stake (NPoS) model is similarly efficient and well-suited for multi-chain edge networks, where various blockchains need to operate seamlessly.

C. Proposed Algorithm for Ethereum-Based Transactions in Edge Environments

Some studies have introduced algorithms to optimize blockchain integration in edge computing, focusing on reducing latency and managing data consistency. Common approaches include distributed consensus mechanisms [1], off-chain scaling solutions like Plasma and state channels [2], and layered architectures that rely on cloud support for intensive processing [3]. While effective, these methods often depend on centralized infrastructures, potentially limiting decentralization.

Our proposed algorithm presents a decentralized transaction validation framework tailored for Ethereum-based systems in edge environments. By utilizing local consensus among edge nodes and Layer 2 scaling, the algorithm minimizes cloud dependency, enabling autonomous, secure transaction processing directly at the edge.

Algorithm 1: Decentralized Transaction Validation at the Edge Using Ethereum

Objective: Efficiently process transactions on Ethereum using edge devices while maintaining security and minimizing latency.

Input: Transaction data (Tx), Node ID (Edge Node), Blockchain state (S)

Output: Validated transaction and updated blockchain state (S')

1. Edge device (Node) receives a transaction request (Tx).
2. Node verifies transaction data (Tx) using Ethereum's cryptographic validation method [5].
3. If the transaction is valid:
 1. Node checks its local blockchain state (S) to ensure consistency.
 2. Transaction is added to a local temporary block.
4. The temporary block is broadcasted to nearby nodes in the edge network for additional validation (Consensus) [6].
5. Upon achieving consensus among edge devices, the validated block is appended to the blockchain.
6. If Layer 2 rollup is enabled, the batch of transactions is compressed and sent to the main Ethereum chain for final settlement [7].
7. **Output:** Updated blockchain state (S') is stored across all edge nodes.

This algorithm utilizes the proximity of edge devices for transaction validation, minimizing reliance on centralized servers or cloud infrastructures. By incorporating Layer 2 scaling solutions, such as rollups, the algorithm offloads part of the computational workload to off-chain solutions, optimizing resource use in constrained edge devices. Consensus

mechanisms within the edge network ensure data consistency before broadcasting the validated block to the larger Ethereum blockchain, balancing security and speed [2].

D. Regulatory and Privacy Concerns

While the technical feasibility of deploying Ethereum in edge environments is promising, regulatory challenges must also be addressed. Ethereum's decentralized nature complicates data privacy and ownership, particularly when considering laws like the General Data Protection Regulation (GDPR) in Europe. Key regulatory issues include:

- **Data Privacy:** Since Ethereum transactions are publicly visible, storing sensitive data (e.g., personal or medical information) on the blockchain may violate privacy regulations. Solutions like Zero-Knowledge Proofs (zk-SNARKs), which allow verification of transactions without revealing sensitive information, could mitigate this issue [9].
- **Cross-Jurisdictional Compliance:** With edge devices deployed globally, ensuring compliance with different legal frameworks across borders poses a significant challenge [11].
- **Data Sovereignty:** Edge environments often operate in localized settings, and the transmission of data to global blockchains raises concerns about who controls and owns that data [11].

E. Future Research Directions

The integration of Ethereum within edge computing environments presents promising opportunities, yet several key challenges in efficiency, security, and scalability remain to be addressed. Future research directions that may significantly advance this field include the following:

1) *Hybrid architectures:* Investigating hybrid architectures that combine both edge and cloud resources could enhance the performance of Ethereum-based applications in edge environments. A proposed approach involves handling time-sensitive, low-latency tasks, such as initial transaction validations, at the edge, while offloading computationally intensive tasks (e.g., complex smart contract execution and large-scale data analysis) to the cloud. This distribution strategy optimizes the limited resources of edge devices while leveraging cloud computational power to handle more demanding tasks, leading to more efficient operations across edge-cloud ecosystems [24].

2) *Improved consensus mechanisms:* To promote scalability and energy efficiency in edge environments, developing lightweight consensus protocols tailored for edge computing is essential. Traditional consensus algorithms, such as Proof of Work (PoW), are highly resource-intensive and unsuitable for resource-constrained edge devices. Future research should explore alternative protocols, such as Proof of Authority (PoA) or adapted Byzantine Fault Tolerance (BFT) models, which could reduce computational overhead and energy requirements while maintaining security. Such protocols would enable resource-constrained edge devices to

participate effectively in Ethereum networks without compromising network performance [13].

3) *Security enhancements*: Enhancing security for Ethereum-edge networks is critical, given the vulnerabilities of edge devices to cyber threats. Research into advanced cryptographic techniques, including secure multiparty computation and zero-knowledge proofs, may strengthen data privacy and integrity within decentralized edge networks. Additionally, exploring post-quantum cryptographic methods is crucial for ensuring resilience against potential future threats from quantum computing, ultimately securing Ethereum-edge networks for long-term operation [2].

4) *Decentralized data storage solutions*: The adoption of decentralized storage solutions offers a pathway to securely manage and distribute data across edge environments. Technologies such as the InterPlanetary File System (IPFS) provide secure, distributed data storage without centralized dependencies. Integrating IPFS with Ethereum could enable edge networks to store data with higher redundancy and fault tolerance, even in disconnected or remote environments [1]. This is particularly advantageous for Internet of Things (IoT) ecosystems, where data generated by edge devices requires secure, decentralized storage and accessibility [2].

Expanding research in these areas could substantially enhance the efficiency, scalability, and security of Ethereum applications within edge computing environments. By addressing these challenges, researchers can contribute to building a robust foundation for decentralized applications capable of operating reliably across distributed edge-cloud ecosystems.

VI. CONCLUSION

The integration of blockchain and edge computing presents a transformative opportunity to improve data security, computing efficiency, and data privacy across various industries, particularly in IoT environments where real-time data processing and secure transactions are crucial. This paper has explored the foundational principles of these two technologies, examined their synergies, and highlighted the significant benefits of their convergence. By leveraging blockchain's decentralized, tamper-resistant structure alongside edge computing's ability to process data locally, this integration promises substantial advancements in performance, reducing latency and enhancing security in decentralized networks. Furthermore, the Ethereum-based case study offered practical insights into how blockchain can be deployed in edge environments, illustrating both its feasibility and the challenges that arise in ensuring efficiency and scalability.

The contributions of this research lie in its systematic analysis of the integration of blockchain and edge computing, with a particular focus on the practical implications of their convergence. The study provides valuable perspectives on the ways these technologies can enhance security through blockchain's immutability and edge computing's localized data processing, while also addressing the challenges related to computational demands, scalability, and managing decentralized networks. The Ethereum case study served as a

critical example of the potential applications in edge environments, but it also underscored the importance of addressing challenges like computational load, network management, and the reliance on cloud infrastructure for certain tasks. This paper contributes to the ongoing dialogue in the field by identifying these key challenges and providing a foundation for future studies focused on optimizing these systems.

Despite the promising potential of blockchain and edge computing, this research is not without its limitations. The case study was limited to Ethereum, and while it provided useful insights, it may not fully represent the diverse array of blockchain platforms with differing consensus mechanisms or resource requirements. Additionally, the study focused primarily on the computational aspects of blockchain at the edge, leaving out considerations around hardware diversity in edge devices, which could impact the integration's performance across different environments. While the hybrid edge-cloud architecture discussed in this paper provides a practical solution, the potential challenges of security, privacy, and the added complexity of cloud dependencies need further investigation, particularly when considering large-scale deployments.

Looking ahead, future research should address several critical areas to optimize the integration of blockchain and edge computing. First, there is a need for the development of lightweight consensus mechanisms that can reduce the computational burden on edge devices, ensuring that blockchain systems remain secure and decentralized while being efficient enough to operate in resource-constrained environments. Second, scalable architectures that can support the growing demands of decentralized edge computing systems should be explored. These architectures should balance the need for real-time processing with the constraints of limited edge resources, while also maintaining the integrity of the blockchain. Finally, further exploration of data privacy solutions is essential, particularly in decentralized systems. Techniques such as zero-knowledge proofs and advanced encryption methods could help ensure privacy without sacrificing performance, enabling secure blockchain operations in edge environments. Addressing these areas will be key to advancing the integration of blockchain and edge computing, enabling the development of secure, efficient, and scalable decentralized systems for future data-driven applications.

REFERENCES

- [1] Xue, H., Chen, D., Zhang, N., Dai, H.-N., & Yu, K. (2022). Integration of blockchain and edge computing in Internet of Things: A survey. arXiv. <https://arxiv.org/abs/2205.13160>.
- [2] Yang, R., Yu, F. R., Si, P., Yang, Z., & Zhang, Y. (2019). Integrated Blockchain and Edge Computing Systems: A Survey, Some Research Issues and Challenges. *IEEE Communications Surveys & Tutorials*, 21(2), 1508-1532. <https://doi.org/10.1109/COMST.2019.2894727>
- [3] A. Alanhdi and L. Toka, "A Survey on Integrating Edge Computing With AI and Blockchain in Maritime Domain, Aerial Systems, IoT, and Industry 4.0," in *IEEE Access*, vol. 12, pp. 28684-28709, 2024, doi: 10.1109/ACCESS.2024.3367118.
- [4] Gadekallu, T. R., Pham, Q.-V., Nguyen, D. C., Maddikunta, P. K. R., Deepa, N., B. P., Pathirana, P. N., Zhao, J., & Hwang, W.-J. (2021). Blockchain for edge of things: Applications, opportunities, and challenges. arXiv. <https://arxiv.org/abs/2110.05022>.
- [5] Tri Nguyen, Huong Nguyen, Tuan Nguyen Gia, Exploring the integration of edge computing and blockchain IoT: Principles, architectures, security,

- and applications, *Journal of Network and Computer Applications*, Volume 226, 2024, 103884, ISSN 1084-8045, <https://doi.org/10.1016/j.jnca.2024.103884>.
- [6] Nawaz, A., Peña Queralta, J., Guan, J., Awais, M., Nguyen Gia, T., Bashir, A.K., Kan, H., & Westerlund, T. (2020). Edge Computing to Secure IoT Data Ownership and Trade with the Ethereum Blockchain. *Sensors*, 20(14), 3965. <https://doi.org/10.3390/s20143965>
- [7] Narouwa, M., Mendiboure, L., Badis, H., Maaloul, S., Berbineau, M., & Langar, R. (2024). Enabling Network Technologies For Flexible Railway Connectivity. *IEEE Access*, 12, 151532-151547. <https://doi.org/10.1109/ACCESS.2024.3479879>.
- [8] W. Jaafar, K. Jean Romeo Beyara, I. Aouini, J. Ben Abderrazak and H. Yanikomeroglu, "On the Deployment of Blockchain in Edge Computing Wireless Networks," 2022 IEEE 11th International Conference on Cloud Networking (CloudNet), Paris, France, 2022, pp. 168-176, doi: 10.1109/CloudNet55617.2022.9978739.
- [9] Bentayeb, Youness & Badir, Hassan & En-Nahnahi, Nouredine. (2023). Blockchain-Based Cloud Computing: Model-Driven Engineering Approach. 10.1007/978-3-031-26384-2_55.
- [10] Nakamoto, S. (2008). *Bitcoin: A Peer-to-Peer Electronic Cash System*. Available at: <https://bitcoin.org/bitcoin.pdf>
- [11] B. C. Girish Kumar, P. Nand and V. Bali, "Opportunities and Challenges of Blockchain Technology for Tourism Industry in Future Smart Society," 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), Sonepat, India, 2022, pp. 318-323, doi: 10.1109/CCICT56684.2022.00065.
- [12] Yu, W., et al.: A survey on the edge computing for the Internet of Things. *IEEE Access* 6, 6900–6919 (2018).
- [13] Dimou, A., Iliopoulos, C., Polytidou, E., Dhurandher, S.K., Papadimitriou, G., Nicopolitidis, P. (2022). A Comprehensive Review on Edge Computing: Focusing on Mobile Users. In: Nicopolitidis, P., Misra, S., Yang, L.T., Zeigler, B., Ning, Z. (eds) *Advances in Computing, Informatics, Networking and Cybersecurity*. Lecture Notes in Networks and Systems, vol 289. Springer, Cham. https://doi.org/10.1007/978-3-030-87049-2_30
- [14] K. Cao, Y. Liu, G. Meng and Q. Sun, "An Overview on Edge Computing Research," in *IEEE Access*, vol. 8, pp. 85714-85728, 2020, doi: 10.1109/ACCESS.2020.2991734.
- [15] L. Fotia, F. C. Delicato and G. Fortino, "Integrating Blockchain and Edge Computing in Internet of Things: Brief Review and Open Issues," 2021 International Conference on Cyber-Physical Social Intelligence (ICCSI), Beijing, China, 2021, pp. 1-6, doi: 10.1109/ICCSI53130.2021.9736164.
- [16] Wenming Wang, Haiping Huang, Lingyan Xue, Qi Li, Reza Malekian, Youzhi Zhang, Blockchain-assisted handover authentication for intelligent telehealth in multi-server edge computing environment, *Journal of Systems Architecture*, Volume 115, 2021, 102024, ISSN 1383-7621.
- [17] A. C. Baktir, A. Ozgovde and C. Ersoy, "How Can Edge Computing Benefit From Software-Defined Networking: A Survey, Use Cases, and Future Directions," in *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2359-2391, Fourthquarter 2017, doi: 10.1109/COMST.2017.2717482.
- [18] R. Myrzhoshova, S. H. Alsamhi, A. V. Shvetsov, A. Hawbani and X. Wei, "Blockchain Meets Federated Learning in Healthcare: A Systematic Review With Challenges and Opportunities," in *IEEE Internet of Things Journal*, vol. 10, no. 16, pp. 14418-14437, 15 Aug.15, 2023, doi: 10.1109/JIOT.2023.3263598.
- [19] T. R. Gadekallu et al., "Blockchain for Edge of Things: Applications, Opportunities, and Challenges," in *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 964-988, 15 Jan.15, 2022, doi: 10.1109/JIOT.2021.3119639.
- [20] Gao, Q., Xiao, J., Cao, Y., Deng, S., Ouyang, C., & Feng, Z. (2023). Blockchain-based collaborative edge computing: Efficiency, incentive and trust. *Journal of Cloud Computing*, 12(1), 72. <https://doi.org/10.1186/s13677-023-00452-4>.
- [21] Oliveira, Miguel, Sumit Chauhan, Filipe Pereira, Carlos Felgueiras, and David Carvalho. 2023. "Blockchain Protocols and Edge Computing Targeting Industry 5.0 Needs" *Sensors* 23, no. 22: 9174. <https://doi.org/10.3390/s23229174>
- [22] Yuanxing Yin, Xinyu Wang, Huan Wang, Baoli Lu, Application of edge computing and IoT technology in supply chain finance, *Alexandria Engineering Journal*, Volume 108, 2024, Pages 754-763, ISSN 1110-0168, <https://doi.org/10.1016/j.aej.2024.09.016>.
- [23] Bentayeb, Youness & Badir, Hassan. (2024). Blockchain-Based Cloud Computing: A Comparative Study of BoC, CoB, and MBC. 255-260. 10.1007/978-3-031-52388-5_24.
- [24] Andriulo, F.C.; Fiore, M.; Mongiello, M.; Traversa, E.; Zizzo, V. Edge Computing and Cloud Computing for Internet of Things: A Review. *Informatics* 2024, 11, 71. <https://doi.org/10.3390/informatics11040071>
- [25] R. R. Irshad et al., "IoT-Enabled Secure and Scalable Cloud Architecture for Multi-User Systems: A Hybrid Post-Quantum Cryptographic and Blockchain-Based Approach Toward a Trustworthy Cloud Computing," in *IEEE Access*, vol. 11, pp. 105479-105498, 2023, doi: 10.1109/ACCESS.2023.3318755.
- [26] Arya Kharche, Sanskar Badholia, Ram Krishna Upadhyay, Implementation of blockchain technology in integrated IoT networks for constructing scalable ITS systems in India, *Blockchain: Research and Applications*, Volume 5, Issue 2, 2024, 100188, ISSN 2096-7209, <https://doi.org/10.1016/j.bcr.2024.100188>.
- [27] Al Hwaitat, Ahmad K., Mohammed Amin Almaiah, Aitizaz Ali, Shaha Al-Otaibi, Rima Shishakly, Abdalwali Lutfi, and Mahmaod Alrawad. 2023. "A New Blockchain-Based Authentication Framework for Secure IoT Networks" *Electronics* 12, no. 17: 3618. <https://doi.org/10.3390/electronics12173618>
- [28] Endale Mitiku Adere, Blockchain in healthcare and IoT: A systematic literature review, *Array*, Volume 14, 2022, 100139, ISSN 2590-0056, <https://doi.org/10.1016/j.array.2022.100139>.
- [29] Pathak, M., Mishra, K.N. & Singh, S.P. Securing data and preserving privacy in cloud IoT-based technologies an analysis of assessing threats and developing effective safeguard. *Artif Intell Rev* 57, 269 (2024). <https://doi.org/10.1007/s10462-024-10908-x>
- [30] Rupa, Ch & Srivastava, Gautam & Gadekallu, Thippa & Reddy, Praveen & Bhattacharya, Sweta. (2021). A Blockchain Based Cloud Integrated IoT Architecture Using a Hybrid Design. 10.1007/978-3-030-67540-0_36.
- [31] Sittón-Candanedo, I. (2020). RETRACTED CHAPTER: A New Approach: Edge Computing and Blockchain for Industry 4.0. In: Herrera-Viedma, E., Vale, Z., Nielsen, P., Martin Del Rey, A., Casado Vara, R. (eds) *Distributed Computing and Artificial Intelligence*, 16th International Conference, Special Sessions. DCAI 2019. *Advances in Intelligent Systems and Computing*, vol 1004. Springer, Cham. https://doi.org/10.1007/978-3-030-23946-6_25
- [32] Bo Gan, Yaojie Wang, Qiwu Wu, Yang Zhou, Lingzhi Jiang, EIoT-PBFT: A multi-stage consensus algorithm for IoT edge computing based on PBFT, *Microprocessors and Microsystems*, Volume 95, 2022, 104713, ISSN 0141-9331, <https://doi.org/10.1016/j.micpro.2022.104713>.

Enhancing COVID-19 Detection in X-Ray Images Through Deep Learning Models with Different Image Preprocessing Techniques

Ahmad Nuruddin bin Azhar¹, Nor Samsiah Sani², Liu Luan Xiang Wei³

Center for Artificial Intelligence Technology-Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia, Selangor, 43600, Malaysia^{1,2,3}

Abstract—The identification of COVID-19 using chest X-ray (CXR) images plays a critical role in managing the pandemic by providing a rapid, non-invasive, and accessible diagnostic tool. This study evaluates the impact of different image preprocessing techniques on the performance of deep learning models for COVID-19 classification based on COVID-19 Radiography Database, which includes 10,192 normal CXR images, 6012 lung opacity (non-COVID lung infection) images, and 1345 viral pneumonia images. Along with the images, corresponding lung masks are also included to aid in the segmentation and analysis of lung regions. Specifically, three convolutional neural network (CNN) models were developed, each using a distinct preprocessing method: Contrast Limited Adaptive Histogram Equalization (CLAHE), traditional histogram equalization, and no preprocessing. The results revealed that while the CLAHE-enhanced model achieved the highest training accuracy (93.26%) and demonstrated superior stability during training, it showed lower performance in the validation phase, with validation accuracy of 91.31%. In contrast, the model with no preprocessing, which exhibited slightly lower training accuracy (92.98%), outperformed the CLAHE model during validation, achieving the highest validation accuracy of 91.50% and the lowest validation loss. The histogram equalization model demonstrated performance similar to that of CLAHE but with slightly higher validation loss and accuracy compared to the unprocessed model. These findings suggest that while CLAHE excels in enhancing image details during training, it may lead to overfitting and reduced generalization ability. In contrast, the model without preprocessing showed the best generalization and stability, indicating that preprocessing techniques should be chosen carefully to balance feature enhancement with the need for generalization in real-world applications. This study underscores the importance of selecting appropriate image preprocessing techniques to enhance deep learning models' performance in medical image classification, particularly for COVID-19 detection. Histogram Equalization The results contribute to ongoing efforts to optimize diagnostic tools using AI and image processing.

Keywords—X-ray; COVID-19; image enhancement; Contrast Limited Adaptive Histogram Equalization; Histogram Equalization

I. INTRODUCTION

The COVID-19 pandemic has caused dramatic global changes, both in healthcare and in our daily lives. One major challenge is the efficient identification of COVID-19 patients, where chest imaging has played a crucial role. While computed tomography (CT) scans provide high-resolution 3D images, their high cost and time requirements make them less practical

than chest X-rays (CXR) for widespread use. Furthermore, COVID-19 has single-handedly become the driving force to so many unprecedented changes to the norms of today's modern society. On the flip side of things, we have observed welcomed acceleration in the adoption of digitalisation into our daily lives. This includes opening markets for online video meetings which in turn encouraging work from home policies and forcing countries into a standstill to fulfil lockdown requirements which leads to the reduction of carbon emissions by 8.8% (much larger than carbon emission reduction after World War II)[1][7]. Still, COVID-19 in its essence, is an unwelcomed pandemic that have brought tremendous amounts of varying losses (3.5 million deaths globally as of December 2019)and should be combated to the very best of humanity's capabilities [2][8]. Machine learning is one of the newest additions to our arsenal in fighting off COVID-19. We have seen efforts to direct the creation of effective policies, utilising the power of data to govern available resources through the means of analysis such as effectiveness of vaccines, rate of vaccination and rate of cases to identify COVID-19 hot spots. Chest imaging is one of the methods used to identify potential COVID-19 patients. Options include computed tomography (CT), X-ray and ultrasound scans. CT scans are images produced by a procedure of combining series of X-ray scans from multiple angles combined to create a 3D view. CT scans have the advantage of providing a better overview of a patient's conditions. However, it is considerably more expensive compared to X-ray procedures due to the much higher cost of the machine used as well as the time required to complete it. Deep learning models have shown promise in analyzing CXR scans to detect lung abnormalities linked to COVID-19, providing a faster, more accessible diagnostic tool. Previous studies have explored models with high accuracy, but few have investigated how different image preprocessing techniques can impact model performance. Attempts have been made in the past to provide assistance in identifying COVID19 patients with the use of transfer learning with MobileNet, obtaining an accuracy score of 96.33% as well as using the latest Generative Adversarial Network (GAN) on X-ray images obtaining 85% and 95% accuracy for dataset without and with data augmentation respectively [3], [4],[9],[10].

Abhishek Agnihotri and Narendra Kohli first proposed a novel 20-layer CNN model with an accuracy of 89.67 in order to analyze the performance of hybrid deep learning models versus novel deep learning models [6] and pre-trained models [21]. This model performs better than four pre-trained models

(Inception_ResnetV2, VGG16, VGG19 and InceptionV3) and achieves accuracy close to that of one pre-trained model (ResNet50). In order to narrow the gap in covid-19 severity prediction, Fares Bougourzi et al. proposed two methods based on 2D and 3D CNNs respectively. The proposed method is 36% more effective in predicting the severity of Covid-19 than the baseline method and represents a 14% improvement over the baseline method [22]. Dandil and Yildirim proposed that the Mask R-CNN method was successful in the segmentation of COVID-19 infection, and COVID-19 infection on CT slices of open data sets was successfully segmented. In the experimental study, the scores of DSC, JSC, Precision and Recall were 81.93%, 74.19%, 90.27% and 79.47%, respectively [23]. Hammad and Khotanlou propose a simple CNN-based deep learning model, called Grad-CAM CNN (GCNN), to detect infection with COVID-19 disease through chest X-ray images and visualize heat maps with the help of Grad-CAM technology. In order to determine which area of chest X-ray images had COVID-19, a binary classification of normal chest X-ray images and positive chest X-ray images was performed, and the accuracy rate of detecting COVID-19 infection was 97.78%. Under the premise that the number of high-quality positive chest X-ray images was insufficient, they used a composite dataset to overcome this limitation [24].

Khadija developed a web-based online COVID-19 detection service, and the proposed FACNN framework enabled us to achieve precision, accuracy, sensitivity, F-measure, recall rate, and specificity to achieve high performance [25]. Arul Raj. A.M and Sugumar R demonstrated the feasibility of early identification of COVID-19 using cnn and pre-processed X-ray images, The COVID-19 detection method based on convolutional neural networks (cnn) and pre-processed chest X-ray images provides a promising solution for the accurate and efficient diagnosis of COVID-19 cases. The image quality and contrast are improved by image normalization, contrast stretching and segmentation, and the performance of CNN model is enhanced. Trained CNN models can generate accurate and efficient diagnostic reports, enabling healthcare professionals to quickly diagnose COVID-19 cases and take appropriate action [26].

Maddula et al. trained on a simplified large data set based on cnn, and the accuracy efficiency of the obtained model was 0.9835, precision was 0.915, sensitivity was 0.963, specificity was 0.972, and F1 score was 0.987. With ROC AUC of 0.925, this model is better than Random Forest with accuracy of 0.8997 and Naive Bayes with accuracy of 0.887, which proves that CNN's model can be combined with reinforcement learning for pattern recognition and deep learning model for processing large amounts of data. The above methods are helpful to improve the prediction accuracy [27]. Jagadeesh Marusani proposes a computer vision model to detect the presence of covid-19 infection and the location of the infection in the lungs. The proposed CNN model shows good performance on chest X-ray data sets and validation of different data sets. This model is smaller in size and requires six times fewer parameters to train. Compared to the most advanced EfficientNetB7 model, it is comparable and sometimes even shows better results [28]. Renuka Devi SM et al. used deep learning methods to train database images. When given a specific chest X-ray image as

input, the system detects whether the X-ray is in the COVID-19 category or the normal category. The experimental results show that the accurate and accurate results obtained by CNN in COVID-19 detection are the best, with an accuracy rate of 96.8% [29]. Hassam Tahir et al. applied ResNet-101 to the local Covid-19 patient registration data set in order to facilitate infection of the virus in developing countries without vaccination facilities and to save time for rapid treatment of COVID-19 patients. Data from 8009 local chest radiographs were collected. Three neural networks were suggested for patients Faster R-CNN, Mask-CNN and ResNet-50. The faster R-CNN showed the best accuracy at 87 percent. The Mask RCNN was 83% accurate and the resNet-50 was 72% accurate [30]. Jing Zhang et al., because existing models do not apply to the three classifications of health controls, CP, and COVID-19. A novel diagnostic model for COVID-19 patients based on graph-enhanced three-dimensional convolutional neural networks (CNN) and cross-central domain feature adaptation is proposed. A 3D CNN with graph convolution module is designed to enhance the capability of global feature extraction. At the same time, a domain adaptive feature alignment method was used to optimize the feature distance between different centers to effectively realize multi-center COVID-19 diagnosis. Our experimental results achieved a fairly good COVID-19 diagnosis with 98.05% accuracy in the mixed dataset and 85.29% and 87.53% accuracy in cross-center tasks [31].

Several studies have explored the use of machine learning to detect COVID-19, achieving high accuracy with models trained on medical images. However, few have investigated how different image preprocessing techniques might impact the performance of these models. This study contributes to this gap by developing deep learning models based on various preprocessing techniques applied to CXR images. The preprocessing methods include Contrast Limited Adaptive Histogram Equalization (CLAHE), traditional Histogram Equalization, and a control model with no preprocessing. The contributions of this study are listed as follows:

1) *Development of three CNN models:* The study develops CNN models to classify COVID-19 using different image preprocessing techniques (CLAHE, Histogram Equalization, and no preprocessing).

2) *Use of real CXR datasets:* The dataset used consists of real X-ray images from the COVID-19 Radiography Database, obtained from open sources like Kaggle, and includes a large number of CXR images classified into COVID-19, normal, lung opacity, and viral pneumonia categories.

3) *Identification of the most effective preprocessing method:* The study identifies key preprocessing techniques that significantly influence model performance in detecting COVID-19, with Histogram Equalization emerging as the best method for model generalization.

In this paper, we aim to extend existing efforts by evaluating the impact of these preprocessing techniques on deep learning models for COVID-19 detection. Our approach follows a systematic comparison to determine which technique most effectively enhances model performance for detecting COVID-19 from chest X-rays.

II. LITERATURE REVIEW

A. X-ray Scans for COVID-19 Identification

Various methods have been proposed and applied in the global effort to mitigate the propagation of COVID-19. Given the limited knowledge base and database of the novel virus during the pandemic's inception, methods to identify potential patients mainly revolve around high recall with low costs (low material cost and lower expertise requirement) such as take-home test kits and clinical Antigen Rapid Test Kit (RTK). Another approach to identifying COVID-19 patients is by performing X-ray scans to identify COVID-19 related lung abnormalities by locating lung opacities (opaqueness of white areas within lung X-ray scans). The findings in Liqa A. Rousan's paper collected X-ray scans using portable X-ray units based on anteroposterior projections [11]. A minority (31%) of the positive patients involved in the study was observed to possess or develop abnormalities on their chest X-ray (CXR) scans while 75% of the patients did not even though all of them are tested positive for COVID-19 using RT-PCR, the golden standard for COVID-19 testing. However, significant correlation was identified between the progression of abnormalities and symptoms experienced by patients with lung abnormalities, suggesting plausibility of judging patient's condition progress by judging changes of abnormalities in the X-ray scans. Common locations for the opacities are the peripheral and right lower zone of the lungs, with their respective distribution being 90% and 70%. Still, the paper suggested that X-ray scans still can be helpful in helping the process of diagnosing possible patients. Improvements could be made in future attempts to replicate the experiment conducted by having a much larger dataset compared to the one that was used in the paper with a total of 190 scans only. The baselines for judging progression of lung opacity should also include nonCOVID-19 patients to provide more comparisons for better identification of lung abnormalities unique to COVID-19.

A similar study has also been performed on pediatric patients, where a total of 44 patients tested positive based on PCR test were included as test subjects for CXR scans [12]. Results show that only a minority of the children tested (13.6%) has no observable findings in their scans. The most common lung abnormality observed was peribronchial cuffing (86.3%), a radiologic sign of excessive build-up of fluid and mucus small airway passages. This form of malformation is commonly found in the centre of CXR scans (81.8%) followed by 63.3% for peripheral occurrences. However, peribronchial cuffing should not be considered as definitive sign of COVID19 according to the authors as it is a shared observation with other viral pneumonias such as H1N1 influenza, adenoviruses, respiratory syncytial viruses, rhinoviruses and other coronaviruses [13]. Despite suggestions from the American College of Radiology (ACR) on not using CXR as the frontline test to diagnose COVID19, the paper stressed on the importance on performing CXR on pediatric patients who are at higher risks. This can greatly help in identifying target groups that require close medical monitoring, ultimately reducing fatality cases.

Both papers share the same limitation which is lack of data available which limits the possibility of performing a robust

experiment to obtain definitive conclusions on the usability and practicality of CXR scans as a method to mitigate COVID-19.

B. Deep Learning as a Method to Classify X-Ray Scans for Covid-19

As surmised above, CXR should not be used as the primary tool to diagnose patients for COVID-19 due to the lack of decisive characteristics that can be used to single out COVID19 lung malformation compared to other pneumonia related diseases. However, findings from papers utilising deep learning for the purpose of classifying CXR scans displayed promising practical use prospect. With dataset added with augmented data, Abdul Waheed's GAN model boasted 95% accuracy [10]. Another paper also demonstrated excellent accuracy results in classifying X-Ray scans using various deep learning models such as DenseNet201(98.8%), InceptionV3(97.5%) and ResNet101(97.91%) [14]. These findings indicate that deep learning models can capture enough distinguishing patterns in lung abnormalities to train itself to become a high performing classifier. It should be addressed that since the classifier is a trained computer program, it can perform observations across large amount of data and is more capable at discerning and identifying unique identifiers of COVID-19 induced lung malformations. Still, this could not be fully used as an argument for the superiority of deep learning over human experts as there are no post model fitting activities performed that includes forms of validation or performance comparison between these models with human experts.

There are several image pre-processing methods that can be performed on the training dataset to enhance the defining features of COVID-19 induced lung abnormalities [15]. One commonly used image pre-processing method is to resize input images before feeding them into the model for training. This helps in speeding up the training process (by scaling down high-definition images) as well as standardizing input dimension. Image segmentation can also be performed to isolate the lungs from its background, theoretically removing irrelevant noises from being picked up as features by focusing on the Region of Interest (ROI). Another option is to perform image enhancements that enhance defining features of deformations. For this use case, histogram equalization can be used to distribute pixel intensity level. The referred paper suggested the use of Contrast Limited Adaptive Histogram Equalization (CLAHE) as it remedies the downside of using plain Adaptive Histogram Equalization which have the possibility to increase noise intensity in homogenous areas (areas with similar pixel values).

Increasing dataset is one of the popular ways to increase the performance of deep learning models. Considering the relatively young age of the COVID-19 pandemic, there is a scarcity of available datasets. Addition of augmented data can remedy this problem. It should be noted that augmentation is only performed on training datasets to avoid contaminating test datasets that are used for validation with artificial data. Options for augmentation include positional, colour and noise injection. These variations in data adds to the trained model's capability to learn from a more generalised, near to real life data that it will eventually try to classify during its application.

C. Image Pre-processing Methods to Enhance Input Features

Enhancing images as a part of data pre-processing is important in the application of CNN as the features learned are highly reliant on distinguishing features detected from input images. Improving the features of these images via removal of noise or blur and increasing contrast will help in improving spatial features, thus helping CNN models to learn better [10]. Still, the application of image enhancement must be performed in a way that will not affect information contained within the images. Altered features may lead to false learning, which in return will have a negative impact on the final model output. Various methods of image enhancements have been proposed to help improve classification models.

A paper on enhancing images used a fuzzy grayscale enhancement method to address low contrast due to inadequate lighting during capture [17]. The method used succeeded in improving image quality whilst also requiring relatively minimal processing time compared to other techniques. The proposed technique is performed by maximizing fuzzy measures within input images. Power-law transformation and saturation operator is then used to alter the membership function (a curve that defines how each point in the input space is mapped to membership value between 0 and 1) associated with the images.

A four stage image enhancing solution has also been proposed by M.Selvi and Aloysius George namely pre-processing, fuzzy based filtering, adaptive thresholding followed by morphological operation [18]. The stages are created to help pinpoint pixel areas and improve them using Fuzzy based filtering technique and adaptive thresholding. Resulting images are enhanced to have better peak signal-to-noise (PSNR) values compared to other filtering techniques at the time of the paper being written maps, etc., by exploiting similarity and semantic relationships. The nonlinear representation is further exploited in exploring web image search results.

III. METHODOLOGY

A. Data Preparation

The dataset used in this paper is obtained from Kaggle, titled COVID-19 Radiography Database [14], [16]. The dataset can be found using the following link: <https://www.kaggle.com/datasets/tawsifurrahman/covid19radiography-database>. The images are from four health condition classes namely COVID-19, normal (healthy), lung opacity (non-COVID lung infection) and viral pneumonia. Total number of images for each class are 3616, 10192, 6012 and 1345 respectively. Totalling the numbers gives us 21165. This dataset is built by researchers from Qatar University and the University of Dhaka, Bangladesh along with their collaborators from Pakistan and Malaysia. Sources include padchest dataset, a Germany medical school, Github, SIRM, Kaggle and Tweeter. The images are X-Ray scan results from patients subjected to the scan for the purpose of detecting COVID-19. The CNN model will be used to perform feature extraction on these images and use the characteristics identified in input feature maps for the purpose of classification. Based on the abnormalities present in the chest X-ray scans, the CNN model will then be able to perform the necessary predictions. The augmented training data

generation for the CNN model was performed using the Keras ImageDataGenerator class with several parameters passed for the purpose of data augmentation. Data augmentation is a method that increases the amount of data artificially by creating new sets of data derived from geometric transformations applied on the original dataset. Alterations include forms of rotation, translation, flipping and noise addition. Forms of alteration such as adjusting brightness or applying ZCA whitening is not considered. Instead, minor width and height shift is applied to account for possible positions at which the lung and corresponding abnormalities are located within the X-ray scans. Horizontal flip is also enabled, while vertical flip is not. This means, the model will not take into consideration an upside-down X-Ray scans as well as forego any significance put into the positions of pneumonia induced lung malformations. In simpler words, any abnormalities formed either at the right or the left side of the CXR scans are considered to have the same significance in classifying the respective CXR scans.

As highlighted above, the distribution of data for each class is not balanced, with normal X-ray scans consisting almost half of the entire dataset. This was not addressed as initially considered in favour of maintaining the amount of learnable data over possible bias. Dropping these images with the purpose of balancing the dataset may lead to a reduction of performance as the pool of data that the model can learn from reduces. Geometric transformations are not applied to the data due to the sensitive dependence on pixel locations as well as the minute variations that accounts for the difference in the X-ray classes. Other options such as increasing brightness is also not applied as X-ray scans are generally similar in both currently used dataset as well as in real-life practice.

The usual train test split is used instead of stratified split as the preservation of class proportion is not desired due to the imbalance of samples as mentioned. Aside from training purposes, the training set is also sampled and passed to the same pre-processing function used in training to provide visualizations of CXR scans after being processed.

B. Data Pre-processing

Equations in display format are separated from the paragraphs of the text. Equations should be flushed to the left of the column. Equations should be made editable. Displayed equations should be numbered consecutively, using Arabic numbers in parentheses. See Eq. (1) for an example. The number should be aligned to the right margin.

Based on randomly sampled observations, all images are perfectly collected for training, with no visible defects that might significantly impact the model's performance. Thus, no images have been dropped from the original dataset. Normalization is performed on both training and test datasets by enabling the rescale parameter (rescale = 1./255) which converts the pixels within the range from [0, 255] to [0, 1]. This scaling procedure aids in making images contribute more evenly to the calculated total loss. The low range also helps in increasing the likeliness of the neural network to converge.

The input (training and test) images are also resized uniformly to 256 by 256 by specifying the input shape parameter in the first layer of the Keras sequential model.

Three pre-processing methods have been chosen as the differentiating variable for each model, which are CLAHE, histogram equalization [20] and no pre-processing. CLAHE and histogram equalization are image enhancement methods used to bring out distinguishing features that might be important for deep learning models to capture thus building their knowledge base on the classes within the dataset. Both pre-processing methods are implemented using the OpenCV API. CLAHE stands for Contrast Limited Adaptive Histogram Equalization and is a subset of adaptive histogram equalization. It is used primarily to improve contrast in images akin to the usual histogram equalization with a slight difference in approach. Instead of using the entire image, CLAHE computes multiple histograms by focusing on small regions within an image called tiles. These tiles correspond to local areas within distinct sections of the image which redistributes the lightness value of the image. This results in improved contrast, further enhancing the boundaries or edges that will be useful in capturing distinguishing shape of lung abnormalities within the CXR scans. CLAHE is an improved version of adaptive histogram equalization, in which it solves the problem of over amplifying noise in homogenous areas. This is done by the introduction of contrast limit that clips calculated histogram. The clipping process is performed before the calculation of Cumulative Distributions Function (CDF) [19]. This clipping variable is set in the pre-processing of input images phase for the first model by specifying the clip limit (CL) to be 4 using the OpenCV API. The number of tiles were kept the same as the default value specified by OpenCV which is (8,8).

Histogram equalization is an umbrella term that can be used to refer to various subsets of image enhancement using similar methodologies. However, in this experiment histogram equalization used as the manipulated variable for the second model refers to the traditional form of histogram equalization. The method contrasts itself with its aforementioned subset, CLAHE where it only uses one histogram that represents the whole image for the purpose of enhancement. These histograms are then equalized, resulting in a distribution of intensity values across the entire image. Regions with lower contrasts will then be able to increase its contrast as a result. Histogram equalization works particularly well with images with dark background and lighter coloured subjects such as XRay scans. It also has the advantage of being computationally cheaper compared to its more complex subset due to its fairly straightforward approach. The CDF of a standard histogram is given as $H'(i)$:

$$H'(i) = \sum_{0 \leq j < i} H(j) \quad (1)$$

This equation is then used to remap the histogram by normalizing $H'(i)$ to have a maximum value of 255. Next, the intensity values for the histogram equalized image can be obtained using the following equation:

$$\text{equalized}(x,y) = H'(\text{src}(x,y)) \quad (2)$$

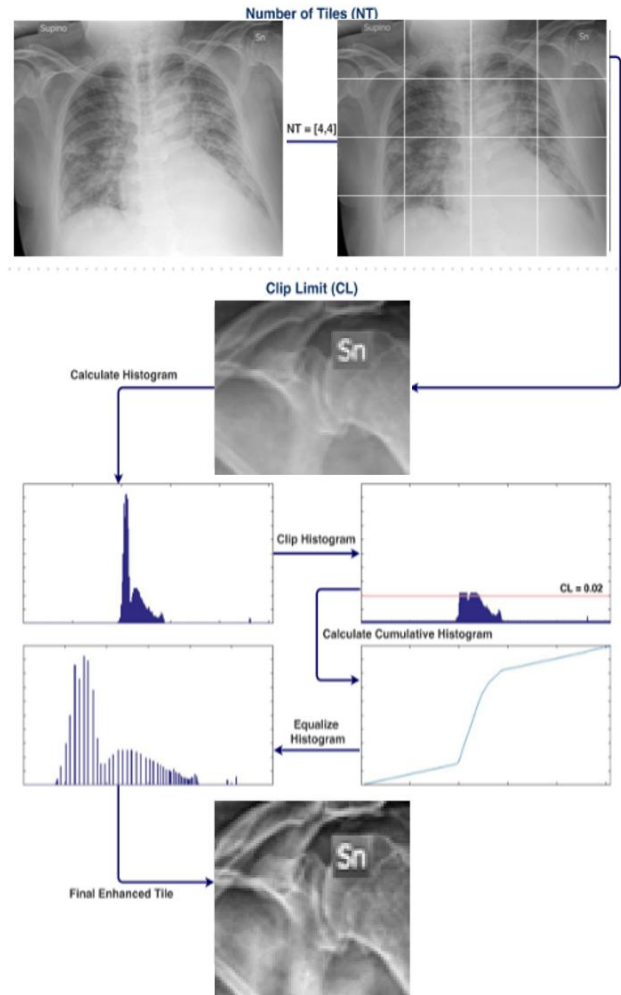


Fig. 1. The process of applying CLAHE.

All models are fed with CXR scans that have been processed into a grayscale image using the OpenCV `cvtColor` method. Fig. 1 demonstrates the process of applying CLAHE (Contrast Limited Adaptive Histogram Equalization) to a chest X-ray image. Below is a detailed explanation of each step in the diagram: The first step involves dividing the chest X-ray image into a 4x4 grid, resulting in 16 smaller tiles.

This division is done to allow local contrast enhancement, which is the main feature of CLAHE. In the diagram, the original chest X-ray image is shown, and the grid overlay highlights the individual tiles. The next step involves setting a clip limit (CL). The clip limit controls the amount of contrast enhancement applied during CLAHE. A higher clip limit leads to greater contrast enhancement. In this flowchart, the clip limit is set to $CL = 0.02$. For each tile, a histogram is calculated. A histogram represents the distribution of pixel intensity values in the image. This step is important because CLAHE works by

manipulating the image's histogram to enhance the contrast in areas of the image with low contrast. After calculating the histogram, the clip histogram step follows. In this step, the histogram is clipped to the defined clip limit. If any part of the histogram exceeds the clip limit, it is clipped off and redistributed, effectively limiting the maximum intensity range. This helps in preventing over-enhancement of certain regions and preserving the details in the image.

This cumulative distribution function (CDF) represents the cumulative sum of the clipped histogram. It helps in redistributing the pixel intensities across the entire range, which further contributes to improving the local contrast. The equalization step follows the cumulative histogram calculation. Here, the pixel intensities are redistributed based on the cumulative histogram. This process enhances the contrast in the image by stretching the pixel intensity values over a wider range. The equalized histogram allows for a more balanced distribution of pixel values, making the image more visually appealing and improving the visibility of important features. The final step shows the enhanced tile after applying CLAHE. The local contrast of the selected tile has been enhanced, making the details of the image more visible. In this case, the tile with the "Sn" label (possibly representing a specific area of interest in the X-ray image) is shown as the final enhanced tile.

C. Descriptive Analysis

Samples of processed images have been by extracting images from the training dataset and passing it through the same image processing methods used in the training of the model. We can see the difference in the resulting images after being passed through different image processing methods as following Fig. 2:

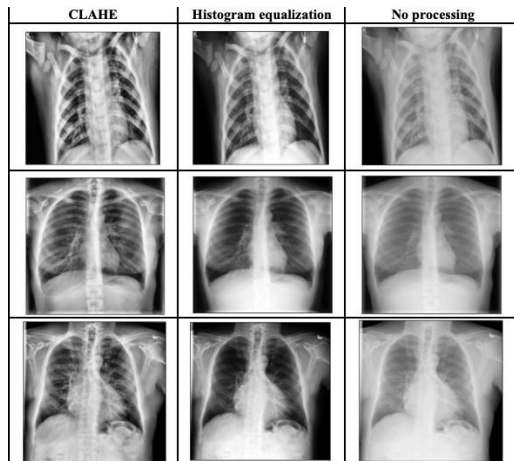


Fig. 2. Images after pre-processed with different image enhancement methods.

Through observation, both CLAHE and histogram equalization has helped in highlighting the edges and the shape of lung opacities when present in CXR scans. Differences in between CLAHE and histogram equalization are as expected, where histogram equalization tends to subdue the intensity of homogenous areas (pixel with similar values). This can be described based on the third row of sample images, where the ribcages at the center of the histogram equalized image is seen to almost lose its shape due to lowered contrast. CLAHE on the other hand seems to uniformly enhance its features, consistently

providing clearer shape of the ribcages. We can also see that CLAHE tends to highlight the structures within white areas better compared to histogram equalized images. This may improve its chance in building a better predictor. It may also backfire as no mention of inner structures of lung opacity is mentioned to be a signal or indicator that can be used to distinguish different pneumonia diseases. This means that these enhanced inner structures might become irrelevant features in the process of identification.

D. Modelling to Data

Three different CNN models have been developed to fulfil the purpose of this paper. All models are constructed using the sequential model class from Keras with the same structure which consists of two convolutional layers, two max pooling layers, one flattening layer and two dense layers.

The convolutional layers are purposely built to have increasing number of filters as the inputs are passed deeper into the CNN model. This is done to capture larger numbers of patterns to enable the model in identifying greater nuances within the CXR scans. Convolutional layers are all proceeded by max pooling layers throughout the structure. Max pooling is immediately applied to the first layer of the CNN structure to downsample input images, reducing dimensions and learnable parameters. This helps in decreasing the amount of time needed to train the classification model. The overall structure can be visualized using the VisualKeras library as follows in Fig. 3:

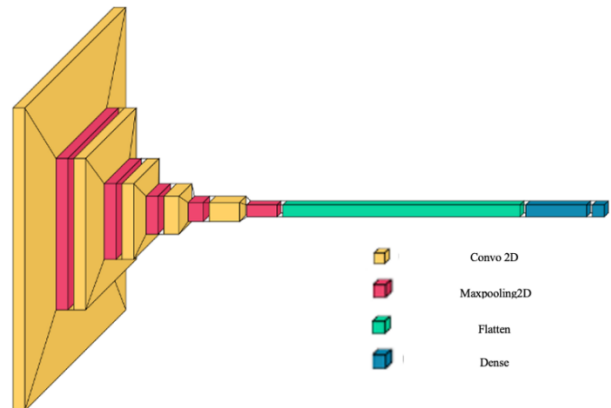


Fig. 3. Resized structure of the CNN models used in the experiment.

Normalization: Normalization is a process in data preprocessing which is used to change the range of numerical data so that it is located in a specific cell, such as [0,1] or [1,1]. In image processing, normalization is a common practice. The essence of the method is some layer input data of the neural network that is preprocessed with zero mathematical expectations and unit variance with the intention of improving the stability and efficiency of the training process [10]. For FER tasks, normalization can give different features similar to ranges. Unnormalized data may lead to unstable gradient problems during model training. In deep learning models, normalized data may lead to a gradient that is too large or too small, thus affecting the learning effect of the model. The normalization in FER2013 dataset is shown in Fig. 8:

Over categorical in this experiment due to the mutually exclusive nature of the dataset classes. This means that the true classes (Y_i) are encoded as standalone integers instead of one-hot encoded. Examples of true classes for sparse categorisation are [1], [2], [3] while one-hot encoded true classes are [1,0,0], [0,1,0], [0,0,1]. The true classes in sparse categorisation refer to the indices of the classes. Linking the class prediction of a model is done by taking the ground truth. For example, if a model output is [0.5, 0.2, 0.4], the prediction will be class 1 if the class indexation starts from 1. The cross-entropy equation is the same, with the only difference being the format of the true class labels:

$$J(w) = -\frac{1}{N} \sum_{i=1}^N y_i \log(\hat{y}_i) \quad (3)$$

Where: y_{-i} = true label, \hat{y}_i = predicted label, W = model parameters.

To avoid overfitting, several options have been considered for data regularization such as adding a dropout layer, adding a normalization layer after the input layer, as well as adding a kernel regularizer in the last layer. Ultimately, it has been decided the only regularizer that will be used is the Ridge Regression regularizer (L2 regularization). The decision to not use normalization layer is due to significant drop in performance for all models when applied. Dropout layer has also been experimented but also disregarded due to similar drop in performance. Dropout layer has also been found to be more effective when used with deeper deep learning structures. The L2 regularization was chosen over L1 regularization as it does not have the tendency to completely remove features deemed as irrelevant. This behaviour is because L1 is capable of forcing coefficients to be exactly zero if given high enough tuning parameter value (usually denoted by λ). While this might be good in reducing the possibility of overfitting in most cases, L1 is not used in this experiment to preserve every single parameter no matter how insignificant and only measure their respective importance by their coefficient values. This decision is made to avoid removing possibly important nuances that might help in the final classification process.

Three options of optimizers have been considered, namely Adaptive Moment Optimization (Adam), stochastic gradient descent (SGD) and Root Mean Squared Propagation (RMSProp). After performing multiple iterations using a controlled model with the same structure, Adam has been chosen due to its excellent performance. It has also been chosen based on its efficiency when working with large datasets. Adam inherits the same concept of momentum from gradient descent with momentum, usually denoted by m_t . The formula is given as follows:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \left[\frac{\delta L}{\delta w_t} \right] \quad (4)$$

Where: m_t = aggregate of gradients at time t [current], m_{t-1} = aggregate of gradients at time $t-1$ [previous], w_t = weights at time t , δL = derivative of Loss Function, δw_t = derivative of weights at time t , β_1 = moving average parameter.

It also inherits the use of exponential moving average from RMSProp, giving us a new variable called sum of square of past gradients, denoted by v_t .

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) \left[\frac{\delta L}{\delta w_t} \right]^2 \quad (5)$$

Where: v_t = sum of square of past gradients, β_2 = moving average parameter.

Adam further improves on these variables by computing and using bias corrected versions of the variables. These new variables are given as follows:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (6)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (7)$$

The weights are then updated using the following equation:

$$w_t = w_t - \hat{m}_t \left(\frac{\alpha}{\sqrt{\hat{v}_t + \epsilon}} \right) \quad (8)$$

Where: ϵ = a small + ve constant to avoid 'division by 0' error.

When training a deep learning model, the training accuracy rate and validation accuracy rate are important indicators to measure the model performance. The training accuracy rate refers to the proportion of correct predictions of the model on the training data set during the training process. It reflects the model's performance on known training data. The validation accuracy rate refers to the prediction accuracy rate of the model on the validation set. Validation sets are data that have not been seen before in the training process and are mainly used to evaluate the generalization ability of the model.

$$\text{Train Acc} = \frac{\sum_{i=1}^n 1(\hat{y}_i = y_i)}{n} \times 100 \quad (9)$$

$$\text{Val Acc} = \frac{\sum_{i=1}^m 1(\hat{y}_i = y_i)}{m} \quad (10)$$

Where, for each training sample x_i the model's prediction \hat{y}_i , y_i is compared with the true label. If the prediction is correct, that counts as a correct prediction.

Due to the sheer variations of model that can be generated by varying the value of hyperparameters in CNN, cross-validation and hyperparameter tuning using GridSearchCV or RandomSearchCV has not been applied. This decision was made considering the computational costs involved as well as time constraints. However, hyperparameter tuning has been done manually to increase the performance of each model. The final values of hyperparameters are listed as below in Table I:

TABLE I. FINAL HYPERPARAMETERS CHOSEN FOR CNN MODELS

Hyperparameter	Value
Batch size	32
Epoch	25
Adam learning rate	0.001
Adam β_1	0.9
Adam β_2	0.999
Adam ϵ	0.0000007
L2 regularizer λ	0.01
CLAHE clip limit	4

E. Communicating and Visualizing the Results

As shown in Fig. 4, both Train Loss and Val Loss decrease gradually with the training rounds. The validation loss fluctuates at some points, but the overall trend is also downward. The CLAHE-enhanced model showed a good decreasing trend of training and verification loss, indicating that the model gradually learned the features on the training set and verification set. The training accuracy (Train Acc) and validation accuracy (Val Acc) both increased steadily, and basically became stable when they approached 20 epochs, indicating that the verification accuracy and training accuracy were close to each other, indicating that the model avoided overfitting well, and CLAHE enhancement helped the model learn image features better [5].

As shown in Fig. 5, training losses and validation losses decreased gradually, and validation losses also fluctuated in some epochs, but the overall trend was downward. Compared with the CLAHE-enhanced model, the validation loss fluctuation is slightly larger, indicating that the model has a slightly weak generalization ability on the validation set and may need further tuning. Training accuracy and validation accuracy rise rapidly in the initial phase and level off near 20 epochs. The verification accuracy is slightly lower than the training accuracy, which indicates that the performance of the model on the verification set is slightly worse than that on the training set, and there is a certain tendency of overfitting, but the overall performance is still good.

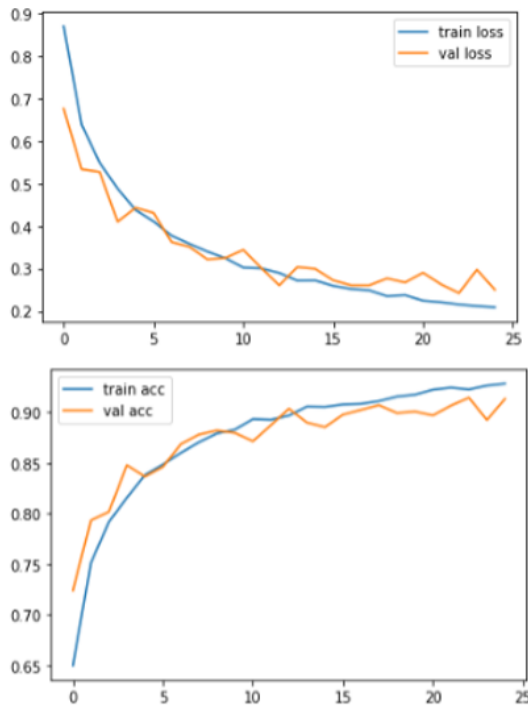


Fig. 4. Images after pre-processed with different image enhancement methods (I).

As shown in Fig. 6, Training losses and validation losses also decrease with epoch, and validation losses also fluctuate at some points, but less so. The loss of the model without preprocessing decreased relatively gradually, especially after 10 epochs, and the validation loss was sometimes slightly higher than the training loss, indicating that the model needed longer training

time to reach a stable state without preprocessing. Despite the high accuracy on the validation set, it performed slightly worse than models preprocessed with CLAHE and histogram equalization, suggesting that data without preprocessing may result in a limited ability of the model to learn features as well as it would have done with preprocessing techniques.

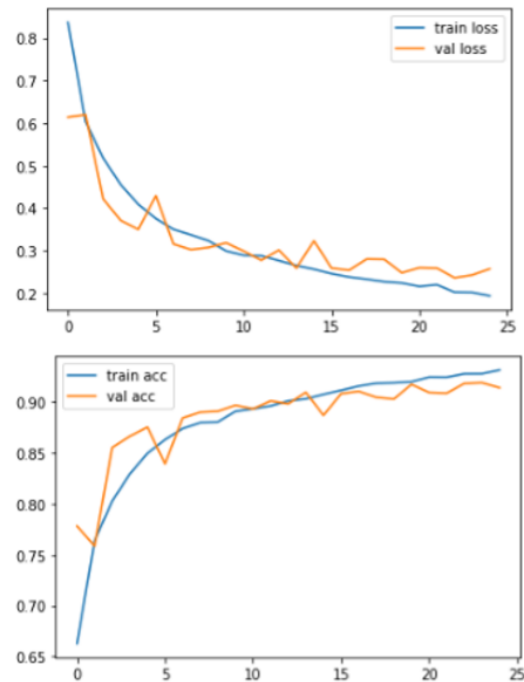


Fig. 5. Images after pre-processed with different image enhancement methods (II).

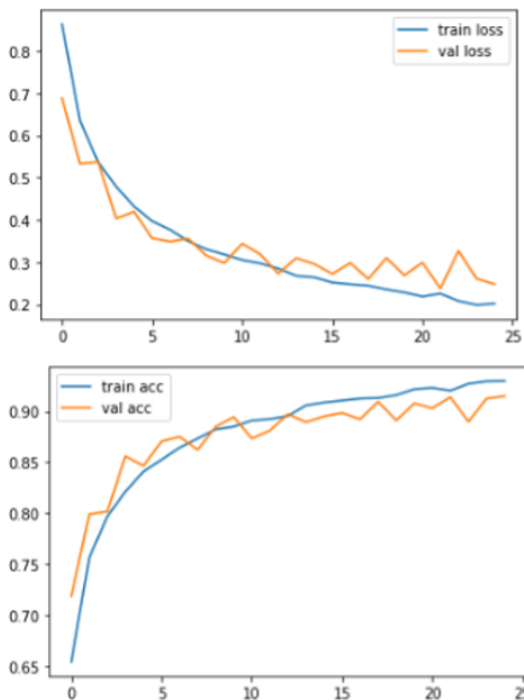


Fig. 6. Images after pre-processed with different image enhancement methods (III).

Image preprocessing has a significant effect on the performance of the model. CLAHE enhancement technology works best at improving local contrast, helping models better learn useful features to improve accuracy and reduce overfitting.

Histogram equalization came in second, while models without any preprocessing performed poorly. It is suggested to use appropriate image preprocessing technology in practical application to improve the generalization ability and overall performance of the model.

IV. RESULTS

The Table II outlines the performance of each model based on the predefined evaluation metrics above.

TABLE II. PERFORMANCE OF EACH MODEL

Image enhancement	Accuracy (%)	Validation accuracy (%)	Loss	Validation loss
CLAHE	93.26	91.31	0.1987	0.2503
Histogram equalization	93.16	91.42	0.1935	0.2569
No preprocessing	92.98	91.50	0.2020	0.2479

Based on the final results obtained, we can observe that all four evaluation metric scores for all of the models are relatively the same.

As shown in Table III. CLAHE performs the best at training, with the highest accuracy score and second lowest loss. The difference between models using histogram equalized inputs and no pre-processing is relatively minor, where no preprocessing scores 0.18 lower accuracy and 0.0085 higher loss. CLAHE scored 0.28 and 0.1 higher in training accuracy compared to no pre-processing and histogram equalized models respectively. CLAHE also has the second-best loss with a 0.0052 and 0.0032 margin compared to no histogram equalized and no pre-processing models. However, evaluation metrics based on the validation data set tells a different story, with CLAHE being the worst performer, with validation accuracy of 91.31 and a 0.0024 higher loss compared to no pre-processing. Interestingly, the model trained using inputs without any form of image processing became the best performer with a validation accuracy score of 91.50 and validation loss of 0.2479. This contradicts with initial hypothesis based on literature reviews, that CLAHE would be the best performer in both training and validation phase. Upon consideration, histogram equalization is chosen as the best method to enhance CXR scans for this very specific CNN model. The main reason to this choice is due its overall performance during both training and validation phase. Models that are better at validation usually signifies a better capability to generalise. Specifically for the context of this experiment, histogram equalization enables the model to better identify distinguishing features that characterise COVID-19 inflicted CXR scans instead of ‘memorizing’ the features through training datasets without actual ‘understanding’. The fact that CLAHE is better at enhancing the minute details of lung abnormalities might the drawback to its ability to generalize as well as histogram equalized model and no preprocessing model.

TABLE III. EVALUATION METRIC MEAN AND STANDARD DEVIATION CLAHE

Metric	Mean	Standard deviation
Training accuracy	0.869348	0.079467384
Training loss	0.351748	0.189184685
Validation accuracy	0.872868	0.044526815
Validation loss	0.34308	0.107216724

Table IV shows the histogram equalization model. the mean value of the training accuracy rate is 0.880868, and the standard deviation is 0.061541603, showing a high training accuracy rate, while the standard deviation is small, indicating that the training process is relatively stable. The average verification accuracy is 0.885632, and the standard deviation is 0.040508844. The verification accuracy is high and the fluctuation is small, indicating that the model has good generalization ability. The validation loss is 0.32322 with a standard deviation of 0.102009803, showing large fluctuations on some validation data, but still performing well overall. Histogram equalization improves the performance of the model, especially in the validation accuracy, but the validation loss fluctuates greatly, suggesting that its stability is slightly lower than that of CLAHE.

TABLE IV. EVALUATION METRIC MEAN AND STANDARD DEVIATION HISTOGRAM EQUALIZATION

Metric	Mean	Standard deviation
Training accuracy	0.880868	0.061541603
Training loss	0.323184	0.148779061
Validation accuracy	0.885632	0.040508844
Validation loss	0.32322	0.102009803

Table V shows that the mean training accuracy of the model without preprocessing is 0.876968, and the standard deviation is 0.063891847, indicating that the consistency in the training process is good, but the accuracy of the model is slightly lower than that of other preprocessing methods. The average validation accuracy is 0.875144, and the standard deviation is 0.044533499, which is close to the validation accuracy compared with other methods, but slightly lower than CLAHE and histogram equalization methods. The verification loss is 0.343188 and the standard deviation is 0.104845224. The fluctuation of the verification loss is large, which shows the instability of the model on the verification set.

TABLE V. EVALUATION METRIC MEAN AND STANDARD DEVIATION WITH NO-PREPROCESSING

Metric	Mean	Standard deviation
Training accuracy	0.876968	0.063891847
Training loss	0.335916	0.155654603
Validation accuracy	0.875144	0.044533499
Validation loss	0.343188	0.104845224

V. DISCUSSION

During the training phase, the CLAHE-enhanced model outperformed both the histogram equalization and no preprocessing models in terms of accuracy, achieving a training accuracy of 93.26%. This suggests that CLAHE excels in enhancing the minute details of lung abnormalities, which could be crucial in detecting subtle features associated with COVID-19 infections. The model's relatively low loss (0.1987) further emphasizes its ability to minimize errors during training.

However, the histogram equalization method, while slightly less effective than CLAHE in terms of training accuracy, performed well with a mean training accuracy of 93.16%. This method helped improve the contrast and brightness of CXR images, which may have helped the model more effectively learn the distinguishing features of the images. Notably, the training loss for histogram equalization (0.1935) was also lower than that of the CLAHE-enhanced model, suggesting that while CLAHE improves feature details, histogram equalization might be more effective at optimizing the overall model performance by reducing error rates during training.

The model without preprocessing, while achieving a slightly lower training accuracy (92.98%), showed stable training consistency. With a training loss of 0.2020, it demonstrated that even without preprocessing, the CNN model could still effectively learn to classify the CXR images, albeit with less precision than the other methods. This highlights that while preprocessing enhances model performance, it is not an absolute requirement for effective training.

The validation phase results presented a different picture, where the CLAHE-enhanced model performed the worst in terms of validation accuracy (91.31%) and exhibited the highest validation loss (0.2503). This is in contrast to the initial hypothesis, which anticipated that CLAHE would perform well in both training and validation phases. The discrepancy between training and validation performance suggests that while CLAHE is effective in fine-tuning the model's ability to capture minute details in CXR images, it might lead to overfitting. The enhanced features could cause the model to 'memorize' training data without fully generalizing to unseen validation images, thus impairing its performance on the validation set.

In contrast, the model trained without any preprocessing achieved the best validation accuracy (91.50%) and the lowest validation loss (0.2479), despite its lower training accuracy compared to CLAHE. This indicates that the lack of preprocessing enabled the model to generalize better, as it did not overfit the specific features of the training set. The validation results for this model demonstrate that preprocessing methods like CLAHE and histogram equalization might enhance feature extraction but at the cost of generalization ability. These findings highlight the importance of balancing training performance with generalization, especially in medical image classification, where the model must perform well on unseen data. The histogram equalization model, while showing good validation accuracy (91.42%), also exhibited noticeable fluctuations in validation loss (0.2569). While histogram equalization improved the model's ability to generalize better than CLAHE, it still presented challenges in terms of stability during the validation phase. The slightly better performance of the histogram

equalization model, compared to CLAHE, underscores its ability to enhance image contrast while maintaining reasonable generalization.

VI. CONCLUSION

This study evaluated the impact of different image preprocessing techniques—CLAHE, traditional histogram equalization, and no preprocessing—on the performance of a convolutional neural network (CNN) for COVID-19 classification using chest X-ray (CXR) images. The experimental results demonstrated that all preprocessing methods improved model performance during the training phase, but the validation phase revealed distinct trade-offs between accuracy, loss, and generalization ability.

The CLAHE-enhanced model achieved the highest training accuracy (93.26%) and exhibited strong stability, but it showed poor generalization in the validation phase, with the lowest validation accuracy (91.31%) and higher validation loss (0.2503). This suggests that while CLAHE helps capture detailed image features, it may lead to overfitting, affecting the model's ability to generalize effectively. In contrast, the model without preprocessing achieved the best validation performance, with a validation accuracy of 91.50% and the lowest validation loss (0.2479), highlighting its superior generalization ability. However, its training accuracy (92.98%) was slightly lower compared to the other methods. This finding emphasizes that while preprocessing enhances feature extraction, a simpler, unprocessed approach can sometimes yield better generalization.

The histogram equalization method, while not the best in training accuracy, provided a good balance between training performance and validation accuracy. With a validation accuracy of 91.42%, it demonstrated that traditional image enhancement techniques could improve generalization without overfitting, making it the most suitable preprocessing method for the CNN model in this study.

In conclusion, histogram equalization emerged as the optimal preprocessing method for COVID-19 classification in CXR images, offering the best combination of training and validation performance. Future work could investigate more sophisticated preprocessing techniques or hybrid models to further enhance both performance and generalization in medical image classification tasks.

VII. FUTURE WORKS AND LIMITATION

While this study has provided valuable insights into the effect of image preprocessing techniques on COVID-19 detection using chest X-ray (CXR) images, there are several avenues for future research to further enhance the performance and generalization of deep learning models. 1. Future work could explore combining CLAHE and histogram equalization to leverage the strengths of both methods. A hybrid preprocessing approach could potentially enhance image details while maintaining good generalization ability, addressing the limitations seen when using CLAHE alone. 2. The use of more sophisticated deep learning models, such as ResNet, DenseNet, or Inception networks, may further improve performance, especially in terms of handling complex features in CXR images. These architectures have been shown to excel at feature

extraction and overcoming challenges like overfitting. 3. To address the potential overfitting issues observed, further research could incorporate advanced data augmentation techniques. This could include random rotations, flips, and color jittering, or even synthetic data generation techniques, to create a more diverse training dataset and enhance the generalization capability of the model. 4. Future research should focus on testing these models in real-world clinical environments to evaluate their robustness, scalability, and performance on larger, diverse datasets. This would also include the development of a user-friendly interface for healthcare professionals to easily adopt the models in practice.

This study has several limitations that should be addressed in future work. 1. The model was trained using a limited number of CXR images from the COVID-19 Radiography Database. Although the dataset is large, it may not fully represent the variety of CXR images encountered in real-world clinical settings, which could impact the model's ability to generalize to diverse populations and varying image qualities. Expanding the dataset or incorporating additional datasets from other regions or healthcare providers could improve the model's robustness. 2. While different preprocessing techniques were evaluated, the impact of each preprocessing method may vary depending on the dataset used. The methods tested in this study may not perform equally well on other datasets or in clinical settings. Therefore, the generalizability of these findings across different datasets remains an open question. 3. While deep learning models, including CNNs, are powerful for image classification tasks, they are often criticized for their lack of interpretability. Future work should focus on making the models more explainable to healthcare providers. Techniques such as Grad-CAM (Gradient-weighted Class Activation Mapping) can be employed to visualize which parts of the CXR images are contributing to the model's predictions, making the model more transparent and aiding in clinical decision-making. 4. This study focused solely on the detection of COVID-19 from CXR images, and did not account for other variables that may affect the model's performance, such as different scanner types, patient positioning, or image resolution. These external factors can significantly influence model accuracy and should be considered in future studies for a more comprehensive evaluation of the model's real-world effectiveness.

ACKNOWLEDGMENT

Funding: This research was funded by the Universiti Kebangsaan Malaysia (Grant code: GUP-2022-060).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Ahmad Nuruddin bin Azhar and Nor Samsiah sani; data collection: Ahmad Nuruddin bin Azhar; analysis and interpretation of results: Ahmad Nuruddin bin Azhar, Liu Luan Xiang Wei and Nor Samsiah sani; draft manuscript preparation: Mohd Aliff Afira Sani, Liu Luan Xiang Wei and Nor Samsiah sani. All authors reviewed the results and approved the final version of the manuscript.

Conflict of Interest The corresponding author states that there is no conflict of interest on behalf of all authors.

Data Availability: The data used in this study are available from the following resources in the public domain:

<https://www.kaggle.com/datasets/tawsifurrahman/covid19radiography-database>.

REFERENCES

- [1] Dobrojevic, M., Zivkovic, M., Chhabra, A., Sani, N. S., Bacanin, N., & Amin, M. M. (2023). Addressing internet of things security by enhanced sine cosine metaheuristics tuned hybrid machine learning model and results interpretation based on shap approach. *PeerJ Computer Science*, 9, e1405.
- [2] Suwadi, N. A., Derbali, M., Sani, N. S., Lam, M. C., Arshad, H., Khan, I., & Kim, K. I. (2022). An optimized approach for predicting water quality features based on machine learning. *Wireless Communications and Mobile Computing*, 2022(1), 3397972.
- [3] Othman, Z. A., Bakar, A. A., Sani, N. S., & Sallim, J. (2020). Household overspending model amongst B40, M40 and T20 using classification algorithm. *International Journal of Advanced Computer Science and Applications*, 11(7).
- [4] Mohamed Nafuri, A. F., Sani, N. S., Zainudin, N. F. A., Rahman, A. H. A., & Aliff, M. (2022). Clustering analysis for classifying student academic performance in higher education. *Applied Sciences*, 12(19), 9467.
- [5] Holliday, J., Sani, N., & Willett, P. (2018). Ligand-based virtual screening using a genetic algorithm with data fusion. *Match: Communications in Mathematical and in Computer Chemistry*, 80(3).
- [6] Bassel, A., Abdulkareem, A. B., Alyasseri, Z. A. A., Sani, N. S., & Mohammed, H. J. (2022). Automatic malignant and benign skin cancer classification using a hybrid deep learning approach. *Diagnostics*, 12(10), 2472.
- [7] Z. Liu et al., "Near-real-time monitoring of global CO2 emissions reveals the effects of the COVID-19 pandemic." *Nat. Commun.*, vol. 11, no. 1, 2020, doi: 10.1038/s41467-020-18922-7.
- [8] V. J. Jayaraj, S. Rampal, C. W. Ng, and D. W. Q. Chong, "The Epidemiology of COVID-19 in Malaysia," *Lancet Reg. Heal. - West. Pacific*, vol. 17, 2021, doi: 10.1016/j.lanwpc.2021.100295.
- [9] S. Dilshad et al., "Automated image classification of chest X-rays of COVID-19 using deep transfer learning," *Results Phys.*, vol. 28, 2021, doi: 10.1016/j.rinp.2021.104529.
- [10] A.Waheed, M. Goyal, D. Gupta, A. Khanna, F. Al-Turjman, and P. R. Pinheiro, "CovidGAN: Data Augmentation Using Auxiliary Classifier GAN for Improved Covid-19 Detection," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.2994762.
- [11] L. A. Rousan, E. Elobeid, M. Karrar, and Y. Khader, "Chest x-ray findings and temporal lung changes in patients with COVID-19 pneumonia," *BMC Pulm. Med.*, vol. 20, no. 1, 2020, doi: 10.1186/s12890020-01286-5.
- [12] C. Oterino Serrano et al., "Pediatric chest x-ray in covid-19 infection," *Eur. J. Radiol.*, vol. 131, 2020, doi: 10.1016/j.ejrad.2020.109236.
- [13] A.I.K., A. R., U. Patel, and S. K. Joshi, "H1N1 influenza: Characterization of initial chest radiographic findings and prognostic value of serial chest radiographs," *Radiol. Infect. Dis.*, vol. 3, no. 4, 2016, doi: 10.1016/j.rjid.2016.11.005.
- [14] M. E. H. Chowdhury et al., "Can AI Help in Screening Viral and COVID-19 Pneumonia?," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3010287.
- [15] H. Mary Shyni and E. Chitra, "A COMPARATIVE STUDY OF X-RAY AND CT IMAGES IN COVID-19 DETECTION USING IMAGE PROCESSING AND DEEP LEARNING TECHNIQUES," *Comput. Methods Programs Biomed. Updat.*, vol. 2, 2022, doi:10.1016/j.cmpbup.2022.100054.
- [16] T. Rahman et al., "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Comput. Biol. Med.*, vol. 132, 2021, doi: 10.1016/j.cmbiomed.2021.104319.
- [17] K. Hasikin and N. A. M. Isa, "Enhancement of the low contrast image using fuzzy set theory," 2012. doi: 10.1109/UKSim.2012.60.

- [18] M. Selvi and A. George, "FBFET: Fuzzy based fingerprint enhancement technique based on adaptive thresholding," 2013. doi: 10.1109/ICCCNT.2013.6726776.
- [19] W.-N. Mohd-Isa, J. Joseph, N. Hashim, and N. Salih, "Enhancement of digitized X-ray films using Contrast-Limited Adaptive Histogram Equalization (CLAHE)," *F1000Research*, vol. 10, 2021, doi: 10.12688/f1000research.73236.1.
- [20] G. F. C. Campos, S. M. Mastelini, G. J. Aguiar, R. G. Mantovani, L. F. de Melo, and S. Barbon, "Machine learning hyperparameter selection for Contrast Limited Adaptive Histogram Equalization," *Eurasip J. Image Video Process.*, vol. 2019, no. 1, Dec. 2019, doi: 10.1186/s13640019-0445-4.
- [21] A. Agnihotri and N. Kohli, "A Hybrid Deep Neural approach for multi-class Classification of novel Corona Virus (COVID-19) using X-ray images," in *2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, Gharuan, India: IEEE, May 2023, pp. 1–5. doi: 10.1109/InCACCT57535.2023.10141782.
- [22] F. Bougourzi, F. Dornaika, A. Nakib, C. Distant, and A. Taleb-Ahmed, "Deep-Covid-SEV: an Ensemble 2D and 3D CNN-Based Approach for Covid-19 Severity Prediction from 3D CT-SCANS," in *2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, Rhodes Island, Greece: IEEE, Jun. 2023, pp. 1–5. doi: 10.1109/ICASSPW59220.2023.10192927.
- [23] E. Dandil and M. S. Yildirim, "Automatic Segmentation of COVID-19 Infection on Lung CT Scans using Mask R-CNN," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, Ankara, Turkey: IEEE, Jun. 2022, pp. 1–5. doi: 10.1109/HORA55278.2022.9800029.
- [24] H. Hammad and H. Khotanlou, "Detection and visualization of COVID-19 in chest X-ray images using CNN and Grad-CAM (GCCN)," in *2022 9th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*, Bam, Iran, Islamic Republic of: IEEE, Mar. 2022, pp. 1–5. doi: 10.1109/CFIS54774.2022.9756420.
- [25] B. Khadija, "Automatic detection of covid-19 using CNN model combined with Firefly algorithm," in *2022 8th International Conference on Optimization and Applications (ICOA)*, Genoa, Italy: IEEE, Oct. 2022, pp. 1–4. doi: 10.1109/ICOA55659.2022.9934144.
- [26] A. R. A. M and S. R., "Enhancing COVID-19 Diagnosis with Automated Reporting using Preprocessed Chest X-Ray Image Analysis based on CNN," in *2023 2nd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India: IEEE, May 2023, pp. 35–40. doi: 10.1109/ICAAIC56838.2023.10141515.
- [27] P. Maddula, P. Srikanth, P. K. Sree, P. B. V. R. Rao, and P. T. S. Murty, "COVID-19 prediction with Chest X-Ray images using CNN," in *2023 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE)*, Bengaluru, India: IEEE, Jan. 2023, pp. 568–572. doi: 10.1109/IITCEE57236.2023.10090951.
- [28] J. Marusani, B. G. Sudha, and N. Darapaneni, "Small-Scale CNN-N model for Covid-19 Anomaly Detection and Localization From Chest X-Rays," in *2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR)*, Hyderabad, India: IEEE, Mar. 2022, pp. 1–6. doi: 10.1109/ICAITPR51569.2022.9844184.
- [29] R. D. S M, B. S. Rose, S. Akshitha, and P. Niharika, "Comparison of COVID-19 Diagnosis by CNN Model and ResNet Using Chest X-Ray," in *2023 International Conference on Sustainable Communication Networks and Application (ICSCNA)*, Theni, India: IEEE, Nov. 2023, pp. 1569–1574. doi: 10.1109/ICSCNA58489.2023.10370248.
- [30] H. Tahir, A. Ifikhar, and M. Mumraiz, "Forecasting COVID-19 via Registration Slips of Patients using ResNet-101 and Performance Analysis and Comparison of Prediction for COVID-19 using Faster R-CNN, Mask R-CNN, and ResNet-50," in *2021 International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, Bhilai, India: IEEE, Feb. 2021, pp. 1–6. doi: 10.1109/ICAECT49130.2021.9392487.
- [31] J. Zhang *et al.*, "Graph Convolution and Self-attention Enhanced CNN with Domain Adaptation for Multi-site COVID-19 Diagnosis," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, Sydney, Australia: IEEE, Jul. 2023, pp. 1–4. doi: 10.1109/EMBC40787.2023.10340851.

Deep Learning-Based Automatic Cultural Translation Method for English Tourism

Jianguo Liu^{1*}, Ruohan Liu²

Foreign Languages College, Henan University of Science and Technology, Luoyang 471000, Henan, China¹

Marxist Academy, Henan University of Science and Technology, Luoyang 471000, Henan, China¹

School of Electronic Engineering, Xi'an Post and Communications University, Xi'an 710100, Shaanxi, China²

Abstract—The general LSTM-based encoder-decoder model has the problems of not being able to mine the sentence semantics and translate long text sequences. This study presents a neural machine translation model utilizing LSTM with improved attention, incorporating multi-head attention and multi-skipping attention mechanisms into the LSTM baseline model. By adding multi-head attention computation, the syntactic information in different subspaces can be mined, and then attention can be paid to the semantic information in the sentence sequences, and then multiple attentions are computed on each head separately, which can effectively deal with the long-distance dependency problem and perform better in the translation of long sentences. The proposed model is analysed and compared using the WMT17 Chinese and English datasets, newsdev2017 and newstest2017, and the results show that the proposed model improves the BLEU score of the automatic translation of Tourism English Culture and solves the problem of low scores in long sentence translation.

Keywords—LSTM-based encoder-decoder model; tourism English culture; automatic translation; enhanced attention mechanism

I. INTRODUCTION

The process of economic globalisation has further brought about the globalisation of language, and English has become the only protagonist of linguistic globalisation, and English language learning has gradually become a matter of concern [1]. English, as a part of the cultural composition of tourism, has become a necessary skill for people travelling abroad across borders. In order to understand tourism English more conveniently, natural language processing technology based on artificial intelligence algorithms has entered people's life [2]. Natural Language Processing (NLP) is to transform the language humans usually communicate and the text they see into what machines can understand [3]. Natural language processing technology has a wide range of applications, including machine translation [4], sentiment classification [5], robot dialogue [6], text classification [7], etc. With the popularity of deep learning, deep neural networks began to be introduced into NLP tasks and made great progress, while machine translation based on deep neural networks received great attention, and researchers embedded deep neural networks into machine translation tasks, which led to the improvement of the quality of automatic English translation [8]. Deep neural network-based machine translation can effectively promote the future economic and social development, thus enhancing people's satisfaction and sense of access. Therefore, the study of machine translation of

tourism English based on deep neural networks is a meaningful research direction, which is of great significance for globalised economic exchange and cultural output [9].

The primary objective of the application of deep learning technology in the automatic translation of tourism English culture is to acquire a comprehensive understanding of the structure and laws of language through neural network models, thereby enhancing the quality and efficiency of translation [10]. Currently, the automatic translation methods of tourism English culture based on deep learning include Seq2Seq model [11], Attention mechanism [12], Transformer model [13], Pre-training language model [14], Multi-modal data translation [15], Zero Resource Translation [16], Online learning and incremental learning of Neural Machine Translation [17] [18] and so on. Although Neural Machine Translation has made greater progress and is better than Statistical Machine Translation on some public datasets, Neural Machine Translation is still far from the effect of human translation, and there are still the following challenges and problems [10]: 1) data sparsity problem; 2) model optimisation problem; 3) large-scale vocabularies and rare words problem.

This text proposes a method for autonomous cultural translation of Tourist English, addressing the issues of attention computation and model optimization in encoder-decoder architectures utilizing recurrent neural networks, specifically through the implementation of an LSTM-enhanced attention mechanism. This paper's primary contributions are: 1) the introduction of an automatic language translation model and a neural machine translation framework; 2) the investigation and design of a neural machine translation model utilizing an LSTM-enhanced attention mechanism; and 3) a comparative analysis of the proposed model employing the Tourism English dataset.

The structure of this paper is organized as follows: Section I discusses foundational techniques, covering automatic translation systems and neural machine translation frameworks. Section II outlines the key challenges in translating tourism English, such as data sparsity and issues with long-sequence translations. Section III introduces the enhanced attention mechanism based on LSTM, emphasizing multi-head and multi-hop attention for improved performance. Section IV describes the experimental framework, including datasets, evaluation methods, and model comparisons using BLEU scores. Section V presents the findings and analysis, demonstrating the model's advantages. Last Section VI concludes with insights and suggestions for future work.

*Corresponding Author

II. RELEVANT THEORETICAL TECHNIQUES

A. Automatic Language Translation Models

The Automatic Language Translation Model [19] is a tool that uses artificial intelligence techniques to automatically convert text from one language to another by means of a computer programme, as shown in Fig. 1.

Automatic language translation models are usually based on deep learning techniques, especially neural networks, such as Recurrent Neural Networks (RNN), Long Short-Term Memory

Networks (LSTM) and Transformer models, as shown in Fig. 2. They are capable of handling complex linguistic structures and expressions and have achieved good translation quality in several public reviews.

To mitigate the issues associated with one-hot encoding, including context independence and excessive dimensionality, the Neural Probabilistic Language Model (NPLM) [20] derives word vectors by learning word distributions, as illustrated in Fig. 3.

UnitY model architecture

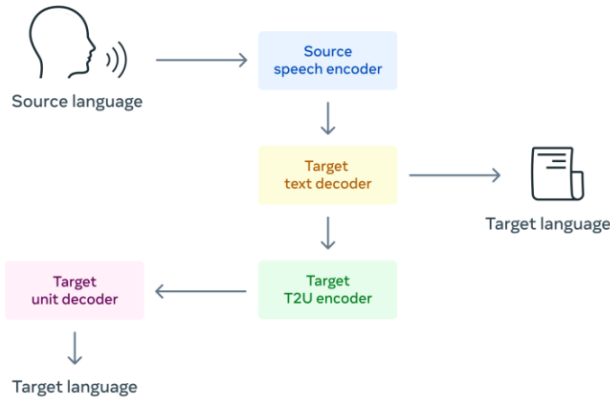


Fig. 1. Model of automatic language translation.

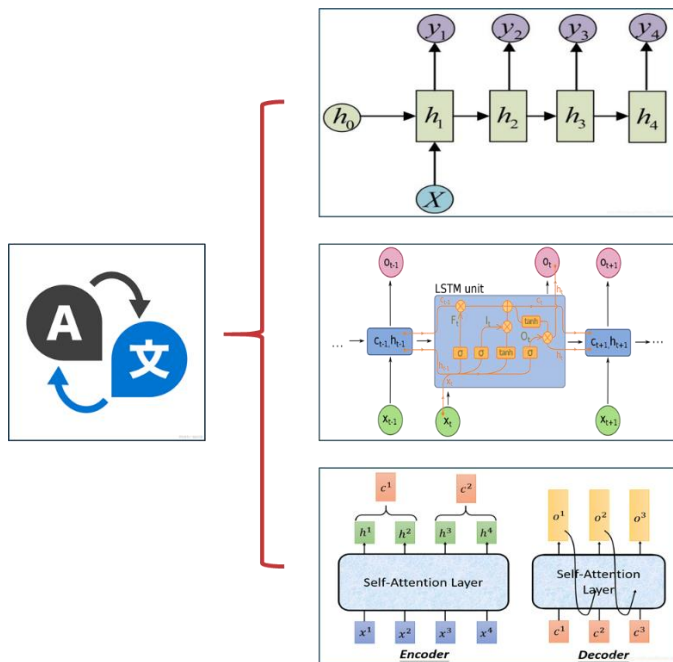


Fig. 2. Classification of automatic language translation models.

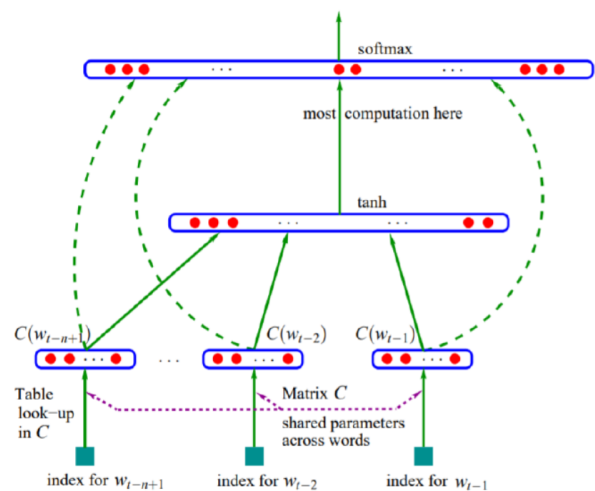


Fig. 3. NPLM structure.

In the NPLM structure, given a text sequence $(w_1, w_2, \dots, w_t, \dots, w_T)$, where w_t is the word in the word list V . The objective function is to build the optimal model f , calculated as follows:

$$f(w_t, \dots, w_{t-n+1}) = \hat{P}(w_t | w_1, \dots, w_{t-1}) \quad (1)$$

$$\sum_{i=1}^{|V|} f(i; w_{t-1}, \dots, w_{t-n+1}) = 1 \quad (2)$$

The structure is divided into two parts, the first part is to build a mapping from any word w_i to a vector in the word list V $C(i)$, the second part has a feed forward neural network g to fit $f(i; w_{t-1}, \dots, w_{t-n+1})$ where $f(i; w_{t-1}, \dots, w_{t-n+1}) = g(i, C(w_{t-1}), \dots, C(w_{t-n+1}))$, the training objective is to maximise the following equation:

$$L = \frac{1}{T} \sum_t \log f(w_t, w_{t-1}, w_{t-N+1}) + R(\theta) \quad (3)$$

where R is the regular term and θ is the parameter of the feedforward neural network g .

B. Neural Machine Translation Framework and Classification

1) *Text feature representation*: Based on the representation of word vectors, the textual feature representation is subsequently obtained, i.e. the Embedding operation [21], the specific structure of which is shown in Fig. 4. The Embedding layer is often used in the first layer of the neural machine translation model, and its role is to map the input sequences into dense vectors of lower dimensions, which are able to characterise the word information effectively.

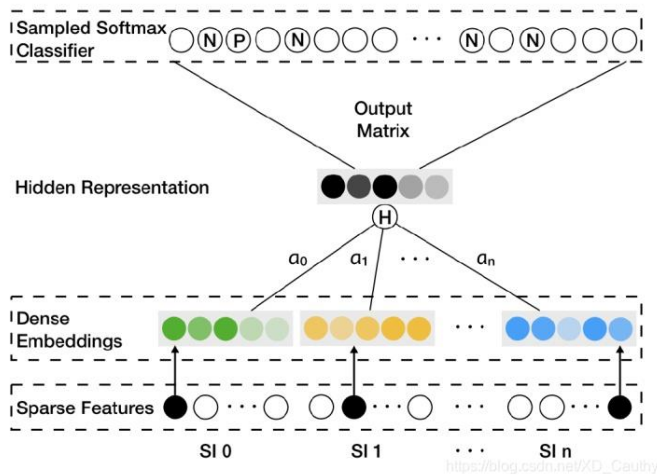


Fig. 4. Embedding operation.

2) *Encoder-decoder structure*: Many neural machine translation models are constructed using the Encoder-Decoder framework [22], which is also referred to as the sequence-to-sequence architecture. The fundamental concept of neural machine translation is exemplified by this framework, which involves the conversion of a source text sequence into a

mathematical problem. The mathematical problem is then solved to produce a target text sequence. Refer to Fig. 5 for the specific structure.

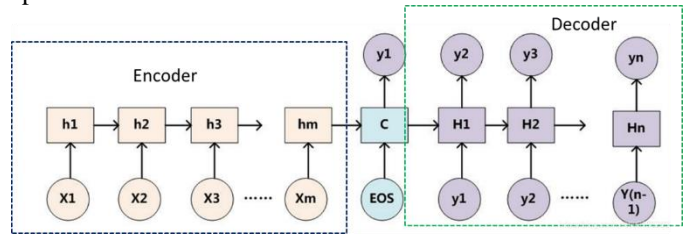


Fig. 5. Encoder-Decoder structure.

Fig. 5 illustrates that the Encoder-Decoder architecture comprises two components: the Encoder, which processes a word from the source sentence at each time step, extracting the informational properties of the source sequence. Subsequent to several time steps, all words will be condensed into the encoder's hidden states, resulting in a context vector C . The decoder receives inputs comprising the context vector C , the prior hidden states, and the previously anticipated output. The outcome of each phase serves as input for the subsequent step.

In this procedure, the decoder functions as a language model; however, this model is conditional, constructed based on the context vector C , therefore referred to as a "Conditional Language Model." The expression is stated as follows:

$$p(y) = \prod_{t=1}^n p(y_t | y_1, y_2, \dots, y_{n-1}, C) \quad (4)$$

where the output target sequence is $y = (y_1, y_2, \dots, y_n)$.

III. LSTM ENHANCED ATTENTION MECHANISM AND APPLICATIONS

A. LSTM Enhanced Attention Mechanism

This study proposes an upgraded neural machine translation model utilizing an LSTM-based attention mechanism, which is an improvement of the RNN encoder-decoder architecture. The model comprises three components:

1) *Encoder*: This component employs a bidirectional LSTM neural network (Bi-LSTM) [23], which combines the sentence information of the source sequence with the future textual information. The word embedding vectors are input into the Bi-LSTM to encode the complete source sentence, which is subsequently processed by the augmented attention module;

2) *Enhanced attention module*: this component receives all the encoder's concealed states after the source sequence is encoded at the encoder side and calculates them in conjunction with the decoder's current state to generate the dynamic context vector for the current moment. This section encompasses both Multi-Head Attention [24] and Multi-Hop Attention [25] methods.

- The multi-attention mechanism is mainly designed to fully mine the sentence information in different subspaces in the model, and the specific structure is shown in Fig. 6.

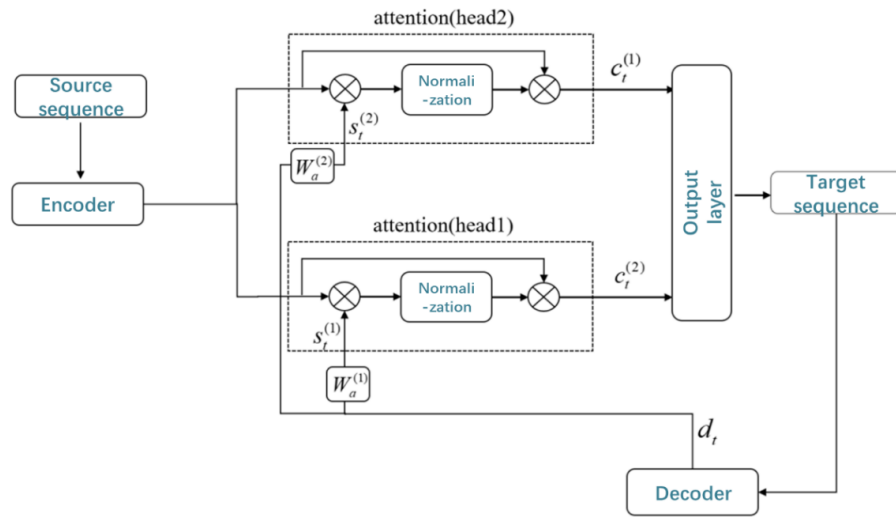


Fig. 6. Multi-attention mechanism structure.

- The multi-hop attentional mechanism mainly performs multiple attentional computations on each HEAD to extend the model's representational capability. The specific structure is shown in Fig. 7.

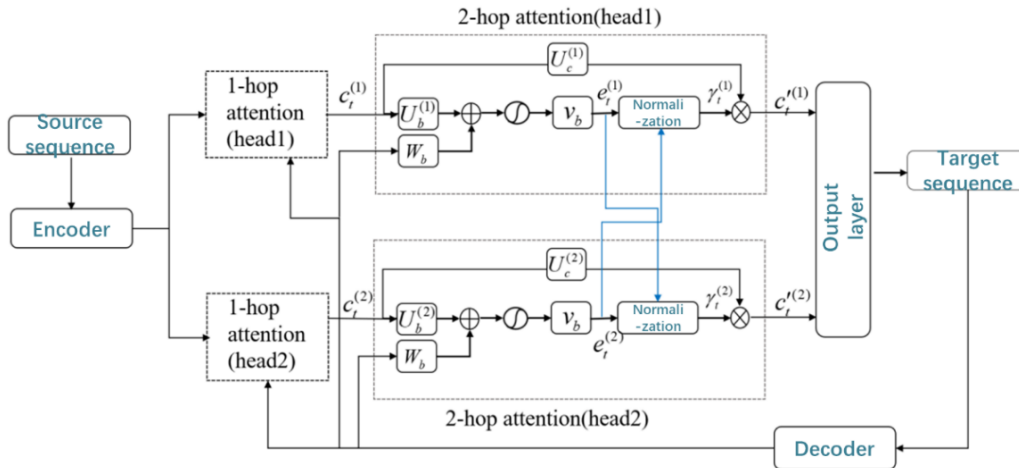


Fig. 7. Structure of multi-hop attention mechanism.

3) *Decoder*: This part is using LSTM network and receives the hidden state information from the encoder.

The overall structure of the model is shown in Fig. 8.

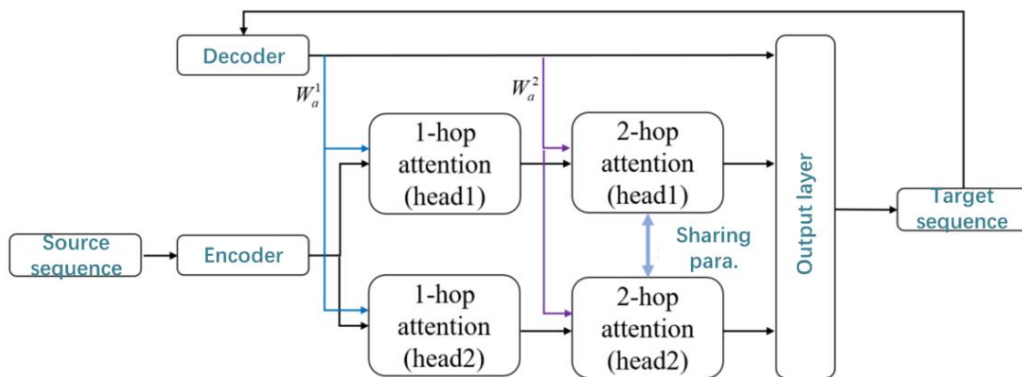


Fig. 8. General structure of the model.

B. Modelling Steps

The operational procedures of the LSTM-based neural machine translation model utilizing an increased attention mechanism are as follows:

1) Input the source sentence sequence $X = (x_1, x_2, \dots, x_T)$ into the encoder with Bi-LSTM, the forward LSTM f_{LSTM} reads the sentence sequence sequentially from x_1 to x_T and computes the forward hidden state $(\vec{h}_1, \vec{h}_2, \dots, \vec{h}_T)$; the backward LSTM \bar{f}_{LSTM} reads the sentence sequence sequentially from x_T to x_1 and computes the backward hidden state $(\vec{h}_1, \vec{h}_2, \dots, \vec{h}_T)$, and splices the forward hidden state and the backward hidden state to get the final hidden state h_t , which is computed by the following formula:

$$h_t = [\vec{h}_t; \vec{h}_t] = [LSTM_{encoder}(x, \vec{h}_{t-1}); LSTM_{encoder}(x, \vec{h}_{t+1})] \quad (5)$$

2) Deliver the hidden state h_t to the Enhanced Attention Mechanism module;

3) Calculate the 1-hop attention score. Based on the hidden state of the target sequence with respect to the previous time node, the output of the LSTM at the current moment is obtained d_t :

$$d_t = LSTM_{encoder}(\hat{y}_{t-1}, d_{t-1}) \quad (6)$$

4) Based on the trained matrix $W_a^{(k)}$, get the hidden state $s_t^{(k)}$ at the current moment:

$$s_t^{(k)} = W_a^{(k)} d_t \quad (7)$$

where k represents the k th head.

5) Calculate the context vector $c_t^{(k)}$ for the k th head:

$$c_t^{(k)} = \text{soft max}(s_t^{(k)} H_{encoder}^T) H_{encoder} \quad (8)$$

6) Calculate the attention fraction of 2-hop:

$$e_t^{(k)} = v_b^T \tanh(U_b^{(k)} c_t^{(k)} + W_b s_t^{(k)}) \quad (9)$$

7) Normalise the attention score for each HEAD to $\gamma_t^{(k)}$:

$$\gamma_t^{(k)} = \frac{\exp(e_t^{(k)})}{\sum_{n=1}^N \exp(e_t^{(n)})} \quad (10)$$

where N represents the total number of heads.

8) The trained parameters $U_t^{(k)}$, $\gamma_t^{(k)}$ and $c_t^{(k)}$ are used to compute the context vector $c_t^{n(k)}$ at the current moment with the following formula:

$$c_t^{n(k)} = \gamma_t^{(k)} U_c^{(k)} c_t^{(k)} \quad (11)$$

9) The context vector $c_t^{(k)}$ is spliced with the output of LSTM d_t , and the text feature vector is obtained by training parameters with Tanh activation function layer:

$$o_t = \tanh\left(W_o \left[d_t; c_t^{(1)}; c_t^{(2)}; \dots; c_t^{(k)} \right] \right) \quad (12)$$

10) The decoder inputs the final text vector o_t to the output layer to get the model prediction result, which is calculated as follows:

$$p(y_t | y_1, y_2, \dots, y_{t-1}, X) = \text{soft max}(o_t) \quad (13)$$

Fig. 9 illustrates the computational process of the neural machine translation model, which is based on the LSTM enhanced attention mechanism.

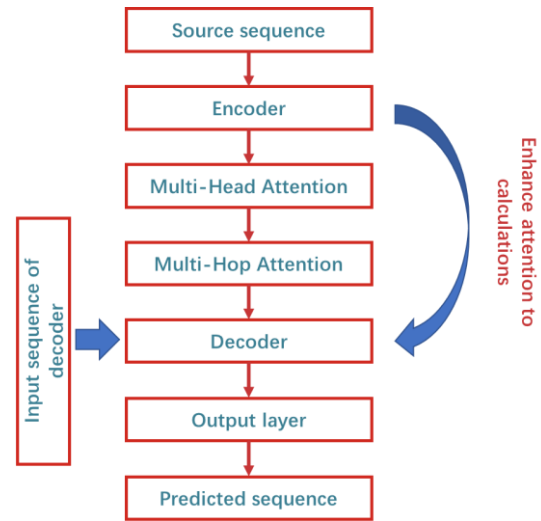


Fig. 9. Computational process of neural machine automatic translation model.

IV. MODEL EXPERIMENTS AND ANALYSES

A. Data Sets

The training dataset used for the experiments in this section is the WMT17 Chinese-English (WMT17zh-en) dataset [26], which is used to train the neural machine translation model based on the LSTM enhanced attention mechanism proposed in this chapter, and newsdev2017 and newstest2017 are used as the validation and test sets [27], respectively, and the introduction about them is shown in Table I.

TABLE I. DATA INFORMATION

Data type	Name (of a Thing)	Magnitude
Training set	WMT17zh-en	227k
Validation set	newsdev2017	4k
Test set	newstest2017	2k

The model is generally set to a fixed text length, so the pad operation is used to supplement sentences of shorter length, as shown in Fig. 10.

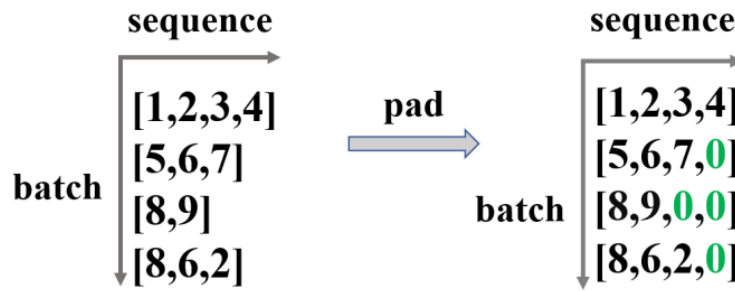


Fig. 10. Pad operation.

B. Indicators for Assessing Translation Effectiveness

In this paper, we adopt an automatic machine translation evaluation method, i.e. BLEU (Bilingual evaluation understudy) [28]. BLEU is the calculation of a similarity score between a given translation generated by a machine translation system, and a reference translation, which is used to measure the performance of this machine translation system, where the range of this score is [0,1]. The specific calculation formula is as follows:

$$BLEU = BP \cdot \exp\left(\sum_{n=1}^N w_n \log P_n\right) \quad (14)$$

$$BP = \begin{cases} 1 & c > r \\ e^{(1-r/c)} & c \leq r \end{cases} \quad (15)$$

where w_n is the weight for different n-grams, P_n is the weight of the corresponding n-element word in the sequence of reference answers, c is the length of the candidate sentence, and

r is the number of words in common between the model-translated sentence and the reference answer sentence.

C. Environmental Settings

This experiment is a deep neural network based translation model, the required experimental environment is shown in Table II, the deep learning framework used for the model experiments in this paper is Pytorch 1.8.1, which contains a large number of libraries internally for the convenience of the researcher. Bi-LSTM [29] and Conv S2S model [30] are used to compare with the proposed model, and the specific model parameter settings are shown in Table III.

V. ANALYSIS OF RESULTS

Table IV presents the BLEU scores for the Bi-LSTM, Conv S2S model, and the neural machine translation model utilizing an LSTM-enhanced attention mechanism across three datasets: WMT17zh-en, newsdev2017, and newstest2017. Table IV illustrates that the BLEU scores of the neural machine translation model utilizing an LSTM-enhanced attention mechanism are the highest across the three datasets: WMT17zh-en, newsdev2017, and newstest2017, with scores of 22.86, 23.64, and 23.14, respectively.

TABLE II. CONFIGURATION OF THE EXPERIMENTAL ENVIRONMENT

No.	Causality	Attribute value
1	CPU	Intel Intel(R) Xeon(R)
2	GPUs	Ge Force RTX 2080Ti
3	memory	24G
4	programming language	Python 3.8.3
5	Deep learning frameworks	Pytorch 1.8.1
6	operating system	Linux

TABLE III. PARAMETER SETTINGS FOR EXPERIMENTAL COMPARISON MODELS

No.	Modelling	Parameterisation
1	Bi-LSTM	Stacked 4-layer LSTM with 1000 hidden layers each, the dimension of word embedding is 1000; the number of neurons in the attention mechanism layer is 1000; the optimiser is SGD, the initial learning rate is 0.005, the batch size is 128 and the number of iterations is 50
2	Conv S2S	The dimensions of the hidden units of the encoder and decoder are 512; the optimiser is SGD; the dropout probability is set to 0.2, the initial learning rate is 0.005, the batch size is 128 and the number of iterations is 50
3	Proposed Method	2-layer Bi-LSTM with word embeddings of dimension 1024, optimiser SGD; initial learning rate 0.005, batch size 128, number of iterations 50

In order to observe the differences between the three models more intuitively, the data in Table IV were visualised as bar charts, as shown in Fig. 11. As can be seen from Fig. 11, on the WMT17zh-en data, the neural machine translation model based on the LSTM enhanced attention mechanism has a higher value of 1.44 BLEU and 0.56 BLEU than the Bi-LSTM and Conv S2S models, respectively; on the newsdev2017 data, the model proposed in this paper has a higher value of 1.44 BLEU and 0.56 BLEU than the Bi-LSTM and Conv S2S models 0.75 BLEU and 0.4 BLEU values, respectively; on newstest2017 data, the model

proposed in this paper outperforms Bi-LSTM and Conv S2S models by 1.28 BLEU and 0.36 BLEU values, respectively.

TABLE IV. EXPERIMENTAL COMPARISON MODEL OF BLEU SCORES

No.	Modelling	WMT17zh-en	newsdev2017	newstest2017
1	Bi-LSTM	21.42	23.89	21.86
2	Conv S2S	22.28	23.24	22.78
3	Proposed Method	22.86	23.64	23.14

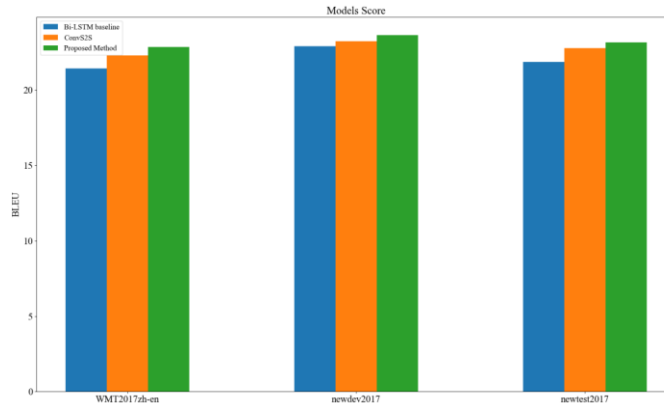


Fig. 11. BLEU scores for three models.

In order to analyse the BLEU scores of the three models under different tourism English sentence lengths, this paper investigates 11 length ranges of tourism English long sentences, and the specific results are shown in Fig. 12. Fig. 12 illustrates that despite the neural machine translation model utilizing the

LSTM-enhanced attention mechanism exhibiting a lower BLEU score compared to the Conv S2S model within the sentence length range of [20,25], it surpasses the BLEU score of the Conv S2S model for sentence lengths around 80 and exceeding 90.

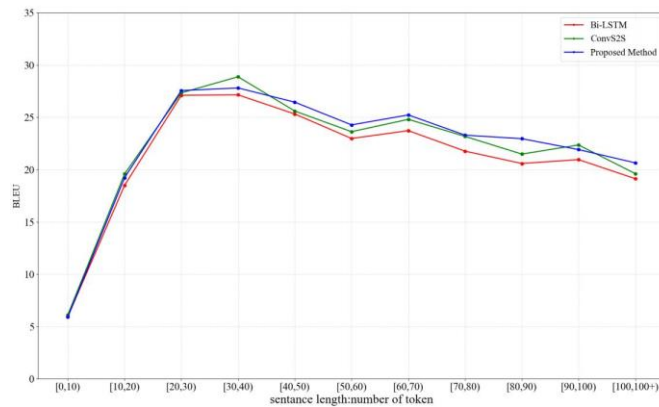


Fig. 12. BLEU scores of the three models for different sentence lengths.

VI. CONCLUSION AND OUTLOOK

This work addresses the issue of automatic translation within the context of English culture, highlighting the shortcomings of the LSTM-based encoder-decoder model, and presents a neural machine translation model that incorporates an increased attention mechanism based on LSTM. By examining pertinent theoretical methodologies and delineating the challenges of English automatic translation, a neural machine translation model utilizing an LSTM-enhanced attention mechanism is developed through the implementation of multi-hop attention computation and multi-head attention procedures. The

experimental part uses WMT17 Chinese and English datasets, newsdev2017 and newstest2017, and introduces the criteria of machine translation evaluation to measure the effect of translation quality through BLEU. The experimental comparative analysis demonstrates the effectiveness of the proposed enhanced attention mechanism, especially for translating long sentences.

Despite its effectiveness, the proposed LSTM-based enhanced attention model for tourism English translation has several limitations. First, the reliance on the WMT17 dataset may limit the model's applicability to other domains or language

pairs, as the dataset might not cover diverse linguistic features or cultural nuances. Second, while the multi-head and multi-hop attention mechanisms improve long-sequence translation, the model's complexity increases significantly, leading to higher computational costs and longer training times. Third, the translation quality heavily depends on the availability of high-quality, domain-specific training data, which remains a challenge in many low-resource contexts. Lastly, the model's performance was evaluated solely with BLEU scores, which might not fully capture the subtleties of cultural translation, such as idiomatic expressions or contextual accuracy. These limitations suggest the need for further improvements in model generalizability, efficiency, and evaluation methods.

To overcome the limitations identified, future research could focus on the following areas: Future research should focus on expanding datasets to include diverse linguistic and cultural contexts, beyond just tourism English. This can involve collecting data from multiple language pairs and incorporating low-resource languages to improve the model's adaptability. A richer dataset will ensure that the translation system captures various idiomatic expressions and cultural nuances, making the model more universally applicable and effective in handling diverse real-world scenarios.

To address the computational burden, future work could explore optimizing the model's architecture to be more lightweight without sacrificing performance. Techniques such as sparse attention mechanisms or pruning can reduce resource usage and training times. This would make the model more scalable and suitable for deployment in environments with limited computing resources, such as mobile devices or embedded systems.

Incorporating more comprehensive evaluation metrics is essential to fully capture translation quality. Beyond BLEU scores, qualitative metrics should be used to assess how well the model handles idiomatic expressions and cultural subtleties. Adding human evaluation to the testing process can provide valuable insights into the model's contextual accuracy and overall fluency, ensuring more reliable and culturally sensitive translations.

REFERENCES

- [1] An G, Tan D A L .Enhancing cross-cultural communication for Chinese tourists in non-native English-speaking destinations: a study of sociolinguistic competence and politeness challenges[J].Global Chinese, 2024, 10(2):215-240.DOI:10.1515/glochi-2023-0041.
- [2] Laskar S R , Manna R , Pakray P B S .A Domain Specific Parallel Corpus and Enhanced English-Assamese Neural Machine Translation[J].computacion y sistemas, 2022, 26(4):1669-1687.
- [3] Goldberg Y .A Primer on Neural Network Models for Natural Language Processing[J].Computer Science, 2016.DOI:10.1613/jair.4992.
- [4] Takahashi K , Sudoh K , Nakamura S .Automatic Machine Translation Evaluation using a Source and Reference Sentence with a Cross-lingual Language Model[J].Journal of Natural Language Processing, 2022, 29(1):3-22.DOI:10.5715/jnlp.29.3.
- [5] Gupta S , Bouadjenek M R , Robles-Kelly A .PERCY: A post-hoc explanation-based score for logic rule dissemination consistency assessment in sentiment classification[J].Knowledge-Based Systems, 2023.DOI:10.1016/j.knosys.2023.110685.
- [6] Ishiguro H .A Preliminary Study on Realising Human-Robot Mental Comforting Dialogue via Sharing Experience Emotionally[J].Sensors , 2022, 22.DOI:10.3390/s22030991.
- [7] Zhong T .A generic multi-level framework for building term-weighting schemes in text classification[J].The Computer Journal, 2024.DOI:10.1093/ comjnl/bxae068.
- [8] Kramov A , Pogorilyy S .Usage of the Speech Disfluency Detection Method for the Machine Translation of the Transcriptions of Spoken Language[J]. NaUKMA Research Papers. computer science, 2023.DOI:10.18523/2617-3808.2022.5.54-61.
- [9] Bilianos D , Mikros G .Sentiment analysis in cross-linguistic context: how can machine translation influence sentiment classification?[J]. Literary & linguistic computing: Journal of the Alliance of Digital Humanities Organizations, 2023.
- [10] Zhou X , Jia W , Shi C .Automatic Translation of English Terms for Computer Network Security Based on Deep Learning[J]. 2024, 20(3s):598-609.
- [11] YANG Donghua, ZOU Development, WANG Hongzhi, WANG Jinbao. SparQL query prediction based on Seq2Seq model[J]. Journal of Software, 2021.DOI:10.13328/j.cnki.jos.006171.
- [12] Gemechu E , Kanagachidambaresan G R .English-Afaan Oromo Machine Translation Using Deep Attention Neural Network[J].Optical memory & neural networks, 2023(3):32.DOI:10.3103/S1060992X23030049.
- [13] Liu W , He Y , Lan T W Z .Research on system combination of machine translation based on Transformer[J].high technology letters, 2023, 29(3):310-317.
- [14] Thin D V , Hao D N , Nguyen L T .Vietnamese Sentiment Analysis: An Overview and Comparative Study of Fine-tuning Pretrained Language Models[J].ACM transactions on Asian and low-resource language information processing, 2023(6):22.
- [15] Ma F .Construction and Evaluation of College English Translation Teaching Model Based on Multimodal Integration[J].Applied Mathematics and Nonlinear Sciences, 2024, 9(1).DOI:10.2478/amns-2024-1774.
- [16] Huang P , Zhao J , Sun S L Y .Knowledge enhanced zero-resource machine translation using image-pivoting[J].Applied Intelligence: the International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 2023, 53(7):7484-7496.
- [17] Uzma F , Shafry M R M , Adnan A .A multi-stack RNN-based neural machine translation model for English to Pakistan sign language translation[J].Neural computing & applications, 2023.DOI:10.1007/s00521-023-08424-0.
- [18] Madi S , Baba-Ali A R .A new hybrid incremental learning system for an enhanced KNN algorithm (hoKNN)[J].Evolving Systems, 2024, 15(3):1001-1019. DOI:10.1007/s12530-023-09531-y.
- [19] Edmundson H P , Oettinger A G .Automatic Language Translation[J].Mathematics of Computation, 2011, 15(74).DOI:10.2307/2004259.
- [20] Maharjan J , Garikipati A , Singh N P , Cyrus, L. , Sharma M, Ciobanu M. OpenMedLM: prompt engineering can out-form fine-tuning in medical question- answering with open-source large language models[J].Scientific Reports, 2024, 14(1).DOI:10.1038/s41598-024-64827-6.
- [21] Li N , Gao C , Jin D , Liao Q. Disentangled Modeling of Social Homophily and Influence for Social Recommendation[J].IEEE Transactions on Knowledge and Data Engineering, 2023(6):35. doi:10.1109/TKDE.2022.3185388.
- [22] Gao X , Li X , Qi G Y .GELU-LSTM-encoder-decoder fault prediction for batch processes based on the global -local percentile method[J].The Canadian Journal of Chemical Engineering, 2024, 102(6):2208-2227.DOI:10.1002/cjce.25170.
- [23] Xu Y , Liu T , Du P .Volatility forecasting of crude oil futures based on Bi-LSTM-Attention model: the dynamic role of the COVID-19 pandemic and the Russian-Ukrainian conflict[J].Resources Policy, 2024, 88.DOI:10.1016/j.resourpol.2023.104319.
- [24] Garg M , Ghosh D , Pradhan P M .Multiscaled Multi-Head Attention-Based Video Transformer Network for Hand Gesture Recognition[J].IEEE signal processing letters, 2023.DOI:10.1109/LSP.2023.3241857.
- [25] Xinyu H , Tongxuan Z , Guiyun Z .MultiHop attention for knowledge diagnosis of mathematics examination[J].Applied Intelligence: the

- International Journal of Artificial Intelligence, Neural Networks, and Complex Problem-Solving Technologies, 2023, 53(9):10636-10646.
- [26] Zheng X, Chen H L, Ma Y Q, Wang Q. A neural machine translation model incorporating dependent syntax and LSTM[J]. Journal of Harbin Institute of Technology, 2023, 28(3):20-27.DOI:10.15938/j.jhust.2023.03.003.
- [27] Sharaff A , Chowdhury T R , Bhandarkar S .LSTM based Sentiment Analysis of Financial News[J].SN Computer Science, 2023, 4:1-8.DOI:10.1007/s42979- 023-02018-2.
- [28] Wołk, Krzysztof, Marasek K .Enhanced Bilingual Evaluation Understudy[J].Computer ence, 2014.DOI:10.12720/lnit.
- [29] Ramadhan T I, Ramadhan N G , Supriatman A .Implementation of Neural Machine Translation for English-Sundanese Language using Long Short Term Memory (LSTM)[J].Building of Informatics, Technology and Science (BITS), 2022.DOI:10.47065/bits.v4i3.2614.
- [30] Tiwari G , Sharma A , Sahotra A , Kapoor R. English-Hindi Neural Machine Translation-LSTM Seq2Seq and ConvS2S[C]//2020 International Conference on Communication and Signal Processing (ICCSP).2020.DOI:10.1109/ICCSP48568.2020.9182117.

A Novel Metric-Based Counterfactual Data Augmentation with Self-Imitation Reinforcement Learning (SIL)

K. C. Sreedhar¹, T. Kavaya², J. V. S. Rajendra Prasad³, V. Varshini⁴

Associate Professor, Department of CSE, Sreenidhi Institute of Science and Technology, Hyderabad, India¹
Student, Department of CSE, Sreenidhi Institute of Science and Technology, Hyderabad, India^{2, 3, 4}

Abstract—The inherent biases present in language models often lead to discriminatory predictions based on demographic attributes. Fairness in NLP refers to the goal of ensuring that language models and other NLP systems do not produce biased or discriminatory outputs that could negatively affect individuals or groups. Bias in NLP models often arises from training data that reflects societal stereotypes or imbalances. Robustness in NLP refers to the ability of a model to maintain performance when faced with noisy, adversarial, or out-of-distribution data. A robust NLP model should handle variations in input effectively without failing or producing inaccurate results. The proposed approach employs a novel metric called CFRE (Context-Sensitive Fairness and Robustness Evaluation) designed to measure both fairness and robustness of an NLP model under different contextual shifts. Next, it projected the benefits of this metric in terms of experimental parameters. Next, the work integrated counterfactual data augmentation with help of Self-Imitation Reinforcement Learning (SIL) to reinforce successful counterfactual generation by enabling the model to learn from its own high-reward experiences, fostering a more balanced understanding of language. The integration of SIL allows for efficient exploration of the action space, guiding the model to consistently produce unbiased outputs across different contexts. The proposed approach demonstrates the effectiveness of our method through extensive experimentation and compared the results of the proposed metric with that of WEAT and SMART testing, and showed a significant reduction in bias without compromising the model's overall performance. This framework not only addresses bias in existing models but also contributes to a more robust methodology for training fairer NLP systems. Both the proposed metric and SIL showed better results in experimental parameters.

Keywords—Natural language processing; fairness, robustness; Word Embedding Association Test (WEAT); SMART testing

I. INTRODUCTION

Natural Language Processing (NLP) serves as a linchpin in enabling seamless human-computer interaction, fostering intuitive communication through interfaces like voice assistants and chatbots. It empowers the automation of text analysis, expediting tasks such as sentiment assessment, document summarization, and content categorization with unparalleled efficiency. By transcending linguistic barriers, NLP promotes global interconnectivity, facilitating multilingual translation and cultural localization.

Its contributions to AI advancements are transformative, powering sophisticated systems like personalized virtual assistants and predictive analytics. NLP is instrumental in extracting actionable insights from unstructured textual data, supporting informed decision-making in critical domains like healthcare, finance, and governance. Furthermore, it champions inclusivity by fostering the development of assistive technologies, such as speech-to-text systems and screen readers, to accommodate individuals with disabilities.

By addressing linguistic diversity and automating complex textual processes, NLP is not merely a technological tool but a catalyst for innovation and inclusivity in the digital age.

Natural Language Processing, a subfield of Artificial Intelligence, has become pivotal in automating and enhancing communication, yet its deployment raises pressing concerns around fairness and robustness. At its core, fairness in NLP pertains to the equitable and unbiased performance of language models across diverse demographic and linguistic groups. Robustness, conversely, measures a model's resilience to adversarial inputs, distributional shifts, or unexpected variations in data. Together, these dimensions are critical to ensuring the ethical and reliable use of NLP technologies.

One of the primary fairness challenges arises from biased training datasets, which reflect historical inequities, stereotypes, or regional disparities. These biases, embedded in language corpora, can perpetuate societal injustices when reflected in model outputs. For instance, gendered pronoun resolution systems may reinforce occupational stereotypes by associating women with caregiving roles and men with leadership positions.

Robustness, on the other hand, is tested when models face adversarial attacks or operate in low-resource settings. Subtle manipulations in input texts—like typos or syntax changes—can disproportionately degrade model performance. Similarly, underrepresentation of certain languages, dialects, or sociolects exacerbates the risk of exclusionary AI systems that fail to generalize effectively.

The interplay of these issues creates a dual imperative: to mitigate inherent biases while enhancing models' adaptability across varied scenarios. Ethical considerations are further compounded by the lack of standardized benchmarks for measuring fairness and robustness. Solutions often involve trade-offs, as techniques that improve robustness, like data augmentation, may inadvertently amplify biases.

Addressing these challenges requires a multi-faceted approach. Incorporating diverse, high-quality datasets and developing fairness-aware training algorithms are pivotal steps. Furthermore, interdisciplinary collaboration—spanning computational linguistics, ethics, and social sciences—can provide nuanced perspectives to inform NLP research. Regular audits, explainable AI methods, and inclusive design principles are essential to embedding trustworthiness into language technologies.

In conclusion, fairness and robustness are not merely technical hurdles but societal imperatives in the age of pervasive AI. As NLP systems permeate sensitive domains like hiring, healthcare, and legal adjudication, ensuring their ethical and equitable deployment becomes a moral obligation. The paper is organized as follows. Section I gives introduction the problem of bias in NLP. Section II gives explains types of bias in NLP. Section III gives various existing metrics for measuring bias. Section IV explains briefing, challenges of robustness and robustness contextual evaluation respectively. Experimental results is given in Section V and finally, the paper is concluded in Section VI.

1) *The Problem of Bias in NLP*: Bias in Natural Language Processing (NLP) refers to the systematic favoritism or prejudice exhibited by language models, often stemming from imbalances or stereotypes present in their training data. This phenomenon undermines the equity, reliability, and ethicality of NLP systems, leading to unintended discriminatory consequences. Bias is particularly critical in applications influencing high-stakes decisions, such as hiring algorithms, legal systems, and healthcare tools, where such predispositions can perpetuate societal inequities [1-3].

At its root, bias arises from the data-driven nature of NLP models, which inherit the flaws, prejudices, and imbalances embedded in the corpora used for training. When these systems process text, they often reinforce or amplify existing stereotypes, inadvertently perpetuating harm against underrepresented or marginalized groups. Addressing bias is a multifaceted challenge that requires understanding its various types and manifestations.

II. TYPES OF BIAS IN NLP

1) *Representation bias*: This form of bias originates in training datasets that over represent certain groups or perspectives while neglecting others. For example, texts predominantly authored in English may marginalize speakers of minority languages or dialects, perpetuating cultural hegemony.

2) *Stereotypical bias*: Models can perpetuate harmful stereotypes, such as associating certain professions with specific genders or ethnicities. For instance, a model might predict "nurse" as a woman or "engineer" as a man based on biased correlations in training data.

3) *Historical bias*: Historical biases reflect long-standing societal inequities embedded in data. Even if collected neutrally, datasets often capture systemic inequalities, such as

racial or gender disparities, which are then reflected in the model's predictions.

4) *Selection bias*: This bias arises from skewed data collection processes. If a training dataset is predominantly drawn from urban populations, for instance, the resulting model may fail to generalize to rural or less technologically advanced contexts.

5) *Aggregation bias*: When data from diverse groups are aggregated into a single dataset, the unique characteristics of minority groups may be overshadowed by majority trends, leading to homogenized outputs that overlook nuanced needs.

6) *Interaction bias*: This bias emerges during user interaction with NLP systems. For example, users' queries can introduce biases that models then propagate, such as autocomplete suggestions that reinforce prejudiced or inappropriate language.

7) *Temporal bias*: Temporal bias stems from the use of outdated data that fails to account for societal evolution. For instance, older datasets might include terms or perspectives that are now considered offensive or obsolete.

8) *Implicit bias*: Implicit biases are more subtle and embedded within the model's architecture, often surfacing in nuanced contexts such as sentiment analysis or content moderation, where subjective judgments are involved.

A. Metrics for Assessing Bias

Quantifying bias in NLP systems is a multifaceted task that requires metrics capable of identifying disparities, imbalances, and stereotypical tendencies. These metrics enable researchers to evaluate the degree of bias and its impact, facilitating informed strategies for mitigation. Below is an overview of commonly used metrics for measuring bias in NLP, along with their mathematical formulations:

1) *Statistical Parity Difference (SPD)*: This metric evaluates whether the outcomes for different demographic groups are equally distributed.

$$SPD = P(Y=1|G=g1) - P(Y=1|G=g2)$$

- Y : Model outcome (e.g., positive or negative sentiment).
- G : Demographic group ($g1, g2$ represent different groups, e.g., male and female).
- A value of 0 indicates perfect fairness, while deviations suggest bias.

2) *Equal Opportunity Difference (EOD)*: This metric focuses on the equality of true positive rates across groups, ensuring that all groups have equal chances of achieving favorable outcomes when eligible.

$$EOD = P(\hat{Y}=1|Y=1, G=g1) - P(\hat{Y}=1|Y=1, G=g2)$$

- \hat{Y} : Predicted outcome.
- Ensures fairness specifically for eligible or qualified individuals.

3) *Conditional Demographic Disparity (CDD)*: This metric measures bias in model predictions while controlling for specific contextual variables.

$$CDD = P(\hat{Y} = 1 | X=x, G=g1) - P(\hat{Y} = 1 | X=x, G=g2)$$

- X: Contextual variables, such as input features.
- Helps identify disparities conditional on input attributes.

4) *Word Embedding Association Test (WEAT)*: This metric quantifies bias in word embeddings by measuring the association between target words and attribute word sets.

$$WEAT = \frac{\text{mean}(s(w, A) - s(w, B))}{\text{std}(s(w, A) - s(w, B))}$$

- w: Target word.
- A, B: Two sets of attribute words (e.g., male- and female-associated words).
- s(w, A): Cosine similarity between w and words in set A.
- A high WEAT score indicates stronger associations, reflecting potential biases.

5) *Bias Amplification Index (BAI)*: This measures the extent to which a model amplifies existing biases in data.

$$BAI = \frac{\text{Bias in Model Output}}{\text{Bias in Training Data}}$$

Ratios greater than 1 indicate that the model exacerbates bias.

6) *Directional Bias Metric (DBM)*: This metric evaluates bias in sentence or text-level outputs by analyzing directional shifts in embeddings.

$$DBM = \frac{\sum_{i=1}^n \cos(\vec{e}_i, \vec{d})}{n}$$

\vec{e}_i : Embedding of Sentence i

\vec{d} : Bias direction vector

n: Total sentences

7) *Mutual Information Difference (MID)*: This metric captures the disparity in the information shared between model predictions and sensitive attributes.

$$MID = I(\hat{Y}; G=g1) - I(\hat{Y}; G=g2)$$

- I: Mutual information between predictions \hat{Y} and group G.
- A high MID score reflects unequal representation of sensitive attributes in predictions.

8) *KL Divergence for Demographic Representation (KLD)*: This measures the divergence between the distributions of outcomes for different demographic groups.

$$KLD(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

- P(i): Outcome distribution for group g1
- Q(i): Outcome distribution for group g2
- Lower divergence values indicate better fairness.

9) *Bias Direction Magnitude (BDM)*: This quantifies the degree of separation between different demographic groups in embedding space.

$$BDM = ||\text{mean}(\vec{e}_{g1}) - \text{mean}(\vec{e}_{g2})||$$

$\vec{e}_{g1}, \vec{e}_{g2}$: Embeddings for groups g1 and g2.

10) *Token Probability Disparity (TPD)*: This metric measures bias in token-level predictions for specific sensitive terms.

$$TPD = P(\text{token}|G=g1) - P(\text{token}|G=g2)$$

Highlights disparities in word usage or token generation probabilities.

These metrics provide nuanced perspectives on bias in NLP systems, addressing its various dimensions, such as representation, prediction fairness, and embedding neutrality. Combining multiple metrics is essential for comprehensive evaluation, as bias often manifests in subtle and multifaceted ways.

B. Robustness in NLP

Robustness in NLP refers to the ability of a model to maintain performance when faced with noisy, adversarial, or out-of-distribution data. A robust NLP model should handle variations in input effectively without failing or producing inaccurate results.

C. Example of Robustness Challenges

1) *Adversarial attacks*: An NLP model trained to classify movie reviews as positive or negative might be tricked by inserting inconspicuous typos or irrelevant phrases. For example, changing "The movie was great!" to " The moovie was gr8!" should ideally still yield a positive classification.

2) *Context sensitivity*: An NLP system that performs well on one data distribution (e.g., news articles) may fail on another (e.g., social media text) if it's not robustly trained.

D. Robustness Improvement Techniques

1) *Adversarial training*: Including perturbed or adversarial examples during training so that the model learns to be resilient.

2) *Augmentation with noisy data*: Training on data that has been altered to include variations such as different spelling, slang, or paraphrasing helps models generalize better.

3) *Balancing fairness and robustness*: Improving fairness often involves altering the data or the training process to mitigate biases, which can sometimes reduce robustness if not done carefully. Conversely, making a model highly robust through general training methods may not necessarily address

inherent biases. The challenge lies in designing approaches that optimize both.

E. SMART Testing

SMART Testing is a methodological paradigm for systematically evaluating NLP systems across diverse dimensions, emphasizing their fairness, robustness, and adaptability. The acronym SMART encapsulates Sensitive attributes, Multiple subpopulations, Artifacts, Reasoning abilities, and Temporal changes, reflecting the multifaceted nature of NLP evaluation. While the framework does not have a universally fixed mathematical formulation, key metrics and equations can be used to assess these dimensions.

1) *Sensitive attributes (Fairness metrics)*: This component assesses disparities in performance between demographic groups with respect to sensitive attributes like gender or ethnicity. A commonly used fairness metric is Statistical Parity Difference (SPD):

$$SPD = |P(\hat{Y} = 1 | G = g1) - P(\hat{Y} = 1 | G = g2)|$$

Where:

- \hat{Y} : Model's predicted outcome.
- G: Demographic groups (g1 and g2 represent different groups).

A value closer to zero denotes minimal bias.

2) *Multiple subpopulations (Subgroup disparities)*: This dimension examines model performance across distinct subpopulations within the data. Disparities are quantified using subgroup metrics such as accuracy variance:

$$Variance = \frac{\sum_{i=1}^n (P_i - \mu)^2}{n}$$

Where,

- P_i is Model performance for subgroup i.
- μ is mean performance across all subgroups

A high variance indicates uneven performance among subgroups.

3) *Artifacts (Sensitivity to spurious patterns)*: Artifacts represent unintended correlations in training data that can lead to spurious model predictions. Artifact sensitivity can be measured by comparing performance on artifact-augmented data to baseline data:

$$Artifact\ Sensitivity = \frac{Performance_{artifact}}{Performance_{baseline}}$$

Ratios significantly deviating from 1 suggest a susceptibility to artifacts.

4) *Reasoning abilities (Cognitive robustness)*: This evaluates the model's logical and linguistic reasoning abilities under adversarial transformations or complex scenarios. Robustness against transformations is defined as:

$$Robustness\ Score(RS) = \frac{Post\ Transformation\ Accuracy}{Baseline\ Accuracy}$$

It is stated that higher scores signify greater resistance to input perturbations.

5) *Temporal changes (Adaptability over time)*: This aspect assesses how well the model performs as linguistic norms evolve. Temporal robustness is evaluated by measuring performance deviation across time-stamped datasets.

$$Temporal\ Deviation(TD) = |Performance_{t1} - Performance_{t2}|$$

It is stated that smaller deviations reflect higher adaptability to temporal variations.

6) *Aggregated SMART score*: To provide a unified view, an aggregated score can be computed as a weighted combination of the individual dimensions:

$$SMART\ Score = w1 \cdot SPD + w2 \cdot Variance + w3 \cdot Sensitivity + w4 \cdot Robustness\ Score + w5 \cdot Temporal\ Deviation$$

Where w1, w2, w3, w4, and w5 are weights reflecting the relative importance of each dimension.

III. PROPOSED NOVEL METRIC-CONTEXT-SENSITIVE FAIRNESS AND ROBUSTNESS (CFRE)

The proposed Context-Sensitive Fairness and Robustness Evaluation (CFRE) metric is designed to measure both fairness and robustness of an NLP model under different contextual shifts [3-9]. Below is the mathematical formulation of the proposed metric:

A. CFRE Metric Components

1) *Fairness Impact Score (FIS)*: The Fairness Impact Score evaluates the difference in output distributions when the model is tested with original data (O_{orig}) and perturbed data (O_{pert}) across different demographic or context groups (G_i).

$$FIS = \frac{1}{|G|} \sum_{i=1}^{|G|} D_{KL}(P(O_{orig}|G_i) || P(O_{pert}|G_i))$$

Where

- D_{KL} is Kullback-Leibler (KL) divergence.
- $P(O_{orig}|G_i)$ and $P(O_{pert}|G_i)$ are probability distributions of outputs for groups G_i in original and perturbed cases respectively.
- $|G|$ is number of distinct groups being evaluated.

2) *Robustness Contextual Evaluation (RCE)*: The robustness contextual evaluation (RCE) measures the stability of model predictions by computing the cosine similarity between output vectors from original and perturbed data (O_{orig}) and (O_{pert}) respectively.

$$RCE = \frac{1}{N} \sum_{j=1}^N \frac{O_{orig}^j \cdot O_{pert}^j}{\|O_{orig}^j\| \|O_{pert}^j\|}$$

where

- N is the number of samples.
- O_{orig}^j and O_{pert}^j are the output vectors for j^{th} sample in the original and perturbed data sets.

3) *Combined CFRE score*: The overall CFRE score can be weighted combination of the FIS and RCE to balance fairness and robustness.

$$CFRE = \alpha * FIS + \beta * RCE$$

Where α and β are weights that control importance of each component.

This formulation allows us to assess not just how fair is model across different contexts but also how consistently it performs when subject to contextual variations.

In the context of the CFRE metric, the interpretations for FIS RCE, and combined CFRE are given as below.

a) Fairness Impact Score (FIS):

- Interpretation: A higher FIS value indicates a greater divergence between the original and perturbed model outputs, suggesting that the model's fairness is more sensitive to contextual changes. This can mean the model exhibits potential biases when tested with varied input conditions, highlighting fairness issues.
- Lower FIS: Implies that the model maintains fairness across different demographic or context groups, showing resilience to contextual shifts.

b) Robustness Contextual Evaluation (RCE):

- Interpretation: This score reflects how similar the model's outputs remain under perturbations. A higher RCE value means the model is more robust, maintaining consistent behavior even when inputs are contextually modified.
- High RCE: Indicates strong robustness, where the model produces stable outputs across different contexts.
- Lower RCE: Suggests that the model's predictions are more context-dependent and can vary significantly with slight input changes.

c) Overall CFRE Value:

- Combined Score: The weighted sum of FIS and RCE allows us to evaluate both fairness and robustness together.
- High CFRE with balanced weights: Implies that the model is sensitive to contextual shifts (indicating fairness issues) but also robust in maintaining consistent outputs under certain conditions.
- Lower CFRE: Indicates that the model is more fair and robust across various tested contexts, demonstrating resilience and equitable behavior.

IV. INTEGRATING CFRE METRIC INTO SELF IMITATION LEARNING (SIL)

A. Introduction to Self-Imitation Learning (SIL)

Self-Imitation Learning (SIL) is an advanced reinforcement learning technique that enables agents to learn from past experiences, even suboptimal ones, by revisiting previously successful trajectories. Unlike traditional reinforcement learning, which often prioritizes exploration or maximizing immediate reward signals, SIL leverages historical data to reinforce and improve upon earlier decisions. It is particularly effective in complex environments where exploration is expensive or risky, as it capitalizes on self-generated "expert" demonstrations to refine policy optimization. By integrating memory-based learning with reinforcement dynamics, SIL demonstrates resilience in solving tasks requiring long-term planning and precise decision-making [10-18].

B. Main Idea behind Self-Imitation Learning (SIL)

At its core, Self-Imitation Learning revolves around the principle of leveraging an agent's historical successes as pseudo-demonstrations for future improvement. Unlike standard reinforcement learning paradigms, which discard suboptimal trajectories, SIL recognizes that even suboptimal actions can contain valuable information for solving complex tasks. This is particularly important in environments with sparse or delayed rewards, where the exploration of new policies might fail to yield immediate benefits.

SIL achieves this by employing a replay buffer, which stores trajectories (sequences of states, actions, and rewards) that yielded above-average returns. These trajectories are treated as guiding examples, and the agent revisits them during training to imitate its own past successes. This imitation process is formalized through a self-imitation loss function, which adjusts the policy to reproduce actions from successful trajectories.

The central innovation of SIL lies in its ability to balance exploitation and exploration dynamically. While traditional methods often face a trade-off between exploiting known strategies and exploring new possibilities, SIL introduces a mechanism where self-imitation augments learning efficiency without stifling exploration. This enables the agent to improve incrementally, even in scenarios where external rewards are scarce or noisy.

Moreover, SIL is robust to noise and imperfect demonstrations, as it does not rely on external expert input but instead generates its training data from its own interactions with the environment. This self-reliant nature makes it highly scalable and adaptable to diverse tasks, from robotics to game-playing.

In essence, SIL represents a shift from purely reward-driven learning to a hybrid framework that integrates self-guidance, allowing agents to harness the full potential of their past experiences for future success. By embracing both imitation and exploration, it achieves greater sample efficiency and stability in training, setting a new benchmark for learning in complex and uncertain domains.

C. Integrating CFRE with SIL

The CFRE metric is a performance measure designed to evaluate the trade-off between fairness and reward optimization in reinforcement learning. Integrating CFRE into Self-Imitation Learning (SIL) involves modifying the SIL framework to consider fairness explicitly during the learning process. The Algorithm 1 shows the CFRE integrated into SIL. Integrating the CFRE metric into Self-Imitation Learning (SIL) can effectively scale to real-world NLP systems operating in resource-constrained environments by prioritizing fairness and reward efficiency in model training. The approach allows selective reuse of high-reward, fairness-optimized trajectories, reducing computational overhead while maintaining equitable outcomes. By leveraging the CFRE metric's adaptability, the framework aligns with limited-resource constraints, improving both performance and inclusivity without excessive reliance on additional data or computing power. This ensures robust deployment of NLP systems in diverse, real-world scenarios.

Algorithm-1: CFRE-Integrated SIL

1. **Initialize:**
 - a) Define the environment E, action space A, and state space S.
 - b) Initialize SIL's policy $\pi_\theta(a|s)$, replay buffer B and reward function R(s,a).
 - c) Set the CFRE threshold τ , which balances fairness and efficiency.
2. **Collect Experience:**
 - a) Interact with the environment to generate trajectories $\tau = (s_t, a_t, r_t, s_{t+1})$ using the current policy π_θ .
 - b) Add the trajectories to the replay buffer B.
3. **Compute CFRE Metric:**
 - a. For each trajectory τ , compute the FIS and CRE :
 - b. $CFRE(\tau) = \alpha \cdot FIS(\tau) + \beta \cdot CRE(\tau)$ Where:
 - i. α, β : weights balancing fairness and reward efficiency.
 - ii. $CRE(\tau) = \text{Sum of rewards} / \text{Length of Trajectory}$

- iii. FIS (τ): Fairness computed using sensitive attributes or group-specific metrics.

c. Retain trajectories with $CFRE(\tau) \geq \tau$ in B.

4. Update Policy:

Use the retained trajectories from B to compute the SIL loss:

- a) SIL loss:
$$L_{SIL} = -\log(\pi_\theta(a|s)) \cdot (R_{expected}(s) - R_{observed}(s))$$
- b) Apply gradient descent to minimize L_{SIL} .

5. Test Policy:

- a) Evaluate the updated policy using CFRE and track performance metrics such as fairness scores, reward efficiency, and overall task accuracy.

6. Repeat:

Continue the process for a predefined number of episodes or until convergence by repeating steps 2 to 5.

V. EXPERIMENTAL RESULTS

The experiment was conducted using Crow-S pairs data set on Google Colab platform of python version 3.11.8. The Crow-S pairs dataset is a benchmark specifically designed to measure biases in NLP models, focusing on sensitive social attributes like gender, race, and socioeconomic status.

It consists of sentence pairs where one sentence carries subtle bias while the other is neutral, enabling the evaluation of a model's fairness by observing its scoring discrepancies. By systematically exposing latent stereotypes or prejudiced behavior in model outputs, the dataset also tests the robustness of NLP systems against biased linguistic patterns, helping to create more equitable language technologies.

At first, we project the graph showing comparison of original and perturbed scores using CFRE as given in Fig. 1. Next, we project the density over scores of WEAT, SMART testing as given in Fig. 2. Next, we project mean scores for various metrics as given in Fig. 3. Finally, we project graphs for average loss versus epochs and average reward versus epochs as given in Fig. 4. Fig. 1 to Fig. 3 showed significant improvement in results in proposed CFRE metric.

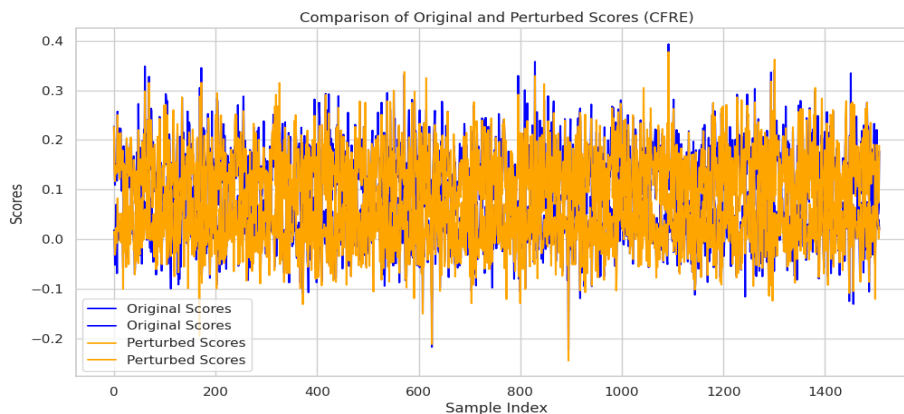


Fig. 1. Comparison of original and perturbed scores for CFRE metric.

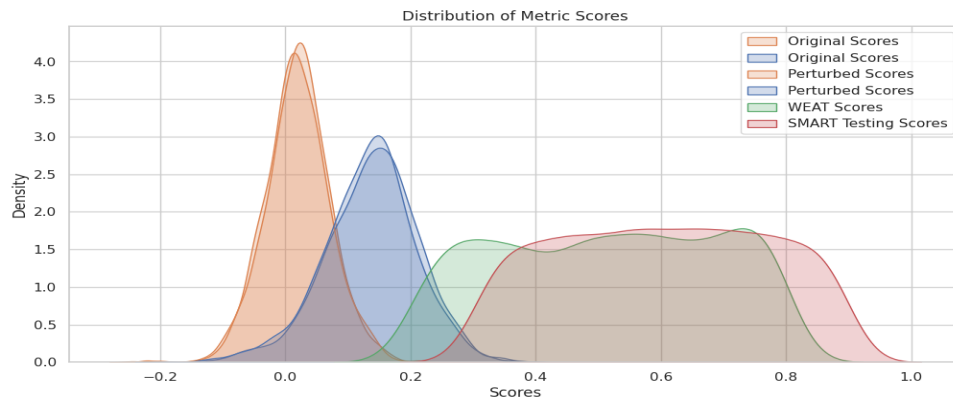


Fig. 2. Density versus scores of various metrics.

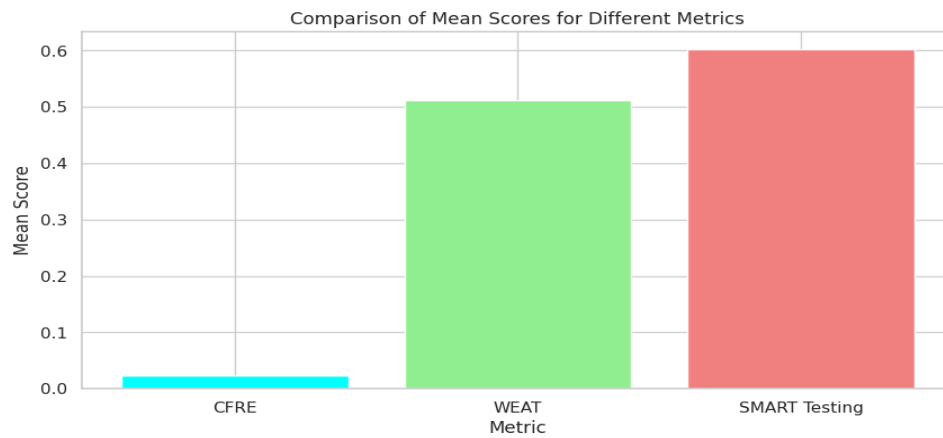


Fig. 3. Mean scores versus various metrics.

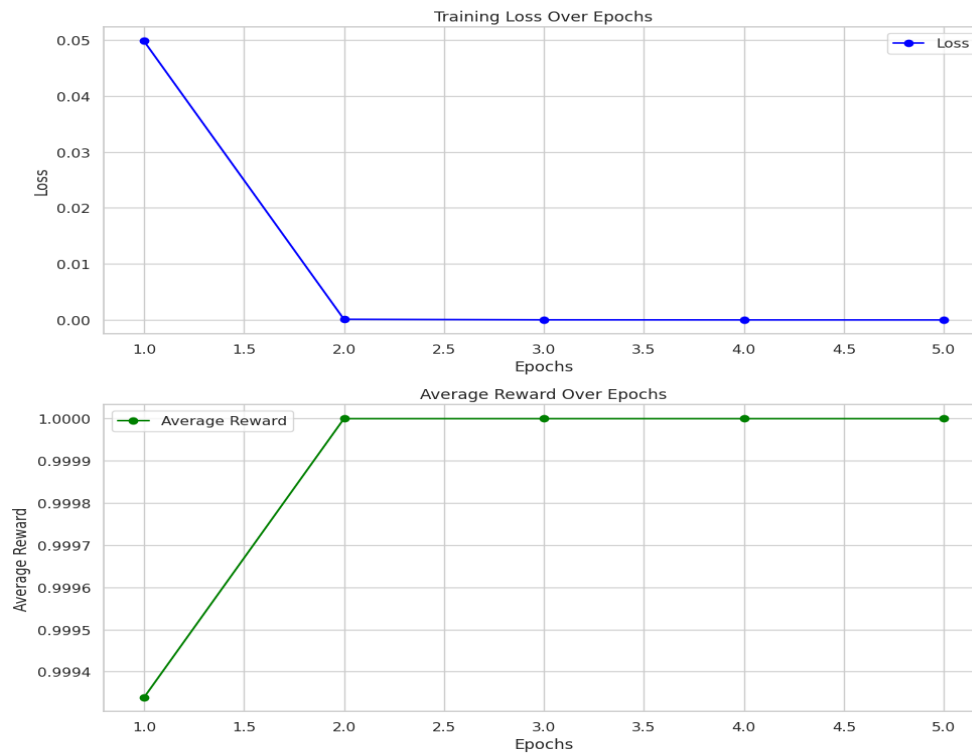


Fig. 4. Graph for loss versus Epochs and Average reward versus Epochs.

VI. CONCLUSION

The integration of the CFRE metric with Self-Imitation Learning (SIL) presents a powerful paradigm for achieving fairness, robustness, and efficiency in reinforcement learning-based NLP systems. This approach ensures that models not only optimize rewards but also address systemic biases, promoting equitable outcomes. By leveraging past successes with fairness-aware constraints, it balances performance and inclusivity, making it especially viable for resource-constrained and real-world applications.

The proposed metric outperformed other existing metrics like WEAT and SMART testing. Also, it got low mean score compared to that of these metrics. The variation between original and perturbed scores serves as a measure of the model's robustness. A narrow difference signifies that the model is resistant to input alterations, showcasing its stability, whereas a wider discrepancy indicates that the model is more vulnerable to adversarial changes or biased modifications in the input data.

REFERENCES

- [1] Ribeiro, M. T., Wu, T., Guestrin, C., & Singh, S. (2020). Beyond accuracy: Behavioral testing of NLP models with CheckList. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 4902–4912. <https://aclanthology.org/2020.acl-main.442>
- [2] Bansal, R. (2022). A Survey on Bias and Fairness in Natural Language Processing. *ArXiv, abs/2204.09591*.
- [3] Rauba, Paulius & Seedat, Nabeel & Luyten, Max & Schaar, Mihaela. (2024). Context-Aware Testing: A New Paradigm for Model Testing with Large Language Models. 10.48550/arXiv.2410.24005.
- [4] Luke Oakden-Rayner, Jared Dunnmon, Gustavo Carneiro, and Christopher Re. Hidden stratification causes clinically meaningful failures in machine learning for medical imaging. In Proceedings of the ACM Conference on Health, Inference, and Learning, pages 151–159, 2020.
- [5] Harini Suresh, Jen J Gong, and John V Guttag. Learning tasks for multitask learning: Heterogenous patient populations in the ICU. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 802–810, 2018.
- [6] Karan Goel, Albert Gu, Yixuan Li, and Christopher Re. Model patching: Closing the subgroup performance gap with data augmentation. In International Conference on Learning Representations, 2020.
- [7] Angel Alexander Cabrera, Minsuk Kahng, Fred Hohman, Jamie Morgenstern, and Duen Horng Chau. Discovery of intersectional bias in machine learning using automatic subgroup generation. In ICLR Debugging Machine Learning Models Workshop, 2019.
- [8] Boris van Breugel, Nabeel Seedat, Fergus Imrie, and Mihaela van der Schaar. Can you rely on your model evaluation? improving model evaluation with synthetic test data. *Advances in Neural Information Processing Systems*, 36, 2024.
- [9] Nabeel Seedat, Fergus Imrie, and Mihaela van der Schaar. Navigating data-centric artificial intelligence with DC-Check: Advances, challenges, and opportunities. *IEEE Transactions on Artificial Intelligence*, 2023.
- [10] Oleg S Pinykh, Georg Langs, Marc Dewey, Dieter R Enzmann, Christian J Herold, Stefan O Schoenberg, and James A Brink. Continuous learning AI in radiology: Implementation principles and early applications. *Radiology*, 297(1):6–14, 2020.
- [11] Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanus Phillips, Irena Gao, Tony Lee, Etienne David, Ian Stavness, Wei Guo, Berton A. Earnshaw, Imran S. Haque, Sara Beery, Jure Leskovec, Anshul Kundaje, Emma Pierson, Sergey Levine, Chelsea Finn, and Percy Liang. WILDS: A benchmark of in-the-wild distribution shifts. In International Conference on Machine Learning, pages 5637–5664. PMLR, 2021.
- [12] Kayur Patel, James Fogarty, James A Landay, and Beverly Harrison. Investigating statistical machine learning as a tool for software development. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 667–676, 2008.
- [13] Lea Goetz, Nabeel Seedat, Robert Vandersluis, and Mihaela van der Schaar. Generalization—a key challenge for responsible ai in patient-facing clinical applications. *npj Digital Medicine*, 7 (1):126, 2024.
- [14] Maire A Duggan, William F Anderson, Sean Altekruze, Lynne Penberthy, and Mark E Sherman. The surveillance, epidemiology and end results (SEER) program and pathology: towards strengthening the critical relationship. *The American Journal of Surgical Pathology*, 40(12):e94, 2016.
- [15] Yeounoh Chung, Tim Kraska, Neoklis Polyzotis, Ki Hyun Tae, and Steven Euijong Whang. Slice finder: Automated data slicing for model validation. In 2019 IEEE 35th International Conference on Data Engineering (ICDE), pages 1550–1553. IEEE, 2019.
- [16] Svetlana Sagadeeva and Matthias Boehm. Sliceline: Fast, linear-algebra-based slice finding for ml model debugging. In Proceedings of the 2021 International Conference on Management of Data, pages 2290–2299, 2021.
- [17] Adebayo Oshingbesan, Winslow Georgos Omondi, Girmaw Abebe Tadesse, Celia Cintas, and Skyler Speakman. Beyond protected attributes: Disciplined detection of systematic deviations in data. In Workshop on Trustworthy and Socially Responsible Machine Learning, NeurIPS 2022, 2022.
- [18] Shi, Zijing & Xu, Yunqiu & Fang, Meng & Chen, Ling. (2023). Self-imitation Learning for Action Generation in Text-based Games. 703–726. 10.18653/v1/2023.eacl-main.50.

Segmentation of Nano-Particles from SEM Images Using Transfer Learning and Modified U-Net

Sowmya Sanan V^{1*}, Rimal Isaac R S²

Research Scholar, Department of Nanotechnology, Noorul Islam Centre for Higher Education,
Kumaracoil, Thuckalay, Tamil Nadu, India¹
Assistant Professor, Department of Nanotechnology, Noorul Islam Centre for Higher Education,
Kumaracoil, Thuckalay, Tamil Nadu, India²

Abstract—Nanomaterials, owing to their distinctive features, are crucial across numerous scientific domains, especially in materials science and nanotechnology. Precise segmentation of Scanning Electron Microscope (SEM) images is essential for evaluating attributes such as nanoparticle dimensions, morphology, and distribution. Conventional image segmentation techniques frequently prove insufficient for managing the intricate textures of SEM images, resulting in a laborious and imprecise process. In this research, a modified U-Net architecture is presented to tackle this challenge, utilizing a ResNet50 backbone pre-trained on ImageNet. This model utilizes the robust feature extraction abilities of ResNet50 alongside the effective segmentation performance of U-Net, hence improving both accuracy and computational efficiency in TiO₂ nanoparticle segmentation. The suggested model was assessed using performance metrics including accuracy, precision, recall, IoU, and Dice Coefficient. The results indicated a high segmentation accuracy, demonstrated by a Dice score of 0.946 and an IoU of 0.897, with little variability reflected in standard deviations of 0.002071 and 0.003696, respectively, over 200 epochs. The comparison with existing methods demonstrates that the proposed model surpasses previous approaches by attaining enhanced segmentation accuracy. The modified U-Net design serves as an excellent technique for accurate nanoparticle segmentation in SEM images, providing substantial enhancements compared to traditional approaches. This progress indicates the model's potential for wider applications in nanomaterial research and characterization, where precise and efficient segmentation is essential for analysis.

Keywords—Nanomaterial; segmentation; ResNet 50; modified UNet; transfer learning; SEM

I. INTRODUCTION

In recent years, nanotechnology has emerged as a revolutionary domain with extensive applications across various industries, including healthcare, energy, electronics, and materials research [1]. Fundamental to several innovations is the capacity to generate and evaluate nanoparticles. Nanoparticles, characterized by at least one dimension ranging from 1 to 100 nanometres, demonstrate distinctive physical and chemical properties that differ from those of their bulk equivalents, attributable to their diminutive size and extensive surface area [2]. Nanoparticles have emerged as a fundamental element of contemporary science and technology, attributable to their distinctive qualities stemming from quantum mechanical phenomena, elevated surface-to-volume ratios, and size-dependent characteristics. These features allow

nanoparticles to demonstrate improved optical, mechanical, magnetic, and catalytic characteristics, rendering them essential for various applications.

In medicine, nanoparticles function as drug delivery vehicles, imaging agents for diagnostics, and tools for tissue engineering [3]. Gold nanoparticles are utilized for targeted drug administration and cancer therapy, whereas silver nanoparticles exhibit strong antibacterial traits, rendering them advantageous in medical equipment and wound dressings. Nanoparticles are employed in energy storage to augment the performance of batteries and supercapacitors by improving conductivity and storage capacity. In materials science, nanoparticles are often incorporated into bulk materials to improve strength, flexibility, and thermal characteristics.

Nevertheless, these advances present the issue of precisely describing and assessing nanoparticles, especially regarding their size, shape, distribution, and surface morphology. To conduct a structural characterization of the particles, electron microscopy (EM) is one of the most commonly employed techniques [4]. A particle-interacting electron beam is used in this type of microscope. The objects are reconstructed via the analysis of these interactions and the signals they generate. Transmission electron microscopy (TEM) and SEM are the two primary methods. The primary distinctions between the two technologies are the resolution and the output dimensions. Though TEM images show 2D projections of objects' interior structures, 3D surface reconstruction from SEM images offers important insights into micro/nanoscale surfaces. The magnification and resolution of TEM surpass those of SEM. Consequently, the SEM is typically employed for morphological characterization, whilst the TEM is utilized for assessing particle size and size distribution, along with other analyses such as phase composition and crystal structure [5].

Analyzing SEM images is a typical procedure used to investigate the outcomes of nanomaterial fabrication processes. By identifying and measuring objects, materials scientists can obtain important morphological information about the material of interest. Image processing procedures for SEM imaging were often executed based on the distinctive qualities of a specific material, including shape, size, brightness, and contrast variations among the observed objects [6]. Nonetheless, the interpretation and analysis of SEM images, particularly for extensive datasets, necessitate effective segmentation techniques capable of precisely delineating nanoparticle

boundaries from intricate backgrounds, and extracting characteristics including dimensions, form, and alignment [7].

The results of segmentation directly affect the precision of subsequent operations, such as nanoparticle counting, size distribution assessment, and morphological investigations. Traditionally, nanotechnologists have depended on manual particle measurement utilizing technologies such as ImageJ, which, while successful for small datasets, becomes labor-intensive and susceptible to discrepancies in large-scale image analysis. Conventional techniques, like template matching, edge detection, and feature extraction-based categorization (utilizing Neural Networks or Support Vector Machines), have proven effective but frequently encounter challenges in generalization as image acquisition methodologies vary. These methods necessitate extensive parameter modifying by experts, making them rigid and time-consuming.

Consequently, conventional techniques frequently fail to yield reliable segmentation outcomes, resulting in erroneous nanoparticle analysis. The drawbacks of current techniques underscore the necessity for a more effective, automated approach to nanoparticle segmentation in SEM images. Recent developments in deep learning have demonstrated significant potential in tackling these challenges. Convolutional Neural Networks (CNNs) have transformed image analysis by facilitating automatic feature extraction and end-to-end learning [8]. CNNs are adept in nanoparticle segmentation tasks, as they can incorporate complex, hierarchical characteristics from raw image data, enabling the collection of subtle details such as nanoparticle borders and textures.

CNN-based models, like U-Net, have shown useful in medical image segmentation; nevertheless, their use for nanoparticle segmentation in SEM images is still inadequately investigated. Moreover, conventional U-Net architectures may inadequately leverage deep learning capabilities for nanoparticle segmentation, as they frequently fail to capture multi-scale features and may experience prolonged training durations when utilized with extensive datasets. This study proposes a novel deep learning (DL)-based segmentation approach to tackle these difficulties. The model is based on the U-Net architecture, known for its effectiveness in biomedical segmentation of images, and integrates a ResNet 50 backbone for improved feature extraction. The incorporation of ResNet50, a robust CNN architecture pre-trained on ImageNet, enables the model to utilize deep residual learning, which has demonstrated efficacy in enhancing the training of extremely deep networks by alleviating the vanishing gradient issue. The main contributions of the proposed research are as follows:

- To introduce a modified U-Net architecture with a ResNet50 backbone, specifically designed for the segmentation of TiO₂ nanoparticles in SEM images.
- By leveraging the pre-trained weights of ResNet50, the model aims to enhance feature extraction capabilities while simultaneously reducing training time and computational resources.
- To highlight the superior performance and applicability of the proposed model nanoparticle analysis compared to existing segmentation techniques.

The subsequent sections of the paper are structured as follows: Section II offers a literature review highlighting existing works and identifying research gaps; Section III elaborates on the proposed model; Section IV discusses the results obtained from the study; A discussion is provided in Section V and finally, a summary of the findings is included in Section VI, which gives a conclusion to the paper.

II. LITERATURE REVIEW

Henrik Eliasson et al. (2024) [9] developed a methodology utilizing two U-Net topologies to independently detect and categorize atomic columns at particle-support interfaces in STEM data. This technique sought to alleviate the problem of noisy data caused by the quick scan speeds required for monitoring nanoparticle movement. The U-Net model was trained on simulated non-physical images and was assessed in comparison to established solutions like AtomSegNet and AtomAI, exhibiting superior performance in both in situ and ex situ time series of diminutive Pt nanoparticles on CeO₂. Experimental time series, captured at 5 frames per second, exhibited dynamic, site-specific displacement of atomic columns. Model training and evaluation were performed on an NVIDIA RTX 4090 GPU, necessitating around 40 minutes for localization and three hours for segmentation. The findings highlighted the method's reliability and precision in examining dynamic nanoparticle behavior.

Nina Gumbiowski et al. (2023) [10] performed an investigation of metallic nanoparticles utilizing a machine learning approach that focused on segmentation, the differentiation of overlapping particles, and individual identification. An approach employing ultimate erosion of convex shapes (UECS) was devised to address particle overlap, enabling the assessment of characteristics such as size, circularity, and Feret diameter within a large particle population. The automated analysis of TEM images successfully extracted shape- and size-related data, including that of nanoscale gold nanoparticles. They employed a DL model, namely a deeplabv3+ network with a ResNet-18 backbone, demonstrating that their method sustained performance over diverse contrast levels, surpassing traditional image processing techniques. This automation markedly decreased analysis duration relative to manual techniques and enabled the extraction of extensive data on particle attributes. Nevertheless, constraints remained in the analysis of significantly overlapping particles, highlighting the difficulties in precisely understanding two-dimensional projections in microscopy.

Jonas Bals and Matthias Epple (2023) [11] employed CNN to independently interpret nanoparticle images acquired from scanning electron microscopy. A framework was built for obtaining secondary electron (SE) and STEM images, enabling quantitative studies of particle size and shape. Utilizing pixel weight loss maps to train CNNs enhanced the segmentation of overlapping particles. Their approaches encompassed the classification of forms, including cubes and spheres, the segregation of particles, and the removal of agglomerates. They attained great accuracy in shape classification utilizing AlexNet and ResNet34, together with efficient segmentation employing UNet++. Nevertheless, the system encountered

difficulties with particles that displayed ambiguous forms or partial obscurity, resulting in a considerable misclassification rate. Moreover, challenges associated with particle overlap and intensity fluctuations in SE images further impeded the efficacy of their approach.

Matthew Helmi Leth Larsen et al. (2023) [12] examined the most effective frame dose for object detection and segmentation in low electron dose TEM imaging. The MSD-net architecture exhibited superior performance compared to the regular U-net, particularly at frame dosages lower than those utilized during training. The MSD-net successfully segmented Au nanoparticles, attaining visibility at dosages as low as 20–30 $e^-/\text{\AA}^2$ and full segmentation at 200 $e^-/\text{\AA}^2$. During training with simulated images, the study highlighted how crucial it is to model the modulation transfer function (MTF), hence improving the network's capacity to identify nanoparticles in low signal-to-noise ratio environments. Through benchmarking the U-net and MSD-net across different frame dosages, the authors demonstrated that the MSD-net can generalize beyond its training data, establishing it as the best option for analyzing noisy images.

Wenkai Fu et al. (2022) [13] developed a machine learning model for predicting temporal sequences of in-situ TEM video frames with a hybrid long-short-term memory (LSTM) algorithm and a features de-entanglement technique. They used deep learning algorithms to predict future video frames from prior ones, offering insights into size-dependent structural alterations in Au nanoparticles under dynamic response settings. The models exhibited notable performance, with a structural similarity value of roughly 0.7, while being trained on limited datasets. The researchers accurately forecasted the shifts of Au nanoparticles from rigid to dynamic forms, underscoring the model's relevance in catalytic science. Although the structural similarity scores were inferior to those from more extensive benchmark datasets, the PhyDNet model proficiently delineated the structural evolution of Au nanoparticles.

Zhongyuan Ji and Yuchen Wang (2022) [14] utilized TEM to examine and evaluate the morphological characteristics of nanoparticles. A multirandom forest algorithm for image segmentation was devised, which markedly surpassed conventional techniques like maximum entropy threshold segmentation and watershed segmentation. Utilizing FCM clustering and the algorithm's ability to manage diverse gray levels in TEM images, they attained segmentation accuracy of up to 95%. Notwithstanding these gains, the methodology exhibited certain limitations, such as sensitivity to fluctuations in sample characteristics and image quality, along with computing difficulties in handling extensive datasets.

Annick De Backer et al. (2022) [15] presented a Bayesian genetic approach to reconstruct atomic models of monotype crystalline nanoparticles from single projections utilizing Z-contrast imaging. They employed atom-counting data from annular dark field STEM images as input for preliminary 3D models. The approach reduced structural energy while integrating past knowledge regarding atom-counting accuracy and neighbor-mass relationships. The results indicated enhanced reliability in reconstructing beam-sensitive

nanoparticles, especially those approximately 3 nm in size, at low electron doses. Their comprehensive simulation analysis objectively assessed the reconstructions, demonstrating a substantial improvement in the precise identification of surface atoms. The study demonstrated that the incorporation of finite atom-counting precision and neighbor-mass relationships significantly enhanced the quality of 3D atomic models, indicating improved predictions of catalytic capabilities for future applications. The algorithm was subsequently utilized to evaluate a time series of experimental photographs of a platinum nanoparticle, demonstrating its efficacy in real-time structural measurement.

Hari Mohan Singh et al. (2022) [16] developed a machine learning regression model, Gradient Boost Regression (GBR), to predict the particle size of aluminum nitride (Al_2N_3), silicon nitride (Si_3N_4), and titanium nitride (TiN) nanoparticles dispersed in ethylene glycol (EG) solution. They found critical factors affecting density, including nanoparticle size, molecular weights, volume concentration, and temperature. The GBR model demonstrated significant accuracy in forecasting nanofluid density for a training dataset, nearly matching experimental values obtained from a DMA 500 density meter across different temperatures and concentrations. The research utilized Gradient Search Optimization (GSO) for hyperparameter optimization to improve model performance. The outcomes show a robust correlation between the GBR predictions and experimental data, highlighting the model's efficacy.

Alexey G. Okunev et al. (2020) [17] investigated the automated identification of metal nanoparticles on highly oriented pyrolytic graphite utilizing deep learning methodologies, particularly the Cascade Mask-RCNN neural network. Their model was trained on a dataset including 23 scanning tunnelling microscopy (STM) images, which included 5,157 labeled nanoparticles, and was subsequently validated on a distinct set of images. The trained network achieved a precision of 0.93 and a recall of 0.78, illustrating its proficiency in identifying nanoparticles with an accuracy range of 0.87–0.99 for mean particle size assessments. The study emphasized the limitations of traditional image processing techniques, which required high-quality images and significant parameter adjustment. Researchers have created the open-access web service "ParticlesNN," enabling researchers to analyze noisy STM images without prior improvement, thereby markedly enhancing nanoparticle identification and quantification across diverse imaging scenarios.

Horwath et al. (2019) [18] examined sophisticated deep learning techniques for the segmentation of nanoparticles in EM images. They discovered that although the speed and quality of image segmentation had enhanced, the application of deep learning methods to precisely capture physical attributes continued to pose difficulties. The model's generalization was limited by the necessity for pixel-level annotations and the class imbalance in the training datasets, necessitating meticulous preparation. The effectiveness of segmentation on high-resolution images was further compromised by noise and light fluctuations, requiring more focus during training. Their experiments with various CNN configurations highlighted the importance of batch normalization and kernel size in enhancing

model accuracy and stability. Nevertheless, problems persisted, including sensitivity to fluctuations in image resolution and a tendency to overfit the training data.

Yi-Chi Wang et al. (2019) [19] examined the issues associated with employing STEM tomographic imaging to delineate 3D elemental segregation in nanoparticles, mainly arising from electron dose constraints in conventional techniques. They used a method known as spectroscopic single particle reconstruction (SPR), derived from structural biology, to assess PtNi nanocatalysts. They effectively identified nanoparticles with a diameter of 20 ± 2 nm and a platinum concentration of 56 ± 6 atom% by integrating both STEM-EDS and STEM-HAADF images. The significant diversity in nanoparticle size, shape, and composition required rigorous selection criteria for accurate characterization in the SPR technique.

Although methods for segmenting and detecting nanoparticles in electron microscopy have advanced, there are still a number of significant drawbacks. Numerous current methodologies, including deep learning techniques, encounter difficulties in precisely segmenting overlapping nanoparticles, particularly under fluctuating imaging settings like low signal-to-noise ratios and elevated electron doses. Conventional segmentation techniques often rely on extensive pixel-level annotations and encounter class imbalance, leading to overfitting and insufficient generalization across diverse datasets. Furthermore, traditional algorithms can inadequately consider the complexities of nanoparticle morphology, such as irregular shapes and agglomeration effects, limiting classification accuracy. Dependence on certain imaging modalities sometimes limits these models' ability to include diverse types of electron microscopy data. To address these limitations, we present a novel methodology designed to enhance segmentation accuracy and robustness across various nanoparticle types and imaging conditions, thus facilitating more precise analysis in materials science and nanotechnology applications.

III. MATERIALS AND METHODS

The proposed modified U-Net architecture, incorporating a ResNet50 backbone, was chosen due to its superior feature extraction capabilities, which are critical for accurately segmenting TiO₂ nanoparticles in SEM images. The deep residual learning framework of ResNet50 enables the model to effectively capture intricate textural details and morphological variations, thereby addressing the limitations of conventional segmentation approaches that struggle with complex nanoparticle structures. Furthermore, the integration of ResNet50 with U-Net enhances segmentation accuracy while maintaining computational efficiency, leveraging pre-trained ImageNet weights to expedite training and improve generalization across diverse SEM datasets. In contrast, traditional U-Net models, while effective for biomedical image segmentation, lack the depth required for precise nanoparticle boundary detection, often resulting in suboptimal segmentation performance. Moreover, standard CNN-based approaches necessitate extensive manual feature engineering and fail to adequately handle the high variability in SEM image textures. High-complexity architectures, such as DeepLabV3+, although

capable of achieving high accuracy, impose significant computational overhead, making them less practical for real-time applications. The proposed modified U-Net architecture effectively mitigates these challenges, providing a robust, scalable, and precise segmentation framework that significantly outperforms existing methodologies in the domain of nanomaterial characterization.

The approach seeks to improve the segmentation of nanoparticles in high-resolution SEM images by incorporating transfer learning and a modified U-Net (mUNet) architecture. A pre-trained ResNet50 model is employed, originally constructed on the ImageNet dataset, to extract significant features from SEM images depicting various nanoparticle shapes, distributions, and sizes. Each image in the dataset is paired with a ground truth mask that delineates the nanoparticles, enabling supervised learning. Pre-processing techniques, such as image scaling and normalization, enable the standardization of inputs for the model. The ResNet50 backbone serves as a feature extractor, acquiring multi-scale, intricate representations of the SEM images, which the mUNet model then refines for accurate segmentation. The efficiency of the model is evaluated following training by accuracy, precision, recall, Intersection over Union (IoU), and Dice Coefficient to confirm its ability to separate nanoparticles under diverse imaging settings. By overcoming the issues related to SEM images, this method aims to increase the precision and efficacy of nanoparticle segmentation. Fig. 1 displays the suggested framework's block diagram.

A. Dataset

The dataset for the proposed study is obtained from the publicly accessible repository on GitHub (<https://github.com/BAMresearch/automatic-sem-image-segmentation>). The database includes EM micrographs of TiO₂ particles and their related segmentation masks that define the boundaries of these particles. The collection additionally encompasses classifications of the particles according to their visibility and occlusion. The set is organized into subfolders that include raw SEM and TSEM images, along with manually annotated segmentation and classification masks. This extensive dataset establishes a robust basis for training and assessing the mUNet model, ensuring the precision and dependability of the segmentation procedure through the utilization of high-quality, real-world SEM data. Fig. 2 shows the first and last sample SEM images from the dataset along their corresponding manually annotated segmentation mask.

B. Data pre-processing

Pre-processing techniques such as scaling and normalization are employed to normalize the SEM images for uniform input into the neural network. Resizing standardizes images to a consistent dimension, whereas normalization calibrates pixel values to meet the network's specifications, hence improving the model's data processing efficiency. Standardizing the inputs enhances the model's robustness and its capacity to manage changes in nanoparticle morphology within SEM images. The dataset was divided into two categories, with 85% allocated for model training and the remaining 15% allotted for validation. This curated dataset is

essential for training and assessing the mUNet model, enabling precise and resilient segmentation performance.

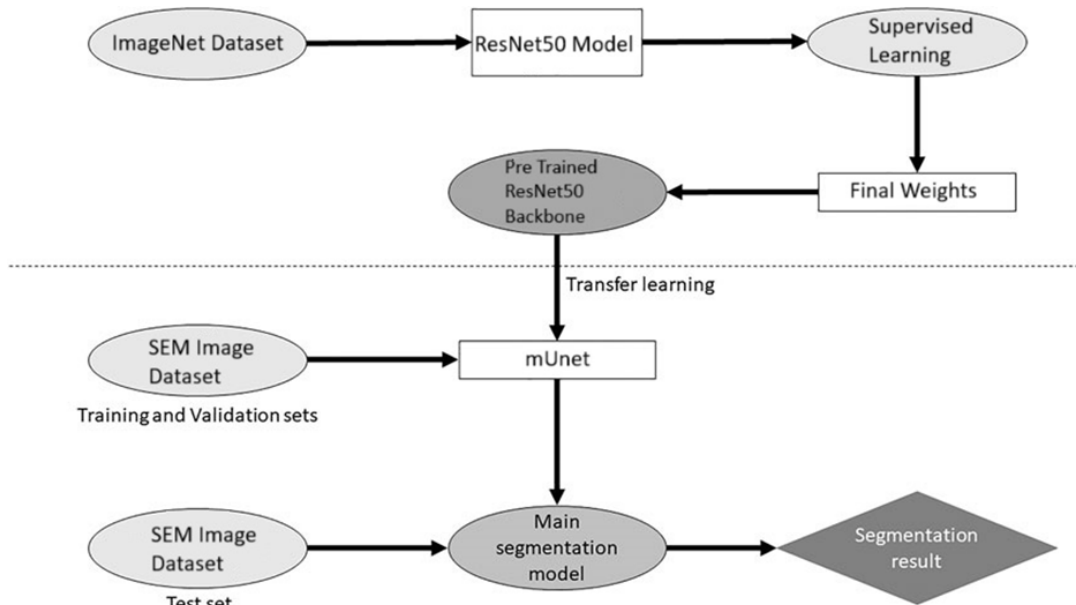


Fig. 1. Block diagram of the proposed model.

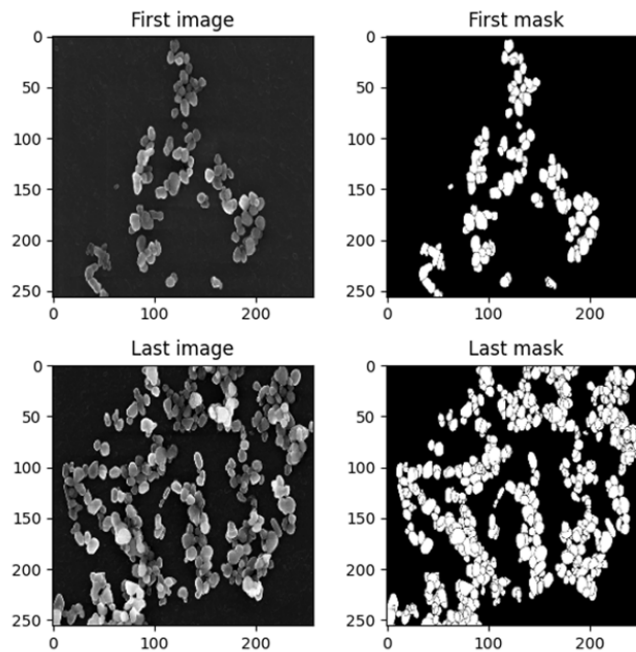


Fig. 2. Sample images from the dataset.

C. Model Development

Segmentation is an essential procedure in image analysis that entails dividing the images into separate sections, facilitating the identification and localization of particular objects within the image. This research utilizes deep learning techniques to improve segmentation accuracy, specifically for the identification of nanoparticles in SEM images. The U-Net design, acclaimed for its efficacy in biomedical image segmentation, underpins the suggested methodology. A mUNet employs a pre-trained ResNet50 encoder to enhance feature

extraction and gather multi-scale information. This combination enables the accurate delineation of nanoparticle borders, enhancing both the precision and efficiency of the segmentation process.

1) *U-Net*: The U-Net is a widely employed CNN model, particularly adept at segmentation tasks, such as nanoparticle segmentation in high-resolution SEM images. UNet is engineered for pixel-level classification, facilitating the differentiation of regions of interest (ROI) from the background in images, which is particularly beneficial in

domains like medical imaging, biological analysis, and materials science [20]. The architecture is characterized by its U-shaped structure, with two main pathways: a contracting pathway and an expansive pathway, as shown in Fig. 3. The

contracting path derives feature maps from the input image using many convolutional layers, whereas the expansive path reconstructs these features into a segmented output by integrating low-level and high-level features.

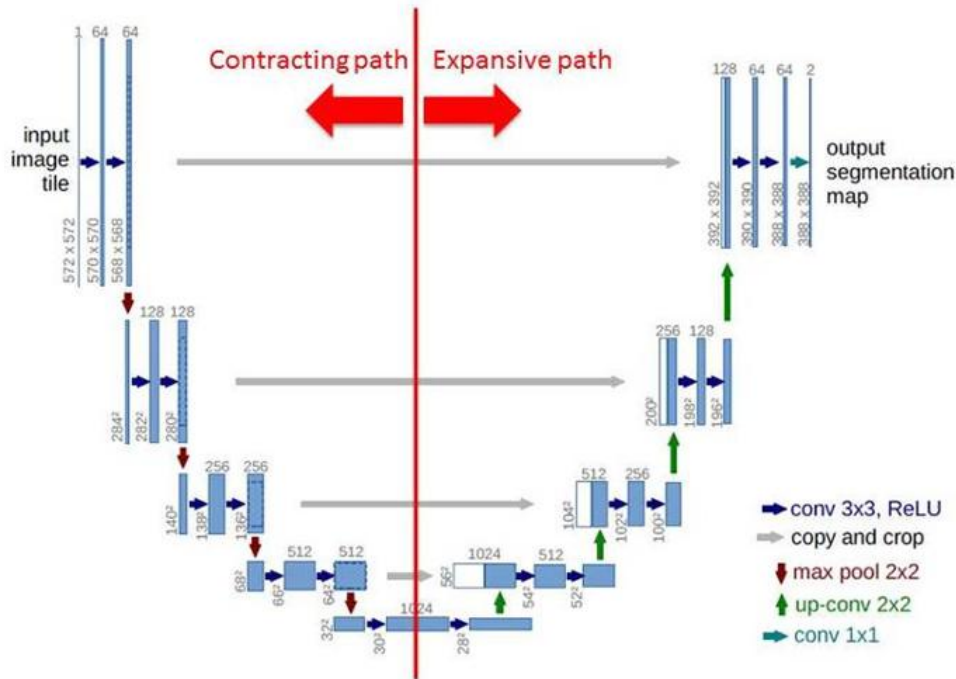


Fig. 3. Basic UNet architecture.

The contracting path, referred to as the encoder, adheres to the conventional architecture of a CNN. The process entails consecutive applications of two 3x3 convolutional layers, succeeded by a ReLU activation function and a 2x2 max-pooling operation. The max-pooling procedure aims to down sample the image, decreasing its spatial dimensions while simultaneously doubling the number of feature channels at each iteration. The convolutional procedure is mathematically expressed as shown in Eq. (1).

$$f(x) = \sigma(W * x + b) \quad (1)$$

where, W denotes the convolutional weights, x signifies the input image, b indicates the bias, and σ represents the activation function.

The expansive path, known as the decoder, mimics the contracting path in reverse, with the objective of reconstructing the image's spatial dimensions while preserving localization accuracy. At each phase of the expanding pathway, the feature maps are subjected to up sampling, generally via a 2x2 transpose convolution process, to revert the image to its original resolution. The transpose convolution is mathematically represented in Eq. (2).

$$\hat{x} = W^T * f(x) \quad (2)$$

Where, W^T denotes the transposed weights utilized in the up-sampling process, while x signifies the upsampled feature map. This step is succeeded by concatenation with the relevant feature maps from the contracting path, enabling the network to merge low-resolution, high-context data with high-

resolution, low-context features. The concatenation technique preserves small information from prior layers, which is essential for precise nanoparticle segmentation.

Skip connections are crucial to the performance of the U-Net [21]. These connections directly associate feature maps from the contracting path with the expanded path, as shown in Fig. 4, enabling the model to preserve intricate spatial information that could be lost during downsampling. By integrating low-level, high-resolution characteristics with up sampled high-level features, the network can generate precise and detailed segmentations.

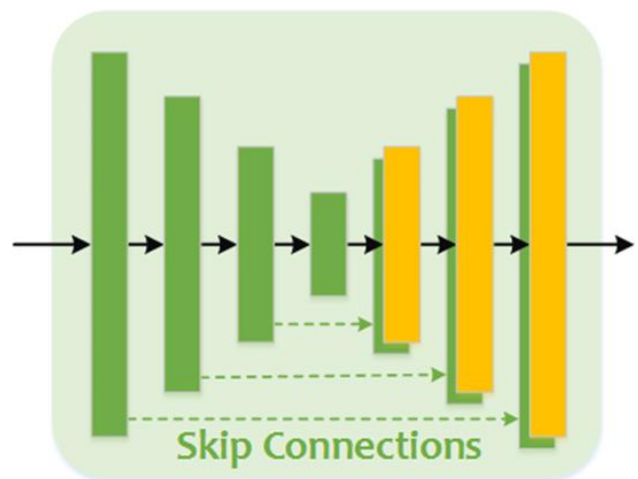


Fig. 4. Skip connections in the UNet architecture.

In the final layer, a 1×1 convolution is utilized to transform the 64-channel feature map into the requisite number of output classes (e.g., foreground and background in binary segmentation), mathematically represented as in Eq. (3). This generates the definitive segmentation mask, categorizing each pixel based on its respective region.

$$y = W_{final} * f(x) + b_{final} \quad (3)$$

Where W_{final} represents the final layer's weights, and b_{final} is the bias term.

A distinctive feature of the U-Net architecture is its capacity to manage huge images via an overlap-tile method. Due to GPU memory limitations on image size, U-Net divides huge images into smaller tiles, processes each tile independently, and then merges them to generate a comprehensive segmentation map. To mitigate the loss of context at the image peripheries, the input image is mirrored throughout the tiling procedure. This technique guarantees precise segmentation of edge pixels, even when analyzing extensive SEM images of nanoparticles.

2) *Proposed modified UNet architecture:* The proposed research enhances the traditional U-Net design by integrating a modified U-Net structure that employs a ResNet50 encoder. This modification seeks to utilize the powerful feature extraction skills of ResNet50 to enhance nanoparticle segmentation performance from SEM images. The modification concentrates on optimizing feature extraction, refining multi-scale representations, and improving the overall accuracy of segmentation results.

ResNet50 is a CNN that uses residual learning to improve the training of deep networks, facilitating the effective acquisition of intricate hierarchical features vital for precise

segmentation tasks. ResNet50 comprises 50 layers organized into many blocks, as shown in Fig. 5, using skip connections, enabling the model to learn residual functions [22]. This architecture mitigates the degradation issue frequently observed in deep networks, wherein greater depth results in reduced performance. ResNet50 mitigates vanishing gradient problems by establishing direct paths for gradients during back propagation, hence enhancing the training of deeper networks. Each residual block in ResNet50 comprises two or three convolutional layers, batch normalization, and ReLU activation algorithms, ending in a shortcut connection that bypasses one or more layers.

In a standard residual block, the input X undergoes a sequence of convolutions, after which the original input is reintegrated into the output, resulting in Eq. (4).

$$Y = F(X) + X \quad (4)$$

Where $F(X)$ denotes the function acquired by the convolutional layers, and Y signifies the output of the block. This architecture enhances the acquisition of identity mappings, hence aiding the training of more profound networks.

The ResNet50 encoder comprises numerous convolutional layers arranged in blocks. Each block generally consists of three convolutional layers: the initial layer applies a 1×1 convolution to diminish dimensionality, the subsequent layer employs a 3×3 convolution for feature extraction, and the last layer utilizes another 1×1 convolution to reinstate the original dimensionality. This setup improves the model's ability to represent spatial hierarchies while preserving computational efficiency. The implementation of batch normalization subsequent to each convolutional layer enhances learning stability by normalizing the inputs to each layer.

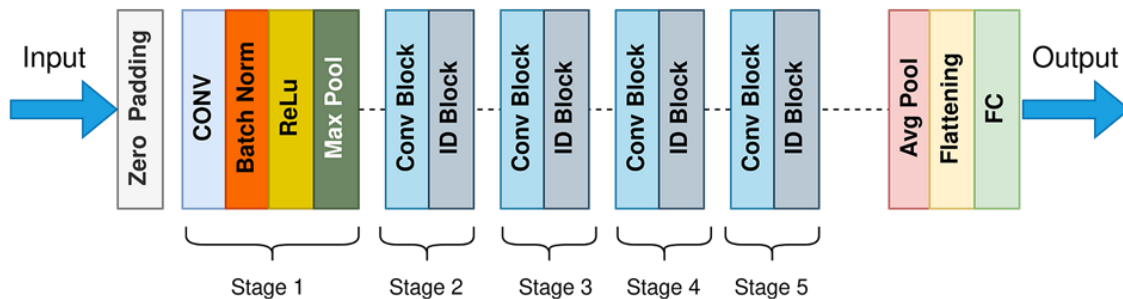


Fig. 5. ResNet 50 architecture.

The downsampling is achieved by strided convolutions and max pooling layers. Strided convolutions diminish the spatial dimensions of feature maps while augmenting depth, efficiently capturing multi-scale characteristics at diverse levels of abstraction. The downsampling process is essential for the encoder, enabling the model to concentrate on more intricate, abstract representations of the input data as it advances through the layers. As the input SEM images traverse the ResNet50 encoder, feature maps are produced at various depths. These feature maps encompass a comprehensive array of attributes that capture both low-level features (like textures and edges) and high-level semantic data (such as patterns and shapes). The hierarchical structure of feature extraction allows the model to

acquire intricate representations essential for precisely segmenting nanoparticles in the images. The formula for feature extraction at any specified layer is represented as shown in Eq. (5).

$$F_i = \sigma(W_i \cdot X + b_i) \quad (5)$$

Where F_i denotes the feature map at layer i , W_i and b_i represents the weights and biases of the convolutional layer, whereas X signifies the input feature map from the preceding layer.

The deepest convolutional layers process the smallest and most abstract feature mappings near the architecture's

bottleneck. This component encapsulates the most complicated illustrations of the input data, enabling the network to proficiently discern complex patterns related to nanoparticles. The architectural design enables the model to manage high-level properties while preserving essential spatial information required for precise segmentation.

During the decoder step, the design utilizes transpose convolutions (deconvolutions) to up sample the feature maps to their original input dimensions. The mUNet employs a concatenation method that integrates the up sampled feature maps with the matching encoder feature maps at different levels, rather than simply copying the feature maps. This procedure ensures the preservation and integration of intricate details and spatial data from preceding layers with enhanced semantic attributes. The concatenation is represented mathematically as in Eq. (6).

$$F_{concat} = F_{upsampled} \oplus F_{encoder} \quad (6)$$

Where F_{concat} denotes the concatenated feature map, $F_{upsampled}$ refers to the feature map subsequent to upsampling, and $F_{encoder}$ signifies the feature map derived from the encoder.

The output layer employs a 1×1 convolution to diminish the feature map's channel number to one, appropriate for binary segmentation applications. A sigmoid activation function is utilized on the output to produce a segmentation mask, resulting in values ranging from 0 to 1, which represent the probability of each pixel being part of the target class (nanoparticles). The final segmentation mask is represented as shown in Eq. (7).

$$M = \sigma(W_{output} \cdot F_{concat} + b_{output}) \quad (7)$$

Where M represents the output mask, and W_{output} and b_{output} denote the weights and bias for the output layer.

Post-processing techniques, including thresholding and morphological processes, are utilized to enhance the raw segmentation outcome. The proposed research achieves enhanced segmentation performance due to the creative utilization of convolutional layers, downsampling techniques, and residual connections, which facilitate an in-depth knowledge of the input data.

D. Hardware and Software Setup

The study utilized a high-performance computational configuration comprising an Intel Core i7 CPU, 32GB of RAM, and an NVIDIA GeForce GTX 1080Ti GPU, facilitating the effective management of demanding computational workloads. The framework was executed with the Keras library, a high-level neural network API based on TensorFlow, recognized for its user-friendly interface and robust functionalities. The training procedure was conducted on Google Colab, a cloud-based Python notebook platform that offers easy accessibility to substantial computational resources, hence facilitating model training.

An essential element of this research was the selection of hyperparameters, which profoundly influence model performance during training. Unlike model parameters that are

derived from the data hyperparameters are predetermined by the user and are crucial in shaping the configuration of the training process to optimize the performance of the nanoparticle segmentation model. The precise hyperparameter selections and model configuration are detailed in Table I.

TABLE I. HYPERPARAMETER SPECIFICATIONS

Hyper parameters	Values
Epochs	200
Learning Rate	0.0001
Optimizer	ADAM
Batch size	4
Loss function	Dice loss

IV. EXPERIMENTAL RESULTS

Initially, several factors are defined to quantify essential performance parameters, as represented in following Eq. (8), Eq. (9), Eq. (10), Eq. (11), and Eq. (13). These metrics, based in the principles of False Positive (FP), True Negative (TN), False Negative (FN), and True Positive (TP), are crucial for evaluating the efficacy of the model.

The calculation of accuracy involves dividing the total number of predictions by the number of right predictions.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

The exactness of a prediction is measured by its precision, or the number of true positives. Instead, recall quantifies completeness, or the number of real positives that were anticipated as positives.

$$Precision = \frac{TP}{TP+FP} \quad (9)$$

$$Recall = \frac{TP}{TP+FN} \quad (10)$$

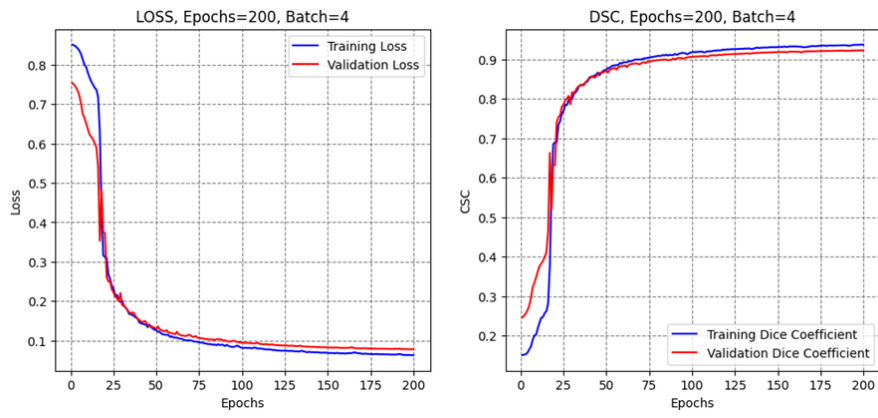
$$F1 - Score = 2 * \left(\frac{Precision * Recall}{Precision + Recall} \right) \quad (11)$$

$$Dice Score = \frac{2 * A \cap B}{|A| + |B|} \quad (12)$$

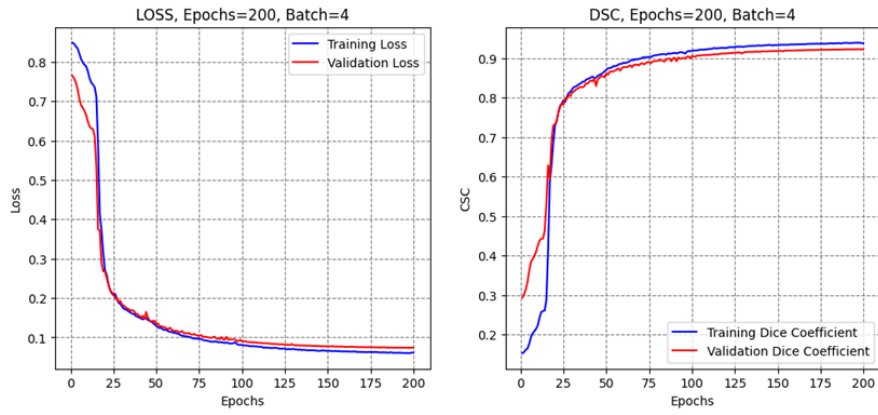
$$IoU = \frac{|A \cap B|}{A \cup B} \quad (13)$$

Where, A and B denote the set of predicted and actual positive instances.

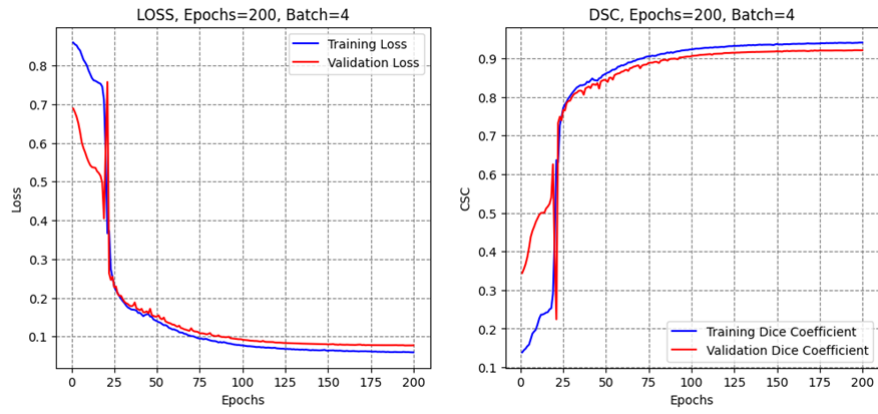
The dataset is partitioned into five folds using KFold, which randomly assigns data to training and validation sets for each iteration. For each fold, the model is established and constructed with five iterations of 200 epochs each. The training procedure includes a 15% validation split, with EarlyStopping callbacks to avert overfitting by ceasing training when the validation loss does not increase over a predetermined number of epochs. Following each training session, the model's learning progression is depicted via the training and validation loss curves, as shown in Fig. 6. Metrics from each iteration are recorded for subsequent analysis, facilitating a thorough assessment of the model's performance across several folds.



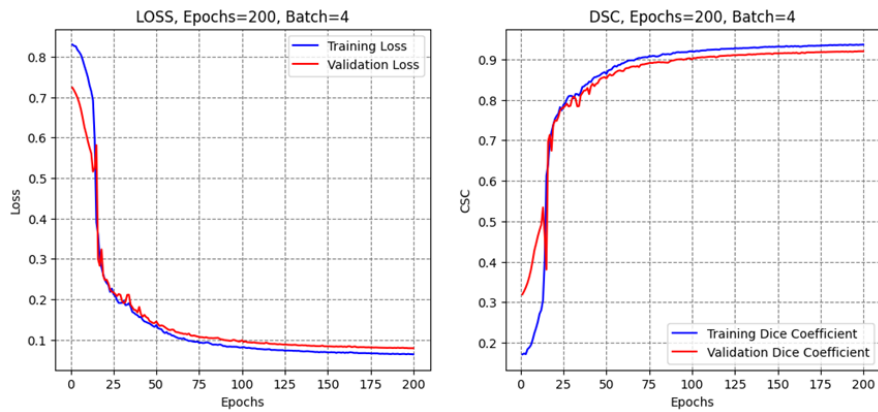
(a) Run 1



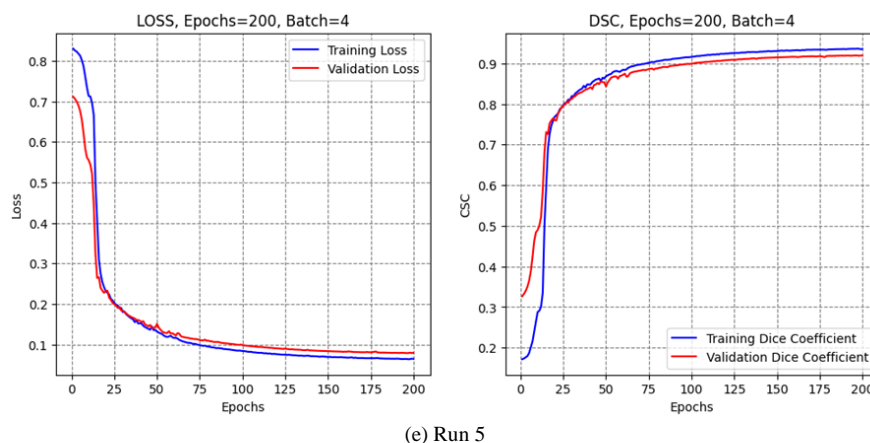
(b) Run 2



(c) Run 3



(d) Run 4



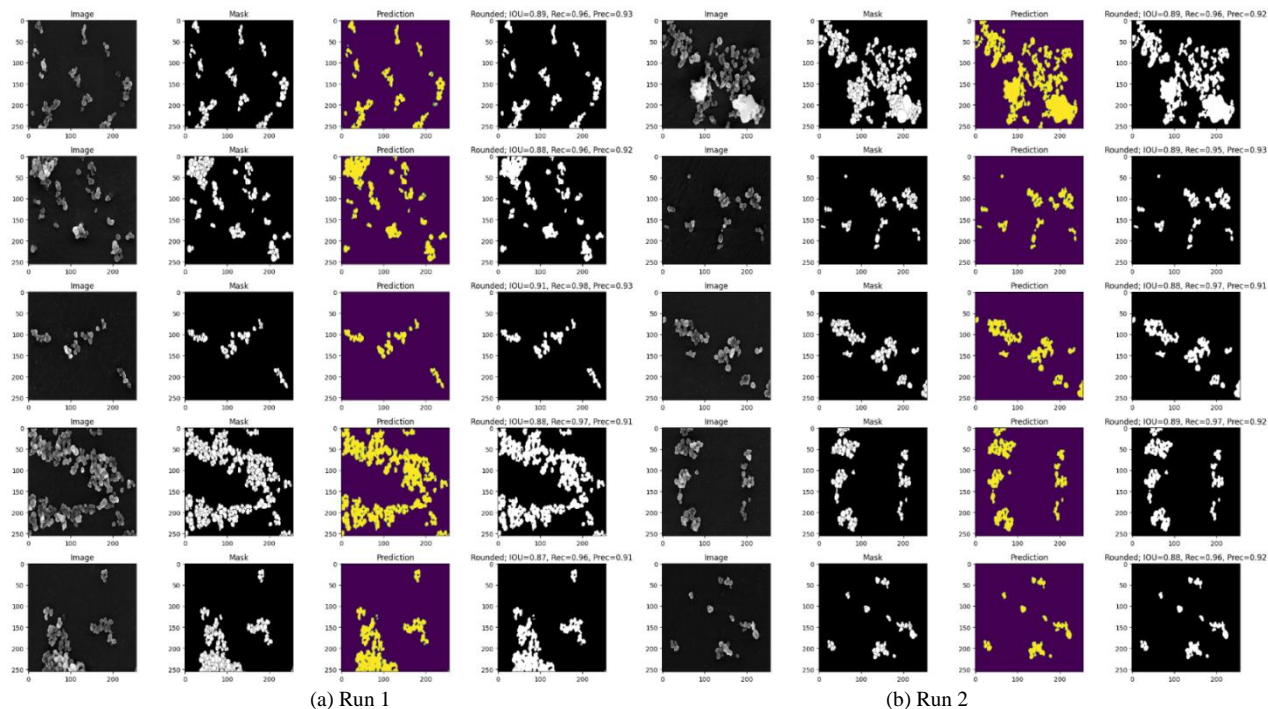
(e) Run 5
Fig. 6. Training and validation dice coefficient of the proposed model.

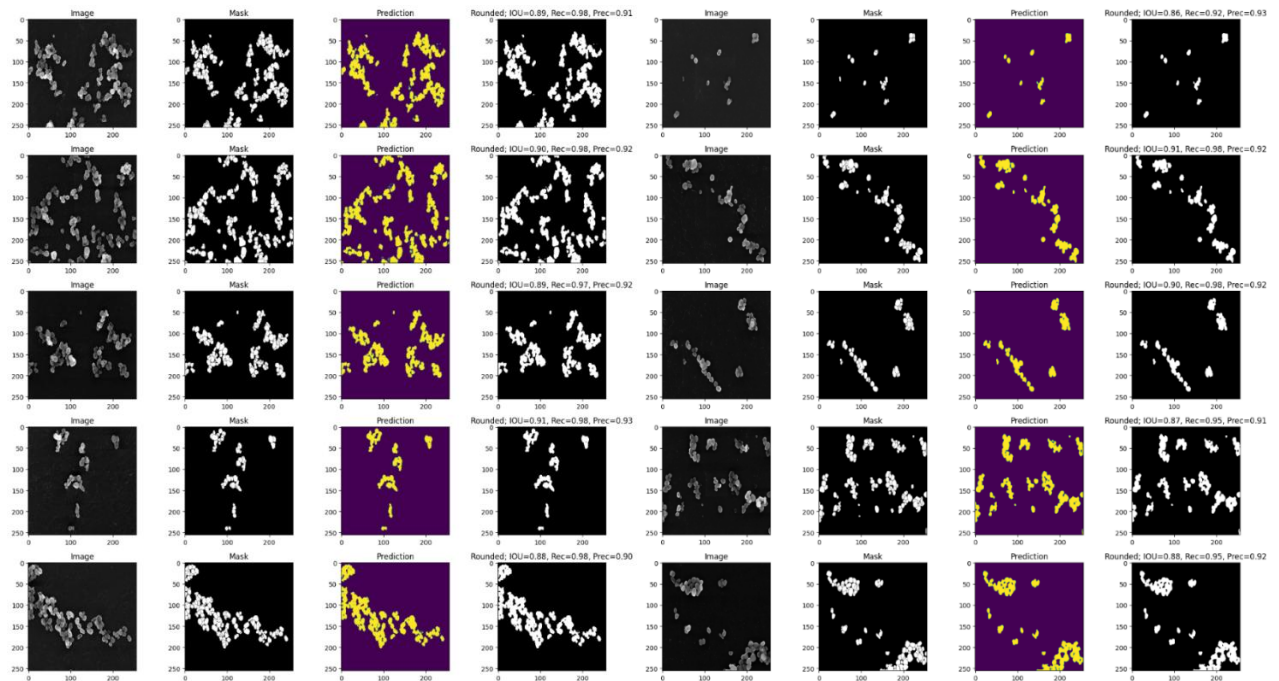
The incorporation of visualizations that compare predicted masks with original images, as illustrated in Fig. 7, improves interpretability and facilitates the evaluation of segmentation quality.

The model's precision metric exhibits a gradual enhancement across the runs, as shown in Fig. 8 (a) and Table II, commencing at 0.907 in the initial run and culminating at 0.926 in the final run. This pattern indicates the model's enhanced capacity to accurately identify true positive nanoparticle boundaries. Notwithstanding a slight decline to 0.900 in the third run, overall precision enhanced, especially in the fourth run, where it attained 0.916. The recall scores consistently stayed elevated during the runs, as shown in Fig. 8 (b), starting at 0.974, decreasing slightly to 0.969 in the second run, and achieving a maximum of 0.979 in the third run. The most recent test registered a decrease to 0.966; however, the

recall metrics continually demonstrated the model's capability in detecting nanoparticle segments.

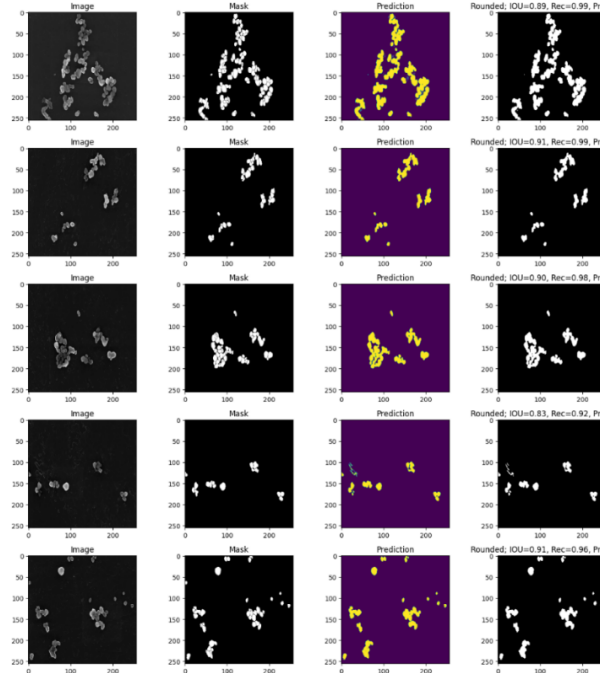
The IoU score demonstrated robust performance, as shown in Fig. 8 (c), commencing at 0.885, with minor fluctuations during the second and third runs, and ultimately increasing to 0.897 in the last run. This enhancement indicates improved precision in delineating nanoparticle boundaries. The DSC commenced at 0.938, as shown in Fig. 8 (d), and reached a peak of 0.946 in the third run, indicating the highest level of segmentation accuracy attained. A decrease in DSC was noted in consecutive trials, culminating in a final value of 0.940. The mUNet model demonstrates robust segmentation capabilities, with significant performance enhancements across multiple parameters, confirming its efficacy in precisely recognizing and characterizing nanoparticle boundaries.





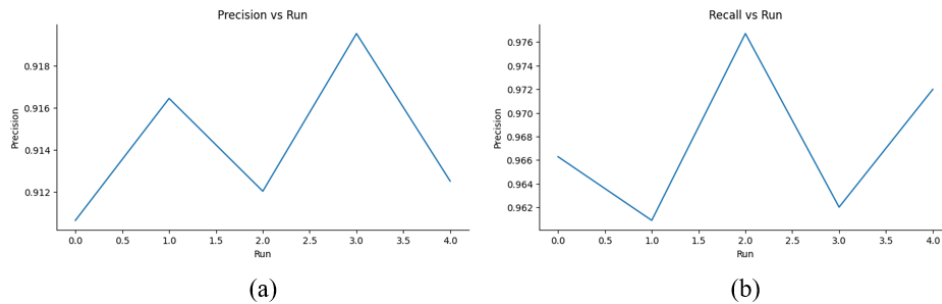
(c) Run 3

(d) Run 4



(e) Run 5

Fig. 7. Prediction outputs.



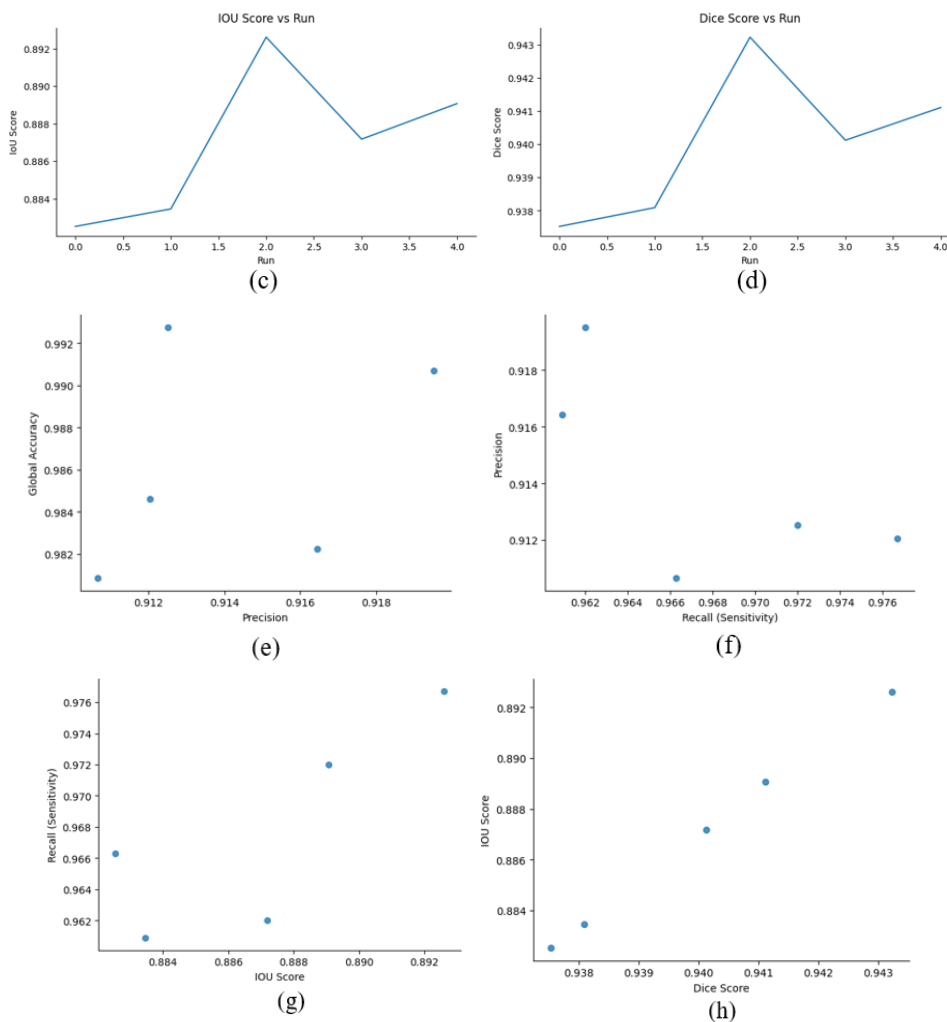


Fig. 8. Visualization of performance evaluation.

TABLE II. PERFORMANCE EVALUATION

Metric	Run 1	Run 2	Run 3	Run 4	Run 5
Dice Score (DSC)	0.939	0.938	0.938	0.943	0.946
Intersection over Union (IoU)	0.885	0.883	0.883	0.892	0.897
Recall	0.974	0.969	0.979	0.972	0.966
Precision	0.907	0.909	0.900	0.916	0.926
Global Accuracy	0.981	0.982	0.983	0.991	0.993
AUC ROC	0.990	0.992	0.988	0.994	0.994

Fig. 9 displays histograms that depict the performance measures of the mUNet model. The precision scores, between 0.912 and 0.916, demonstrate the model's robust ability to effectively separate nanoparticles with minimal false positives. Despite a slight variation in these numbers, the general stability demonstrates the model's efficacy in positive predictive accuracy. The recall scores vary from 0.962 to 0.970, indicating the model's exceptional capability in accurately identifying almost all true positive cases of nanoparticles. The

IoU scores, ranging from 0.884 to 0.892, indicate strong performance, reflecting a significant overlap between expected and actual areas. The uniformity of these scores highlights the model's dependability in sustaining high-quality segmentation across samples. Finally, the Dice scores vary from 0.938 to 0.943, underscoring the model's proficiency in generating precise segmentations. The close proximity of these scores signifies negligible performance variability, hence affirming the model's dependability.

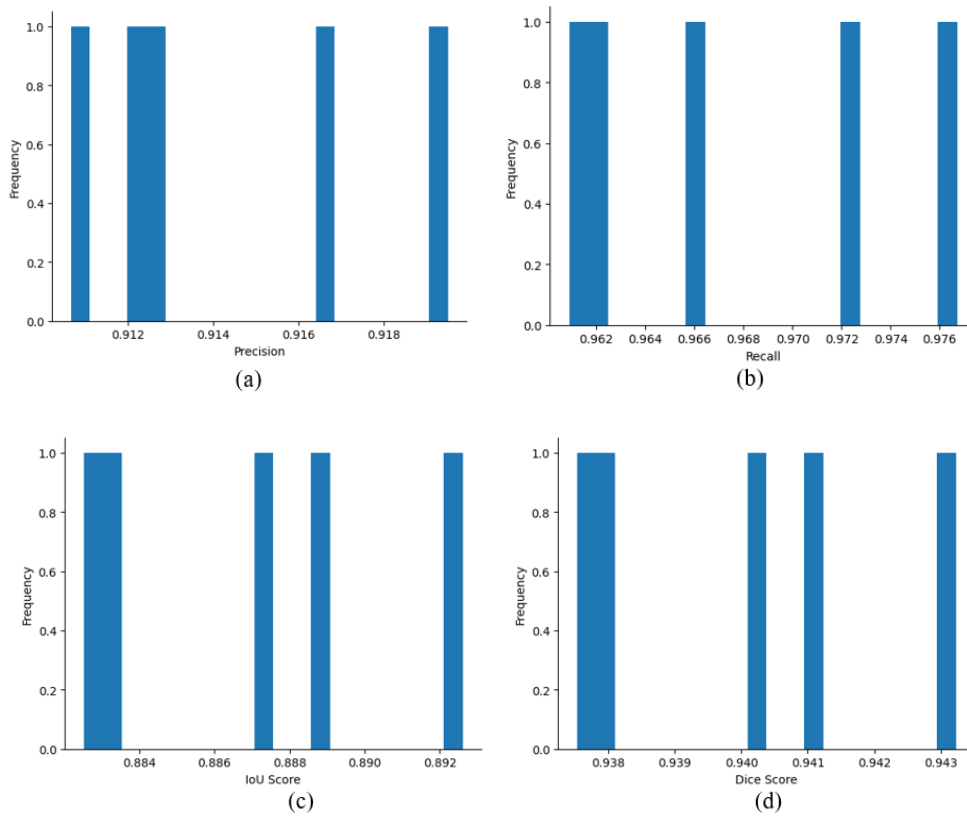


Fig. 9. Histogram plots.

Fig. 10 illustrates the mean values of the primary performance metrics acquired over an epoch of 200. The Dice Score Mean is 0.940, indicating a high degree of accuracy in the overlap between the anticipated and actual nanoparticle areas. The mean IoU score is 0.887, demonstrating strong efficacy in accurately defining nanoparticle regions with few deviations. The mean results together underscore the dependability of the mUNet model in nanoparticle segmentation tasks, demonstrating consistent performance across various parameters.

Fig. 11 depicts the standard deviations computed across 200 epochs. The Dice score demonstrates a minimal standard deviation of 0.002071, indicating consistent segmentation accuracy over repetitions. The IoU score demonstrates a standard deviation of 0.003696, signifying reliable overlap between predicted and real nanoparticle regions. The recall has slightly larger variance, with a standard deviation of 0.006005, indicating modest fluctuations in the model's capacity to recognize all actual nanoparticle occurrences.

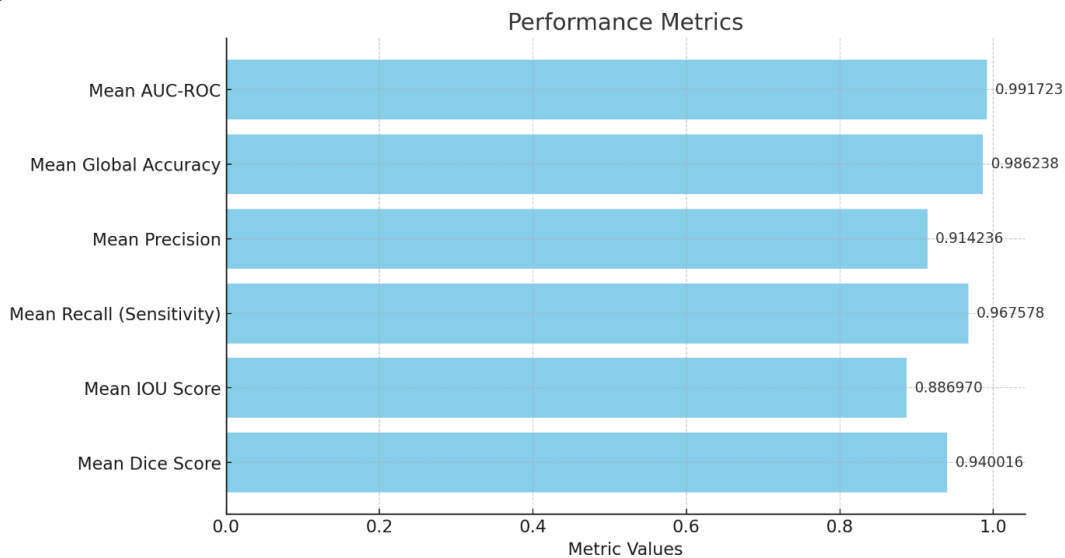


Fig. 10. Mean value of the scores.

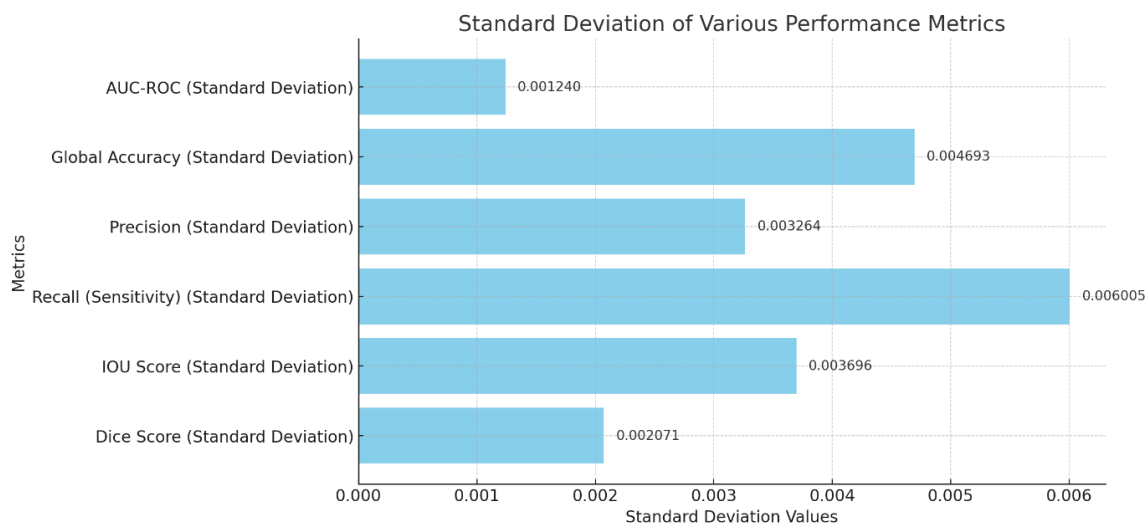


Fig. 11. Standard deviation of the scores.

The precision exhibits a standard deviation of 0.003264, signifying dependable identification of true positives. The standard deviation of global accuracy is 0.004693, indicating uniform performance across all iterations, whilst the AUC-ROC score exhibits the lowest deviation at 0.00124, implying remarkable stability in differentiating between nanoparticle and non-nanoparticle regions. The low standard deviations across all measures indicate that the mUNet model exhibits consistent and dependable performance, with minor variations in its ability to effectively segment TiO₂ nanoparticles throughout numerous iterations.

V. DISCUSSION

A thorough accuracy comparison of the suggested model against a number of cutting-edge segmentation methods is depicted in Fig. 12 and Table III. The suggested framework attained an outstanding accuracy of 99.3%, markedly surpassing conventional approaches such as NSNet (86.2%) and more sophisticated architectures like Deeplabv3+ with ResNet-18 (96.12%), multiple-output convolutional neural networks (96.59%), and U-Net (97.1%).

TABLE III. COMPARATIVE ANALYSIS OF THE PROPOSED MODEL AGAINST EXISTING METHODS

Author	Methodology	Accuracy (%)
Sun et al. [25]	NSNet	86.2
Gumbiowski et al. [10]	Deeplabv3+ network with a Resnet-18	96.12
Oktay et al. [24]	multiple output convolutional neural networks	96.59
Bals et al. [11]	UNET and UNet++	97
Leonid Mill et al. [23]	U-Net	97.1
Proposed model: Modified U-Net with ResNet 50		99.3

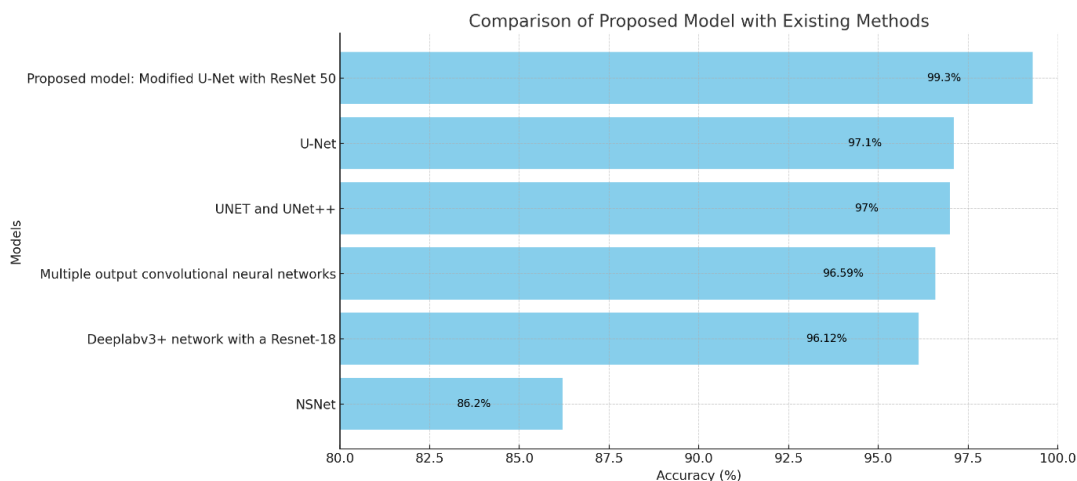


Fig. 12. Comparison of the proposed model with existing methods.

The suggested model exhibits enhanced segmentation performance compared to closely comparable models such as UNet++ (97%) and U-Net (97.1%). The enhancement is due to the incorporation of ResNet50 as the encoder, facilitating superior feature extraction via deep residual learning, along with the U-Net's strong skip connections, which ensure accurate reconstruction of spatial information. The modifications, together with the model's capacity to utilize pre-trained weights for more effective training, lead to a significant improvement in segmentation accuracy for nanoparticle boundaries in SEM images. The modified layout exhibits enhanced proficiency in managing complex textures and diverse sizes of nanoparticle areas, positioning itself as a dependable and effective solution for high-precision segmentation tasks in nanotechnology.

VI. CONCLUSION

The proposed research offers an extensive approach for automating the segmentation of SEM images of TiO₂ nanoparticles utilizing a modified U-Net architecture with a ResNet50 backbone pre-trained on ImageNet. This model effectively overcomes the drawbacks of conventional segmentation techniques, which frequently encounter issues with the intricate and diverse textures present in SEM images. By utilizing the robust feature extraction capabilities of ResNet50 and integrating them with the effective segmentation framework of U-Net, the model exhibits notable enhancements in accuracy, precision, and overall performance. The findings, featuring an average Dice score of 0.940 and an IoU of 0.887, demonstrate the model's proficiency in precisely delineating nanoparticle boundaries. The minimal standard deviations in all performance metrics, such as a Dice score standard deviation of 0.002071 and an IoU standard deviation of 0.003696, underscore the model's consistency and stability throughout various iterations. The incorporation of skip connections and multi-scale feature learning allows the model to preserve spatial details while analyzing abstract, high-level data. This method substantially surpasses conventional strategies in nanoparticle segmentation for accuracy and reliability, as evidenced by comparisons with existing methods. The model's versatility and diminished training duration underscore its practical utility for high-throughput nanoparticle investigation in materials science. The research highlights the efficacy of deep learning models, particularly the mUNet, in enhancing nanomaterial analysis through accurate, automated segmentation methods.

REFERENCES

- [1] Rokunuzzaman, M. K. (2024). The Nanotech Revolution: Advancements in Materials and Medical Science. *Journal of Advancements in Material Engineering*, 9(2), 1-10.
- [2] Fernandez-Garcia, M., Martinez-Arias, A., Hanson, J. C., & Rodriguez, J. A. (2004). Nanostructured oxides in chemistry: characterization and properties. *Chemical reviews*, 104(9), 4063-4104.
- [3] Shariati, L., Esmaili, Y., Rahimmanesh, I., Babolmorad, S., Ziaei, G., Hasan, A., ... & Makvandi, P. (2023). Nanobased platform advances in cardiovascular diseases: Early diagnosis, imaging, treatment, and tissue engineering. *Environmental Research*, 116933.
- [4] Su, D. (2017). Advanced electron microscopy characterization of nanomaterials for catalysis. *Green Energy & Environment*, 2(2), 70-83.
- [5] Inkson, B. J. (2016). Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) for materials characterization. In *Materials characterization using nondestructive evaluation (NDE) methods* (pp. 17-43). Woodhead publishing.
- [6] Mohale, G. T. M., Beukes, J. P., Kleynhans, E. L. J., Van Zyl, P. G., Bunt, J. R., Tiedt, L. R., ... & Jordaan, A. (2017). SEM image processing as an alternative method to determine chromite pre-reduction. *Journal of the Southern African Institute of Mining and Metallurgy*, 117(11), 1045-1052.
- [7] Yao, L., & Chen, Q. (2023). Machine learning in nanomaterial electron microscopy data analysis. In *Intelligent Nanotechnology* (pp. 279-305). Elsevier.
- [8] Yao, X., Wang, X., Wang, S. H., & Zhang, Y. D. (2022). A comprehensive survey on convolutional neural network in medical image analysis. *Multimedia Tools and Applications*, 81(29), 41361-41405.
- [9] Eliasson, H., & Erni, R. (2024). Localization and segmentation of atomic columns in supported nanoparticles for fast scanning transmission electron microscopy. *npj Computational Materials*, 10(1), 168.
- [10] Gumbiowski, N., Loza, K., Heggen, M., & Epple, M. (2023). Automated analysis of transmission electron micrographs of metallic nanoparticles by machine learning. *Nanoscale advances*, 5(8), 2318-2326.
- [11] Bals, J., & Epple, M. (2023). Deep learning for automated size and shape analysis of nanoparticles in scanning electron microscopy. *RSC advances*, 13(5), 2795-2802.
- [12] Larsen, M. H. L., Lomholdt, W. B., Valencia, C. N., Hansen, T. W., & Schiøtz, J. (2023). Quantifying noise limitations of neural network segmentations in high-resolution transmission electron microscopy. *Ultramicroscopy*, 253, 113803.
- [13] Fu, W., Spurgeon, S. R., Wang, C., Shao, Y., Wang, W., & Peles, A. (2022). Deep-learning-based prediction of nanoparticle phase transitions during in situ transmission electron microscopy. *arXiv preprint arXiv:2205.11407*.
- [14] Ji, Z., & Wang, Y. (2022). Application of Multiple Random Forest Algorithm in Image Segmentation of Nanoparticles. *Journal of Nanomaterials*, 2022(1), 4964368.
- [15] De Backer, A., Van Aert, S., Faes, C., Arslan Irmak, E., Nellist, P. D., & Jones, L. (2022). Experimental reconstructions of 3D atomic structures from electron microscopy images using a Bayesian genetic algorithm. *npj Computational Materials*, 8(1), 216.
- [16] Singh, H. M., Sharma, D. P., & Alade, I. O. (2022). GBR-GSO based machine learning predictive model for estimating density of Al₂N₃, Si₃N₄, and TiN nanoparticles suspended in ethylene glycol nanofluids. *The European Physical Journal Plus*, 137(5), 587.
- [17] Okunev, A. G., Mashukov, M. Y., Nartova, A. V., & Matveev, A. V. (2020). Nanoparticle recognition on scanning probe microscopy images using computer vision and deep learning. *Nanomaterials*, 10(7), 1285.
- [18] Horwath, J. P., Zakharov, D. N., Megret, R., & Stach, E. A. (2019). Understanding important features of deep learning models for transmission electron microscopy image segmentation. *arXiv preprint arXiv:1912.06077*.
- [19] Wang, Y. C., Slater, T. J., Leteba, G. M., Roseman, A. M., Race, C. P., Young, N. P., ... & Haigh, S. J. (2019). Imaging three-dimensional elemental inhomogeneity in Pt-Ni nanoparticles using spectroscopic single particle reconstruction. *Nano Letters*, 19(2), 732-738.
- [20] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer International Publishing.
- [21] Wang, H., Cao, P., Wang, J., & Zaiane, O. R. (2022, June). Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 36, No. 3, pp. 2441-2449).
- [22] Prabakaran, J., & Selvaraj, P. (2022, December). Implementation of ResNet-50 with the Skip Connection Principle in Transfer Learning Models for Lung Disease Prediction. In *International Conference on*

- Intelligent Systems and Sustainable Computing (pp. 9-19). Singapore: Springer Nature Singapore.
- [23] Mill, L., Wolff, D., Gerrits, N., Philipp, P., Kling, L., Vollnhals, F., ... & Christiansen, S. (2021). Synthetic image rendering solves annotation problem in deep learning nanoparticle segmentation. *Small Methods*, 5(7), 2100223.
- [24] Oktay, A. B., & Gurses, A. (2019). Automatic detection, localization and segmentation of nano-particles with deep learning in microscopy images. *Micron*, 120, 113-119.
- [25] Sun, Z., Shi, J., Wang, J., Jiang, M., Wang, Z., Bai, X., & Wang, X. (2022). A deep learning-based framework for automatic analysis of the nanoparticle morphology in SEM/TEM images. *Nanoscale*, 14(30), 10761-10772.

Application of Big Data Mining System Integrating Spectral Clustering Algorithm and Apache Spark Framework

Yuansheng Guo

China Mobile Communications Group, Hunan Co. Ltd, Changsha, Hunan, 410001, China

Abstract—Spectral clustering algorithm is a highly effective clustering algorithm with broad application prospects in data mining. To improve the efficient data processing capability of big data mining systems, a big data mining system that integrates spectral clustering algorithm and Apache Spark framework is proposed. It is applied in the big data mining system by combining Hadoop, Spark framework, and spectral clustering algorithm. The research results indicated that after 300 iterations of spectral clustering algorithm, the error value tended to stabilize and drops to 0.123. In different datasets, different error values were displayed, indicating that spectral clustering algorithm had better performance in discrete data processing and smaller testing errors. The minimum time consumed by the comparative system was 37.83 seconds, the maximum time was 55.26 seconds, and the average time was 51.65 seconds. The minimum time consumed by the research system was 18.93 seconds, the maximum time consumed was 32.22 seconds, and the average time consumed was 28.14 seconds. Compared with the comparative system, the research system consumed less time, trained faster, and was more conducive to shortening the clustering running time. The algorithm framework and system raised in the research have good operational efficiency and clustering ability in data mining processing, which promotes the reliability and development of big data mining systems.

Keywords—Spectral clustering algorithm; apache spark; big data; data mining

I. INTRODUCTION

The advent of the big data era has led to a proliferation of big data mining technology across a range of industries. Big data technology takes a critical parts in multiple fields with its massive data information and high-intensity processing capabilities. It not only enables efficient analysis of complex data modules, but also has foresight and predictability, and can extract valuable data in a timely manner [1]. Data mining technology, as an emerging discipline, originated in the 1980s with the initial aim of promoting the development of artificial intelligence technology. Modern data mining technologies focus on in-depth exploration of hidden and valuable data to discover new data patterns and valuable information, which has critical guiding significance for enterprise decision-making. Spark, as a big data processing framework, has the merits of high efficiency, scalability, and high fault tolerance, and is therefore broadly utilized in the field of big data mining [2]. This study will explore big data mining techniques from the perspective of Spark. Spectral Clustering (SC), as a classic data mining algorithm, is a clustering algorithm used in graph theory.

It achieves node clustering by analyzing the eigenvalues and eigenvectors of the Laplacian matrix of the graph. Many experts and researchers have put forward their own opinions on the research and implementation of big data systems. SC is an unattended clustering algorithm that has been broadly applied in the fields of pattern matching and computer vision due to its excellent clustering capabilities. However, the conventional SC algorithms are ill-suited for large-scale data classification such as that required for hyperspectral remote sensing images. This is due to their high computational complexity and the difficulty of representing the inherent uncertainty of the images [3]. Li et al. employed fuzzy anchor points for the processing of hyperspectral image classification and proposed an SC algorithm based on fuzzy similarity measurement. The findings of the experiment on the datasets of hyperspectral remote sensing images demonstrated the efficacy of the enhanced algorithm. The incorporation of a fuzzy likelihood measure led to the generation of a more resilient similarity matrix. The kappa coefficient obtained by the raised algorithm was 2% higher than that of the traditional algorithm. Furthermore, the raised algorithm achieved superior classification results on hyperspectral remote sensing images when compared with existing methods [4]. The advent of wireless communication technology has led to the generation of a substantial corpus of spatio-temporal user tracking data, which is recorded by wireless communication networks as users utilize these networks to meet a range of needs. To enhance the healthy development of students and facilitate the construction of campus-wide information, Guo Y et al. put forth an SC algorithm based on a multi-level threshold and density combined with common nearest neighbors. Several clustering algorithms were used for detecting anomalies, and four assessment indicators were applied to assess the clustering results. The results indicated that the MSTDSNN-C algorithm exhibited better clustering performance [5]. However, the fact that the clustering model is defined only for the original data and not explicitly extended to out-of-sample data is one of the main drawbacks of SC. To improve its efficiency, Shen D et al. proposed a new modular SC method with out of sample extension, combining a new spectral mapping algorithm based on modular similarity measurement and out of sample extension. The experiment outcomes denoted that the research method had better findings compared to other related algorithms on several data sets [6]. A block distributed Chebyshev-Davidson algorithm was developed by Pang Q et al. to solve the problem of large leading eigenvalues in SC. Through the analysis of the Laplacian matrix or normalized

Laplacian matrix in SC, a scalable distributed parallel version was developed. The results demonstrated its efficiency in SC and its advantage in scalability compared to existing feature solvers used for SC in parallel computing environments [7].

Most existing multi-view clustering methods may be affected by data corruption in terms of technology, leading to a sharp decline in clustering performance. Pan Y et al. put forth a multi-pattern SC method which uses robust bar space segmentation. To address the optimization issue of the weak sparse segmentation, an optimization procedure based on the extended Lagrangian multiplier method was developed. The experiment findings on various benchmark sets showed that the raised method performed well relative to several recent advances in clustering methods [8]. High utility itemset mining is a common utilized data mining method for finding useful patterns. Sethi K et al. proposed a new way to mine itemsets using Spark. They tested it on six real data sets and found that it outperformed other algorithms [9]. When managing very large datasets, the high processing cost of mining data for fuzzy rules increased considerably, and in many cases memory overrun faults are triggered. Fernandez-Basso C et al. used the Spark algorithm to process large amounts of heterogeneous data and find interesting rules. They proposed a measure of interest decomposition based on Alpha cuts and demonstrated through experiments that only 10 equidistant Alpha cuts were sufficient to find all the important fuzzy rules. The efficiency and speed of all proposals were compared and analyzed [10]. Ji L et al. proposed an improved SC-based method of detecting anomalies for anomalous data mining in dam safety monitoring, which introduced natural eigenvalues to select data point edges based on traditional SC. The results showed that this method could avoid the algorithm from becoming bogged down in local topology and improve the efficiency of clustering and anomaly detection. It further confirmed that the method could adjust itself well to the case of discrete distribution datasets, and was more accurate than classical SC methods in both the case of labeling and detecting the data points with unusual anomalies [11].

In summary, regarding data mining, existing researchers in the literature review have some involvement and research on data processing, algorithm classification, and dataset clustering improvement. However, the design and application of clustering algorithms for implementing system data mining are not deep enough, such as data relationship description, architecture design of data processing systems, etc. In order to achieve more efficient and large-scale data processing efficiency, a big data mining method that combines spectral clustering algorithm and Apache Spark framework is proposed compared with literature review. It combines distributed computing framework (such as Spark) to optimize spectral clustering algorithm, realizing parallel processing and fast clustering of large-scale datasets. This is similar to the distributed block Chebyshev Davidson algorithm developed by Pang Q et al. And innovatively introduced spectral clustering algorithm applied to data mining systems, designed a big data mining system architecture, and provided a technical foundation for massive data mining and processing.

The article structure of this study is as follows. Introduction is given in Section I. Section II of this study is dedicated to the

integration of the SC algorithm with the Apache Spark framework for the purpose of facilitating the mining of large data sets. This represents a significant area of focus and innovation within the field of big data analytics. Section III presents the experimental verification and analysis of the results obtained from the data set, based on the algorithm designed in the first part. Section IV presents conclusions regarding the experimental results and discusses the limitations of the design, as well as avenues for future research.

II. METHODS AND MATERIALS

The study adopts spectral clustering algorithm as the core clustering method, which can identify sample spaces of any shape and converge to the global optimal solution, especially suitable for clustering convex structured data. And by constructing a similarity matrix, calculating eigenvalues and eigenvectors, and using classical clustering algorithms such as K-means to cluster the eigenvectors, data clustering analysis is achieved. Firstly, this study combines Hadoop and Apache Spark to investigate the processing techniques of big data. Secondly, the SC algorithm is introduced and combined with the Apache Spark framework to design a framework for a big data mining system.

A. Big Data Technology based on Hadoop and Apache Spark Computing Framework

As the advent of the digital age, big data has become a fundamental element for enterprises to compete. Apache Spark has gained widespread attention in terms of processing speed, fault tolerance, and ease of use. Apache Spark is a high-performance, flexible computing engine that is optimized for processing large datasets. Compared to the traditional big data processing framework MapReduce, Spark has a faster processing speed. This is because Spark stores data in memory instead of traditional storage on disk. Another feature of Spark is that it can perform iterative calculations based on memory. Hadoop Distributed File System (HDFS) can work well on inexpensive hardware and is designed to be fault-tolerant. It provides high throughput for accessing application data and enables fast access to large datasets [12]. Hadoop is a distributed system built by the Apache Foundation, and the HDFS is one of its components [13]. The big data ecosystem of Hadoop is shown in Fig. 1.

HDFS is capable of accessing data in the file system in the form of streams, and the fundamental design of this framework is based on HDFS and MapReduce. HDFS provides storage for substantial quantities of data, while MapReduce offers computational capabilities for similarly large data sets [14]. The MapReduce feature of Hadoop can decompose a large and complex task, allocate scattered subtasks to multiple nodes, and then load them as a single dataset into a data warehouse. The distributed architecture of Hadoop enables the big data processing engine to be situated as proximate to the storage facility as possible. This makes the system relatively suitable for batch operations such as ETL, given that the results of such operations may be transmitted directly from the processing engine to storage. The popularity of Hadoop in the area of big data processing can be attributed to its efficacy in data extraction, transformation, and loading.

Spark is an open-source project under the Apache foundation that provides a distributed computing framework for fast processing of large-scale datasets [15]. Compared to traditional MapReduce, Spark uses memory storage to read and write data faster, avoiding frequent disk I/O operations and improving data processing speed. Spark supports multiple programming languages, such as Scala, Java, Python, and R, making it easy for users to choose their familiar programming language for development. It also provides a resilient distributed dataset (RDD), as shown in Fig. 2 for its structure and running process.

Fig. 2 (a) showcases the structure of the RDD dataset, and Fig. 2 (b) showcases the operational flowchart of RDD. RDD is composed of multiple partitions, each of which is a subset of data that can be distributed across multiple machines for parallel computing. Partitioning is the process of grouping data records with the same attributes together according to specific rules, where each partition is equivalent to a segment of the dataset. This partitioning mechanism enables RDD to support parallel processing and improve computational efficiency.

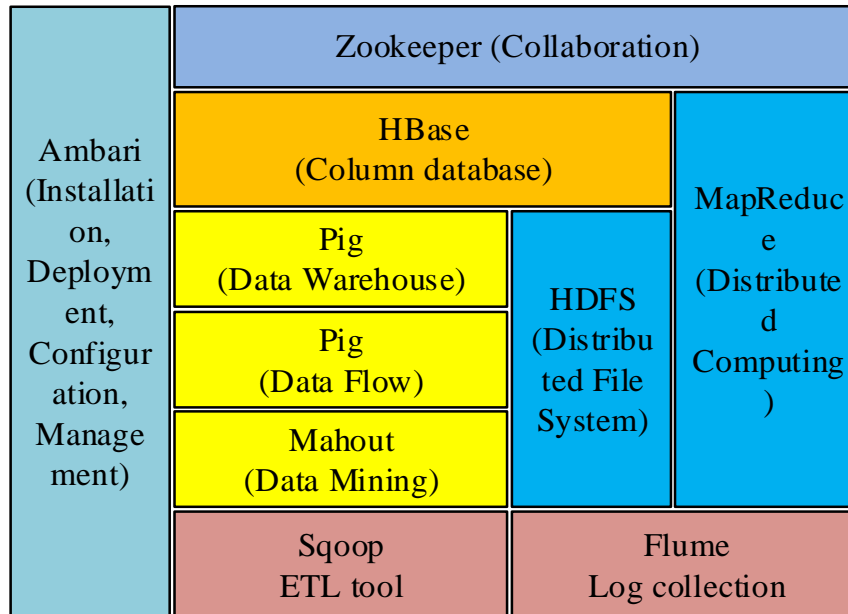


Fig. 1. Hadoop big data ecosystem.

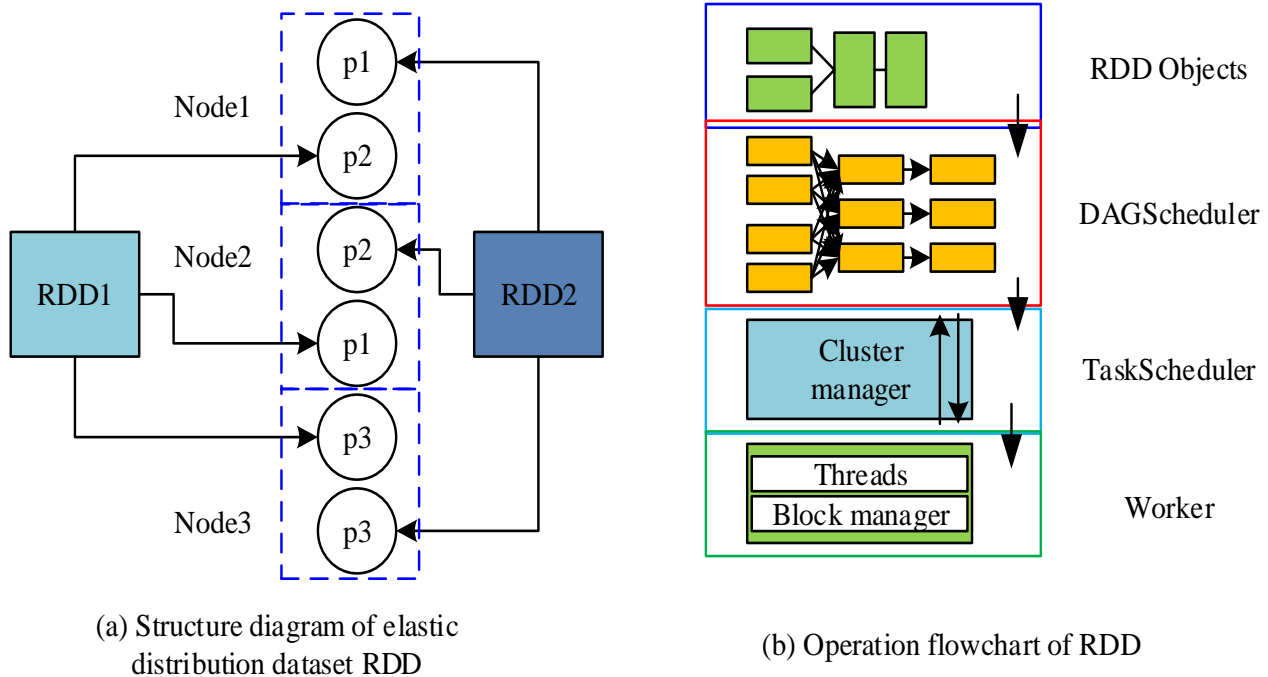


Fig. 2. Structure diagram and operation flowchart of RDD dataset.

The running process of RDD in Spark architecture mainly includes the following steps. Firstly, it is necessary to create an RDD object. Secondly, the dependency relationships between RDDs are calculated and a Directed Acyclic Graph (DAG) is constructed. SparkContext is responsible for calculating the dependency relationships between RDDs and building the DAG. DAG represents the structure of the entire computing task, including the conversion and computation between various RDDs. Then the DAG is decomposed into multiple stages, and the DAGScheduler is responsible for decomposing the DAG graph into multiple stages, each stage containing multiple tasks [16]. The tasks in each stage are executed in order of their dependency relationships to ensure the correctness of the calculation results. Afterwards, each task will be distributed by the task scheduler to the Executors on each work node for execution. After receiving the task, the Executor

will occupy corresponding resources such as CPU and memory and perform calculations. The calculation results will be returned to the Driver for summarization and processing. Finally, there is the summary and output of the results. After all tasks are completed, the Driver will collect all the results, perform necessary summarization and processing, and finally output the results. This can be done by pulling all data back to the driver end using the collect () method.

This process involves the core mechanisms of Spark's distributed computing framework, including resource allocation, task scheduling and execution, as well as result aggregation and output. In this way, Spark can efficiently process large-scale datasets, achieve parallel and distributed computing, and the running process is shown in Fig. 3.

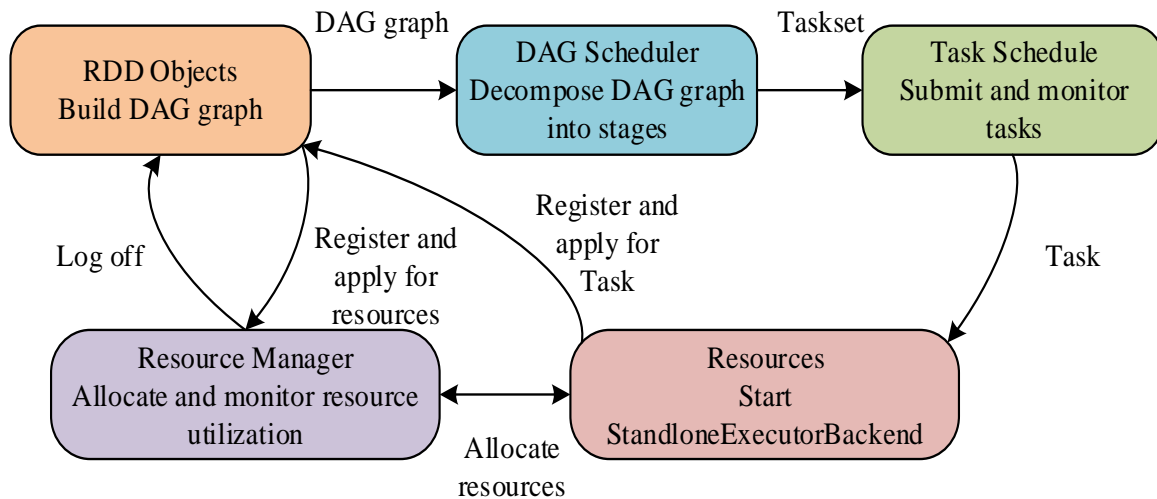


Fig. 3. Spark running process.

The running process of Spark involves environment construction, resource allocation, task decomposition and scheduling, as well as specific behaviors in different running modes, ensuring efficient execution of distributed computing. Firstly, the DAG graph created in the RDD object is decomposed into stages, and Task Schedule is formed through Taskset to submit and monitor tasks.

B. A Big Data Mining System Integrating Spectral Clustering Algorithm and Apache Spark Framework

To achieve efficient mining and analysis of big data, a high-performance SC algorithm is adopted in the study, which can provide better clustering for convex structured data. SC is a clustering method based on graph theory that divides a weighted undirected graph into two or more optimal subgraphs. This is achieved by ensuring that the subgraphs are as similar as possible internally while maximizing the distance between subgraphs [17]. The underlying principle of the SC method is the transformation of the initial clustering problem into an optimal graph partitioning problem. The selection of appropriate eigenvectors for clustering is achieved by calculating the eigenvalues and eigenvectors of the similarity matrix of the sample data points. This method is capable of identifying sample spaces of any shape and converging upon the global optimal solution [18]. The implementation process

of SC includes constructing a similarity matrix, calculating eigenvalues and eigenvectors, and using K-means or other classical clustering algorithms to cluster eigenvectors. The SC algorithm has a wide range of applications, including computer vision, pattern recognition, information retrieval, and other fields. Spectral clustering algorithm treats all data as points in space during the clustering process. By slicing the graph composed of all data points, the edge weights between different subgraphs are minimized, while the edge weights within subgraphs are maximized, thus achieving the purpose of clustering. This method overcomes the disadvantage of traditional clustering algorithms (such as K-Means) that may not be able to obtain the global optimal solution on any shaped sample space.

The study will use a directed unweighted graph to represent the dataset $G = (V, E)$, and describe its relationships using a matrix to transform it into a graph/matrix problem. The similarity of data points will be described using functions, and the relationship equation will be constructed as shown in Eq. (1).

$$w_{i,j} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (1)$$

In Eq. (1), $w_{i,j}$ denotes the similarity between x_i and x_j corresponding to the i row and j column, and the dataset is represented as $\{v_1, v_2, \dots, v_n\}$. x_i and x_j are the data points. A matrix is constructed with a size of $n * n$ based on the relationships between data points. A set matrix that represents the sum of similarity relationships between data points and other points through a degree matrix, as shown in Eq. (2).

$$\begin{cases} d_i = \sum_{j=1}^n w_{ij} \\ D = \begin{pmatrix} d_1 & & & \\ & d_2 & & \\ & & \dots & \\ & & & d_n \end{pmatrix} \end{cases} \quad (2)$$

In Eq. (2), D denotes the degree matrix, and d_i represents the degree of data point x_i . In this study, the similarity matrix is constructed using fully connected connections, and a Gaussian kernel function is utilized to construct the similarity distance, as shown in Eq. (3).

$$W_{ij} = S_{ij} = \left(-\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \quad (3)$$

In SC algorithms, graph problems involve partitioning problems. From the perspective of graph theory, clustering problems are equivalent to partitioning problems of a graph. The similarity between subgraphs is described by dividing them into different subgraphs. the partitioning principles include minimum cut criterion, normative cut criterion, and proportional cut criterion. The objective of partitioning is to reduce the sum of edge weights that are removed, as a smaller sum of edge weights results in a greater dissimilarity between the subgraphs connected by them, and therefore a greater distance between them. Subgraphs with low similarity can be easily cut off from them [19].

The Laplacian matrix is an important component of SC algorithms and is a matrix used to represent a graph. Given a graph $G = (V, E)$ with n vertices, the vertex set V represents each sample, and the weighted edge E represents the similarity between each sample. The non normalized Laplacian matrix is represented by Eq. (4).

$$L = D - W \quad (4)$$

The properties of the non normalized Laplacian matrix are shown in Eq. (5).

$$f^T L f = \frac{1}{2} \sum_{i,j=1}^n w_{ij} (f_i - f_j)^2 \quad (5)$$

In Eq. (5), $f = (f_1, f_2, \dots, f_n)^T, f \in R^n$ is an arbitrary vector. The normalized Laplacian matrix can be divided into two forms: symmetric and random walk normalized matrices, as shown in Eq. (6).

$$\begin{cases} L_{sym} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} W D^{-1/2} \\ L_{rw} = D^{-1} L = I - D^{-1} W \end{cases} \quad (6)$$

In Eq. (6), L_{sym} represents the symmetric normalization matrix, L_{rw} represents the normalization moment of random walks, W represents the adjacency matrix, I represents the identity matrix, and L represents the non normalized Laplacian matrix. The properties of the normalized Laplacian matrix are shown in Eq. (7).

$$f^T L_{sym} f = \frac{1}{2} \sum_{i,j=1}^n w_{ij} \left(\frac{f_i}{\sqrt{d_i}} - \frac{f_j}{\sqrt{d_j}} \right)^2 \quad (7)$$

In Eq. (7), d_i and d_j represent the element values of the matrix. The acquisition of SC algorithm requires the partitioning of the graph, transforming discrete problems into continuous problems. The SC algorithm's acquiring process is shown in Fig. 4.

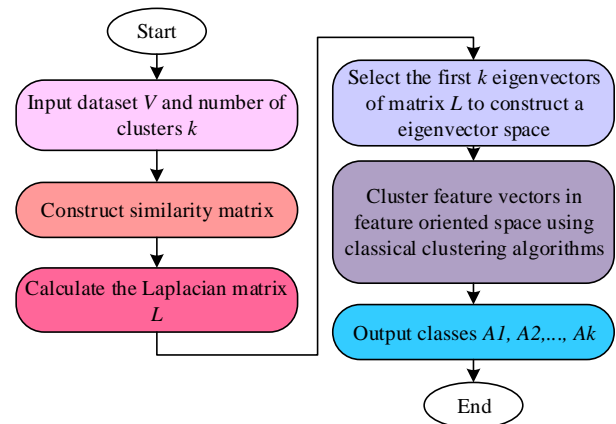


Fig. 4. Spectral clustering algorithm process.

The process mainly includes the following steps. Firstly, it will calculate the similarity between given datasets, and select an appropriate similarity calculation method based on the characteristics of the datasets to build a similarity matrix. On the basis of the similarity matrix, a Laplacian matrix is constructed through regularization processing. The Laplacian matrix can be constructed in two ways: diagonal matrix and adjacency matrix. The eigenvalue decomposition is performed on the Laplacian matrix to obtain a series of eigenvalues and corresponding eigenvectors. The corresponding eigenvectors are selected based on the first K smallest eigenvalues, which form a low dimensional space, and project the original dataset into this low dimensional space [20-21]. The clustering analysis is performed on the projected dataset using the K-means algorithm to get the final clustering results. In addition, to assess the efficacy of SC algorithms, this study uses algorithm time complexity. Firstly, a dataset of n with each data dimension d is set up to construct a corresponding similarity map. After calculating the time complexity, the eigenvalues and eigenvectors of the similarity matrix are calculated. Finally, the corresponding eigenvectors are obtained through dimensionality reduction for clustering. The calculated time map is shown in Fig. 5.

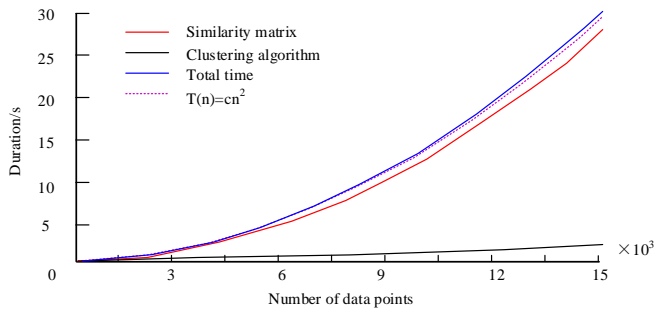


Fig. 5. Calculation time chart of spectral clustering.

Fig. 5 shows the time required for each step of spectral clustering in the dataset. The total time of the algorithm is basically consistent with the fitting function $f(n)=cn^2$, so the total time complexity of the algorithm is $O(n^2)$, and the construction of the similarity matrix stage consumes the most time. This study is based on SC algorithm and Apache Spark framework to design a big data mining system. The system is broken into three layers of architecture, each layer has interface connections, and from bottom to top are the data layer, business layer, and interaction layer. The data layer accesses files in the data system during the homework process to perform read and save operations on data in the database. The main function of the interaction layer is to display data and receive and transmit user data, providing an interactive operating interface for the website's system operation. The business layer identifies and processes user input information, saves it separately, establishes a new data storage method, reads the data during the storage process, and saves the business logic description code. The system architecture is shown in Fig. 6.

The research first uses the Hadoop and Apache Spark computing frameworks for data processing, and utilizes the distributed computing capabilities of the Apache Spark framework to allocate the computing tasks of the spectral clustering algorithm to multiple nodes for parallel execution, thereby improving the efficiency of the algorithm. By utilizing Spark's RDD (Elastic Distributed Dataset) mechanism, distributed storage and parallel processing of data can be achieved, reducing disk I/O operations and accelerating data processing speed.

This study combines spectral clustering algorithm with Apache Spark framework, which not only optimizes the computational efficiency of spectral clustering algorithm, but also enhances the ability of big data processing. This technological fusion provides new ideas and methods for

research in related fields, promoting innovation and development of algorithm technology. By utilizing the distributed computing capabilities of the Apache Spark framework, this study achieved efficient processing and analysis of large-scale datasets. This helps to address the limitations of traditional big data processing techniques in terms of processing speed and fault tolerance, providing strong support for the further development and application of big data technology.

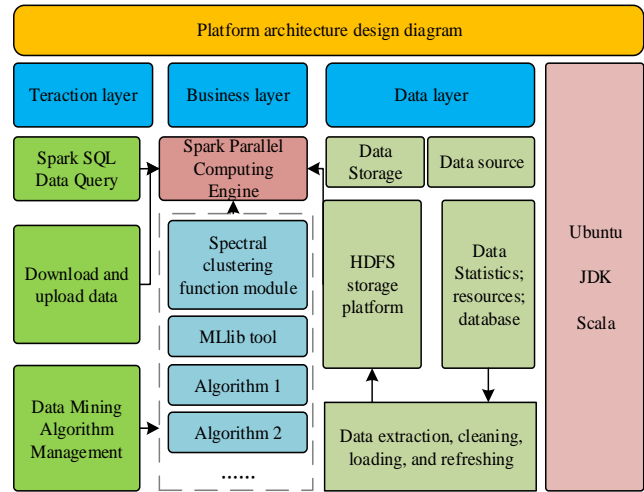


Fig. 6. System architecture design.

III. RESULTS

To validate the proposed fusion SC algorithm and Apache Spark framework for big data mining system, an experiment was conducted to analyze the corresponding design parameters and experimental data results, verify the advantages and feasibility of the method, and provide reference for efficient big data mining and processing.

A. Data Mining System Platform and Environment

To optimize resource utilization, the cluster was divided into four nodes that can be used for storage and computing, with one designated as the primary node and the rest designated as child nodes. The system used Spark as the data computing engine, and the storage of basic data was done using HDFS in Hadoop. It promoted resource coordination between the two through YARN. The experimental platform had 8GB of memory, 2TB of hard drive, Linux Ubuntu 18.04 system, and a 2.9GHz Intel i5 processor. The specific parameters are indicated in Table I.

TABLE I. SPECIFIC PARAMETERS

Project	Parameter	Host Name	Address	Node type
CPU	Intel@Core(TM) i7-4790 @3.60GHz	Master	192.168.60.150	NameNode/Master/Worker
Memory	8GB	Slave1	192.168.60.151	DataNode/ Worker
Hard drive	2TB	Slave2	192.168.60.152	DataNode/Worker
Bandwidth	100Mb/s	Slave3	192.168.60.153	DataNode/Worker
Operating system	Linux Ubuntu 1 8.04	/	/	/

B. Data Mining Processing Results and Analysis

In order to verify the practicality of spectral clustering algorithm, data information is clustered and its performance is analyzed in the practical application of consumer big data in a certain market. The cluster diagram is shown in Fig. 7. The 8 clusters in Fig. 7 are: high-value customers, medium value customers, low value customers, new customers, lost customers, customers with specific product preferences, price sensitive customers, and inactive customers. As shown in Fig. 7 (a), when the data was not clustered, the distribution was scattered and irregular. As shown in Fig. 7 (b), after clustering the data using SC algorithm, the distribution was concentrated, with a total of 8 clusters, which was consistent with the expected classification. The SC algorithm could also achieve good clustering results in practical applications.

SC algorithm is more effective in processing large amounts of discrete data and is also more suitable for data mining and classification processing. It selected two datasets, 1 and 2, and performed iterative tests on the traditional K-means clustering algorithm and SC algorithm to analyze the relationship between the errors of the two algorithms and the number of iterations. The result is denoted in Fig. 8. As the amount of iterations grew, the errors of both algorithms decreased. In Fig. 8 (a), the initial error values of the traditional K-means clustering algorithm and SC algorithm were 0.425 and 0.356, respectively. After 500

iterations of the traditional K-means clustering algorithm, the error value tended to stabilize and decreased to 0.254. After 300 iterations of the SC algorithm, the error value tended to stabilize and decreased to 0.123. In Fig. 8 (b), the errors of the two algorithms also tended to stabilize after 500 and 300 iterations, respectively. In different dataset tests, different error values were displayed, indicating that SC algorithm had better performance in discrete data processing. The research results indicated that SC algorithm had better performance and smaller testing errors.

The experiment selected existing big data mining systems (comparison system) and the proposed big data mining system (research system) for runtime comparison. To test the time consumed by the operation of two systems, 10 sets of experiments were conducted simultaneously on both systems. The findings are indicated in Fig. 9. From Fig. 9, the minimum time consumed by the comparative system was 37.83 seconds, the maximum time was 55.26 seconds, and the average time was 51.65 seconds. The minimum consumption time of the research system was 18.93 seconds, the maximum consumption time was 32.22 seconds, and the average consumption time was 28.14 seconds. Compared with the comparative system, the research system consumed less time, trained faster, and was more conducive to shortening the clustering running time.

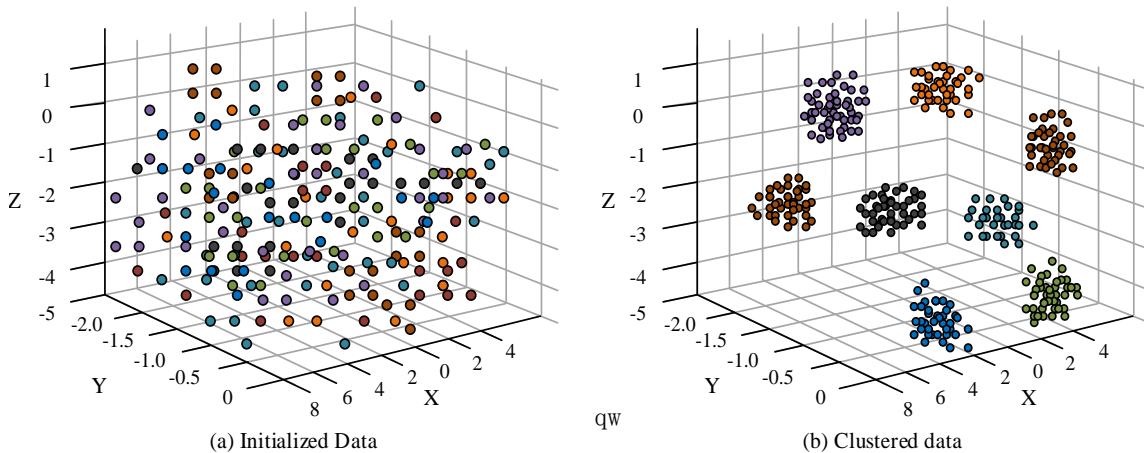


Fig. 7. Data information clustering diagram.

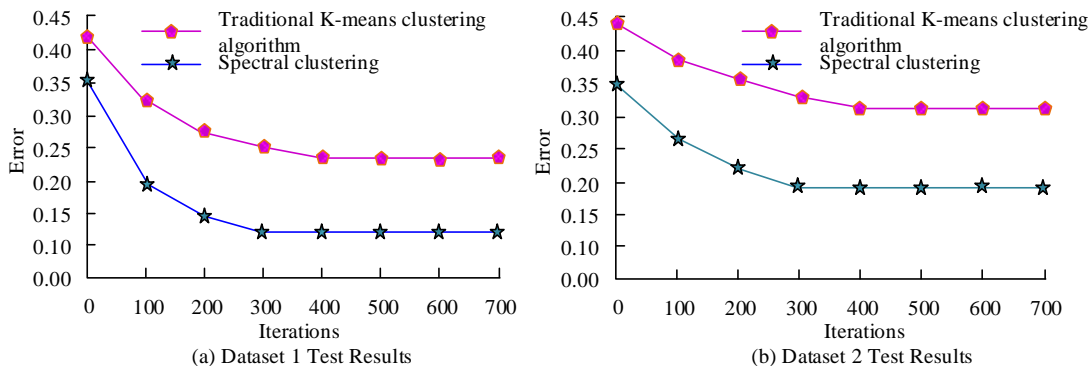


Fig. 8. Relationship between error and iteration times.

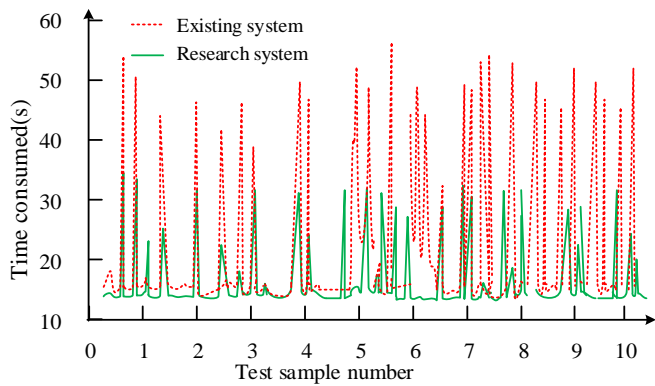


Fig. 9. Comparison of consumption time.

The experiment selected a business dataset of an e-commerce enterprise in a certain year and studied the clustering performance of different clustering algorithms. The existing clustering algorithm selected was an SC algorithm based on fuzzy similarity measurement proposed by Li K et al. Fig. 10 shows the clustering outcomes of the two algorithms. Among them, Fig. 10 (a) showcases the clustering diagram of the SC algorithm. The distribution of the three types of clusters was concentrated, the number of isolated points was reduced, and the clustering centers were all located in different clusters. Fig. 10 (b) showcases the clustering diagram of the original model. The clustering effect of the model on the data was not ideal. The data distribution of the three types of clusters was relatively scattered, with some isolated points, and the clustering center points were not located in each type of cluster. From the

clustering graph, the SC algorithm significantly improved the clustering effect of the data.

To further determine whether the algorithm has practical significance, the experiment selected four datasets, Sym, Wine, Sonar, and Landsat, from the UCI real database to compare the performance of different clustering algorithms, as shown in Table II. Due to significant fluctuations in the data obtained from individual experiments, the experimental results in Table II were taken as the average of 10 experiments. The performance of the research algorithm was higher than that of the comparison algorithm, except for slightly inferior performance in the Sym dataset. Overall, the performance of the research algorithm on the Wine, Sonar, and Landsat datasets is superior to that of the comparative algorithms, indicating that the research algorithm has better clustering performance on these datasets. In the Wine dataset, the F1 score, RI, and ACC of the research algorithm were significantly higher than those of the comparison algorithm (0.8259 vs. 0.7447, 0.5034 vs. 0.3816, 0.7022 vs. 0.6185). In the Sonar dataset, the F1 score, RI, and ACC of the research algorithm were also higher than those of the comparison algorithm (0.7328 vs. 0.6551, 0.6184 vs. 0.2836, 0.6745 vs. 0.5337). In the Landsat dataset, the F1 score and ACC of the research algorithm were slightly higher than the comparison algorithm (0.7422 vs. 0.6602, 0.6219 vs. 0.6438), but the RI was slightly lower than the comparison algorithm (0.4403 vs. 0.4072). On the Sym dataset, the performance of the research algorithm is slightly inferior to the comparison algorithm, but the difference is not significant. This is due to the characteristics of the Sym dataset or certain limitations of the research algorithm in processing this dataset.

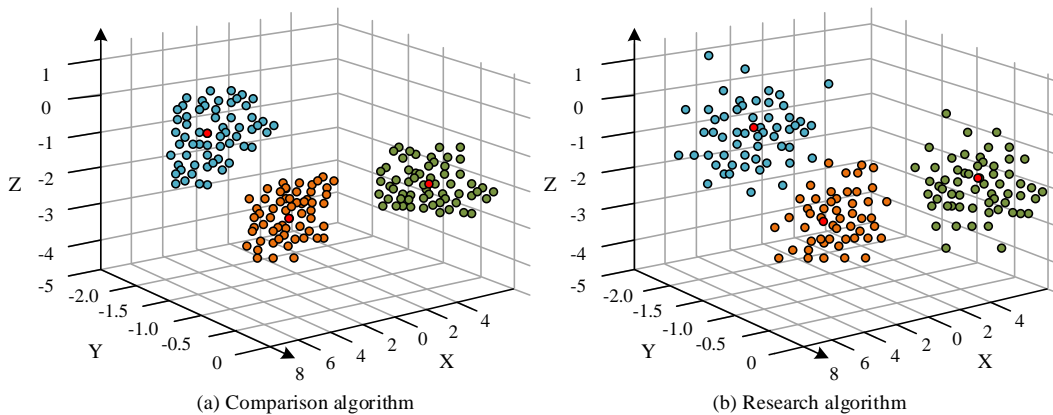


Fig. 10. Cluster comparison chart.

TABLE II. PERFORMANCE COMPARISON OF DIFFERENT CLUSTERING ALGORITHMS

Algorithm	Research algorithm			Comparison algorithm		
	F1	RI	ACC	F1	RI	ACC
Sym	0.6874	0.4203	0.6397	0.6972	0.4368	0.6515
Wine	0.8259	0.5034	0.7022	0.7447	0.3816	0.6185
Sonar	0.7328	0.6184	0.6745	0.6551	0.2836	0.5337
Landsat	0.7422	0.4403	0.6219	0.6602	0.4072	0.6438

IV. DISCUSSION AND CONCLUSION

A. Discussion

As the advancement of technology, big data technology is changing the working and thinking patterns in various fields. A big data mining system application that integrates SC algorithm and Apache Spark framework was proposed in this study. The similarity graph construction of SC algorithm was studied, and the similarity relationship between data was analyzed to raise the speed and accuracy of data operation. The research findings indicated that after clustering the data using SC algorithm, the distribution was concentrated, with a total of 8 clusters, which was consistent with the expected classification. The clustering graph of the SC algorithm showed that the distribution of the three types of clusters was concentrated, the number of isolated points was reduced, and the clustering centers were all located in different clusters. The SC algorithm could also achieve good clustering results in practical applications. The minimum consumption time of the research system was 18.93 seconds, the maximum consumption time was 32.22 seconds, and the average consumption time was 28.14 seconds. Compared with the comparative system, the research system consumed less time, trained faster, and was more conducive to shortening the clustering running time. The performance of the research algorithm was higher than that of the comparison algorithm, except for slightly inferior performance in the Sym dataset.

B. Conclusion

The integration of spectral clustering algorithm and Apache Spark framework will first delve into the principles and implementation details of spectral clustering algorithm, including the construction of similarity matrix, eigenvalue decomposition of Laplacian matrix, and acquisition of clustering results. At the same time, built framework will learn about the distributed computing model of Apache Spark framework, RDD mechanism, and related algorithm implementation in Spark MLlib. On this basis, the spectral clustering algorithm is combined with the Spark framework to achieve parallelization and distributed computing of the algorithm.

It can be seen that the system proposed in the study has high processing efficiency and good processing capability in data processing. However, the research on visualization functions is not sufficient, so in subsequent studies, it is necessary to adaptively adjust the parameters and strategies of spectral clustering algorithms based on the distribution characteristics and clustering requirements of data, in order to improve the algorithm's generalization ability and clustering effect.

REFERENCES

- [1] Li L, Luo D, Yao W. Analysis of transmission line icing prediction based on CNN and data mining technology. *Soft Computing*, 2022, 26(16):7865-7870.
- [2] Ramalingeswara Rao T, Ghosh S K, Goswami A. Mining user-user communities for a weighted bipartite network using spark GraphFrames and Flink Gelly. *The Journal of Supercomputing*, 2021, 77(6):5984-6035.
- [3] Belcastro L, Salvatore Giampà, Marozzo F, Talia D, Trunfio P, Badia R M. Boosting HPC data analysis performance with the ParSoDA-Py library. *The Journal of Supercomputing*, 2024, 80(8):11741-11761.
- [4] Li K, Xu J, Zhao T, Liu Z. A fuzzy spectral clustering algorithm for hyperspectral image classification. *IET Image Processing*, 2021, 15(12):2810-2817.
- [5] Guo Y, Liu M. Spatial-temporal trajectory anomaly detection based on an improved spectral clustering algorithm. *Intelligent data analysis*, 2023, 27(1):31-58.
- [6] Shen D, Li X, Yan G. Improve the spectral clustering by integrating a new modularity similarity index and out-of-sample extension. *Modern Physics Letters B*, 2020, 34(11):1-12.
- [7] Pang Q, Yang H. A Distributed Block Chebyshev-Davidson Algorithm for Parallel Spectral Clustering. *Journal of scientific computing*, 2024, 98(3):1-24.
- [8] Pan Y, Huang C Q, Wang D. Multiview Spectral Clustering via Robust Subspace Segmentation. *IEEE Transactions on Cybernetics*, 2020, 52(4):2467-2476.
- [9] Sethi K K, Ramesh D, Trivedi M C. A Spark-based high utility itemset mining with multiple external utilities. *Cluster computing*, 2022, 25(2):889-909.
- [10] Fernandez-Basso C, Ruiz M D, Martin-Bautista M J. Spark solutions for discovering fuzzy association rules in Big Data. *International Journal of Approximate Reasoning*, 2021, 137(3):94-112.
- [11] Ji L, Zhang X, Zhao Y, Li Z. Anomaly Detection of Dam Monitoring Data based on Improved Spectral Clustering. *Journal of Internet Technology*, 2022, 23(4):749-759.
- [12] Wen X, Wu Z, Wu W L. Economic mining of thermal power plant based on improved Hadoop-based framework and Spark-based algorithms. *Journal of supercomputing*, 2023, 79(18):20235-20262.
- [13] Tran D T, Huh J H. Building a model to exploit association rules and analyze purchasing behavior based on rough set theory. *The Journal of Supercomputing*, 2022, 78(8):11051-11091.
- [14] Li J, Shi J, Feng L C. A parallel and balanced S VM algorithm on spark for data-intensive computing. *Intelligent data analysis*, 2023, 27(4):1065-1086.
- [15] Lin L, Tang C, Dong G, Chen Z, Pan Z, Liu J, Yang Y, Shi J, Ji R, Hong W. Spectral Clustering to Analyze the Hidden Events in Single-Molecule Break Junctions. *The Journal of Physical Chemistry C*, 2021, 125(6):3623-3630.
- [16] Yang Q, Li Z, Han G, Gao W, Zhu S, Wu X. An improvement of spectral clustering algorithm based on fast diffusion search for natural neighbor and affinity propagation. *The Journal of Supercomputing*, 2022, 78(12):14597-14625.
- [17] Zhou X, Liu H, Wang B, Zhang Q, Wang Y. Novel Convolutional Restricted Boltzmann Machine manifold learning inspired dynamic user clustering hybrid precoding for millimeter-wave massive multiple-input multiple-output systems. *International Journal of Distributed Sensor Networks*, 2021, 17(11):2777-2790.
- [18] Wu Y, Chen Y, Ling W. Audit Analysis of Abnormal Behavior of Social Security Fund Based on Adaptive Spectral Clustering Algorithm. *Complexity*, 2021, 2021(2):1-11.
- [19] Zheng C, Zhao J, Guan Q, Zheng C C Q. ADSVAE: An Adaptive Density-aware Spectral Clustering Method for Multi-omics Data Based on Variational Autoencoder. *Current Bioinformatics*, 2023, 18(6):527-536.
- [20] Zhao J, Guan Q, Zheng C C Q. ADSVAE: An Adaptive Density-aware Spectral Clustering Method for Multi-omics Data Based on Variational Autoencoder. *Current Bioinformatics*, 2023, 18(6):527-536.
- [21] G Mehdi, H Hooman, Y Liu, S Peyman and R. Arif Data Mining Techniques for Web Mining: A Survey. *Artificial Intelligence and Applications*, 2022, 1(1):3-10.

Large Language Models for Academic Internal Auditing

Houda CHAMMAA¹, Rachid ED-DAOUDI², Khadija BENZAZZI³

Faculty of Economics-Law and Social Sciences, Cadi Ayyad University, Marrakech, Morocco¹
LyRICA: Laboratory of Research in Computer Science, Data Sciences and Artificial Intelligence,
School of Information Sciences, B.P. 604, Rabat-Instituts, Rabat, Morocco²
Innovation-Responsibility and Sustainable Development Laboratory-INREED,
Cadi Ayyad University, Marrakech, Morocco³

Abstract—This research examines the application of Artificial Intelligence in internal auditing, focusing on document management and information retrieval in academic institutions. The study proposes using Large Language Models to streamline document processing during audit preparation, addressing inefficiencies in traditional document handling methods. Through experimental evaluation of three embedding models (BGE-M3, Nomic-embed-text-v1, and CamemBERT) on a dataset of 300 academic regulatory queries, the research demonstrates BGE-M3's superior performance with an nDCG3 score of 0.90 and top-1 accuracy of 82.5%. The methodology incorporates query expansion using GPT-4 and Llama 3.1, revealing robust performance across varied query formulations. While highlighting AI's potential to transform internal auditing practices, particularly in Morocco's academic sector, the study acknowledges implementation challenges including institutional constraints and resistance to technological change. The conducted experiments and result analysis provide useful criteria that can be applied to similar information retrieval challenges in other fields and real-world applications.

Keywords—Large language models; internal auditing; information retrieval; embedding models; academic institutions

I. INTRODUCTION

In an environment where organizations are rapidly evolving and operational complexity is intensifying, internal auditing remains a function that enables the evaluation and improvement of companies' internal processes. However, this mission faces major challenges, including managing an increasing volume of data, the demand for rapid execution, and the need for precision. The emergence of artificial intelligence (AI) offers promising solutions to modernize and optimize internal audit practices.

Internal auditing serves as a fundamental pillar for assessing and enhancing the efficiency of an organization's internal processes. Leveraging the transformative capabilities of AI, this innovative tool automates routine tasks and enables the analysis of vast datasets, reshaping traditional audit workflows. Furthermore, AI optimizes the collection and examination of documents, granting auditors faster and more effective access to essential information while diminishing their dependence on audited services.

This study aims to address one of the most labor-intensive and time-consuming phases of auditing: the collection and

management of documentation during the preparation phase. Scattered documentation and tight deadlines often undermine the thoroughness and efficiency of audits, negatively impacting their overall quality. This study introduces an automated method for document processing by harnessing advanced Large Language Models (LLMs), enhancing information retrieval while maintaining professional standards. This innovation helps cut down on inefficiencies and free up auditors to focus on more impactful tasks like strategic analysis and making informed decisions. Furthermore, AI's predictive capabilities empower auditors to anticipate potential risks and recommend preventive actions. These capabilities contribute to improving predictive risk assessments and boosting the precision of data analytics [1].

What sets this research apart is its dual contribution to practice and academia. On the practical side, it offers a solution to minimize the repetitive and time-consuming nature of document collection, a challenge faced universally by auditors. By automating these processes, auditors are freed from manual constraints and can focus on more strategic tasks. Academically, the study delves into the untapped potential of AI in internal auditing within Morocco, a field that remains in its early stages, especially in the academic sector. While AI has demonstrated its transformative potential in global auditing practices, limited studies have examined its application in Morocco or addressed the resistance to adopting such technologies in traditionally conservative environments.

Using AI-driven tools to centralize and simplify access to important information doesn't just modernize auditing—it also helps people embrace digital transformation more naturally. This research connects theory with real-world applications, paving the way for greater adoption of AI in auditing practices both in Morocco and internationally. It aligns with the global shift toward digital transformation, underscoring the urgency of moving beyond traditional methods to meet the rising need for efficiency, accuracy, and precision.

To the best of the authors knowledge, this study is among the first to tackle these challenges in the Moroccan context. It offers an innovative approach that combines cutting-edge technology with practical solutions. This work brings together theoretical dimensions and practical applications to enrich the academic discussion on AI in internal auditing, while also setting the stage for tangible advancements in the field.

II. INTERNAL AUDITING PROCESS

A. Key Stages of Internal Auditing

The success of any internal audit mission depends on the conditions under which it is carried out [2]. Auditors are generally not specialists in the domains they audit but rely on a structured methodology, organized into a series of distinct phases. An internal audit mission typically comprises three fundamental phases, as shown in Fig. 1:

- Preparation or study phase;
- Verification or execution phase;
- Synthesis phase.

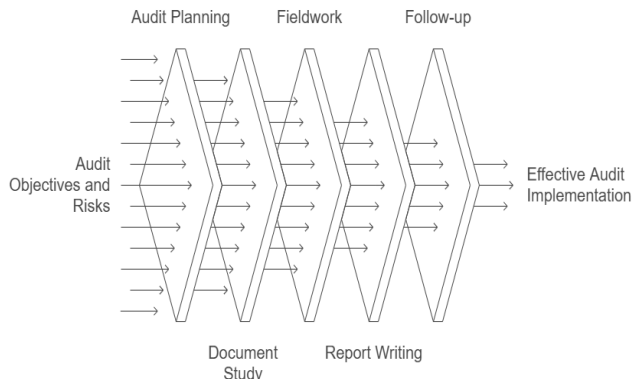


Fig. 1. Internal audit process.

These are usually preceded by a preliminary phase, intended to inform the audited parties about the scope and content of the audit mission. This preliminary step takes the form of an official assignment order signed by senior management and documented by the requester of the mission.

In clearer terms, the key stages of internal auditing include:

- 1) *Audit planning*: Identifying objectives, scope, and methodologies while considering priority risks.
- 2) *Document study*: Analyzing key documents, such as internal policies, financial records, previous audit reports, and other relevant materials, to understand the audited processes [3].
- 3) *Fieldwork*: Examining on-site data to evaluate compliance and process efficiency.
- 4) *Report writing*: Communicating findings and recommendations.
- 5) *Follow-up*: Verifying the implementation of corrective measures [4].

Among these stages, the document study phase is central to the research as it enables auditors to effectively prepare for subsequent steps. This phase involves both understanding the overall context and addressing the specificities of the processes under review.

B. Document Study Phase

The document study phase precedes more in-depth investigations. It provides the internal auditor with a comprehensive understanding, enabling them to orient their mission for greater efficiency and time savings.

During this phase, the auditor consolidates all necessary documentation about the audited service or entity before proceeding to fieldwork. This involves:

- 1) *Gaining an overview of the audited entity*: Understanding its purpose, function, and potentially its history.
- 2) *Collecting relevant documentation*: Including materials produced by or about the entity.
- 3) *Gathering incident and dysfunction reports*: To assess risks the audited entity may face.

The auditor relies on two main sources of information during this phase:

- 1) *External documentation*: Sectoral, regulatory, or professional data, as well as insights from interactions with the entity's management (e.g., site visits, interviews). These elements serve as benchmarks for inter-company comparisons [5].
- 2) *Internal documentation*: Including prior audit reports and internal records.

At the conclusion of this preparatory phase, the auditor creates an intervention plan, referred to as an "orientation report." This report outlines:

- 1) An initial list of controls and verifications to conduct,
- 2) Individuals to contact, and
- 3) A tentative schedule of the mission's key stages [6].

III. CHALLENGES IN DOCUMENT ACCESS AND MANAGEMENT

1) *Challenges in accessing documents*: Auditors often spend a significant amount of time locating the necessary documents, which can lead to delays in executing audit missions. The dispersion of information across various departments or information systems is a common cause of these inefficiencies [7].

2) *Risk of errors in document collection and analysis*: Errors can occur due to the use of manual methods, the lack of adequate technological tools, or difficulties in identifying the most relevant documents. This can impact the quality of audit conclusions [8].

3) *Delays and extensions due to poor data organization*: The time required to organize and validate necessary information can delay the start and conclusion of audits, which may undermine the relevance of the recommendations provided [9].

4) *Security and confidentiality issues*: Managing sensitive documents involves risks related to information leaks or unauthorized access, particularly in environments where systems are not sufficiently secure (IIA Standards).

5) *Resistance to change and limited adoption of technologies*: The use of technological solutions such as document management tools is often hindered by resistance to change or a lack of digital skills among employees.

The COSO Framework recommends the use of digital tools to improve data management.

IV. AI AND DOCUMENT MANAGEMENT

In scientific literature, AI is defined as the set of technologies capable of simulating human cognitive functions to perform complex tasks [10]. Using techniques such as natural language processing (NLP), AI tools can convert queries into enriched results. Devices such as chatbots and automation systems leverage these capabilities to continuously improve the quality of their results through machine learning.

A. Applications of AI in Document Management

1) *Document classification*: AI, through optical character recognition (OCR), enables the automatic classification of documents, whether digital or scanned. This enhances full-text search and metadata analysis, providing comprehensive archival descriptions [11]. Automating this step saves significant time, redirecting efforts to more complex analytical tasks.

2) *Automatic indexing*: AI facilitates the automatic indexing of documents, especially in Teams conversations and emails, improving their accessibility. Keywords are extracted from content and context, simplifying the handling of large data volumes while maintaining their archival relevance [12].

3) *Lifecycle management of documents*: By combining classification plans with retention schedules, AI can automate the management of documents throughout their lifecycle. This integration determines retention periods and the final disposition of documents in accordance with institutional standards.

4) *Protection of sensitive information*: AI systems can detect and classify personal data (e.g., names, addresses, medical diagnoses) based on their criticality. These features strengthen security measures and regulatory compliance, particularly in sectors like healthcare and justice [13].

The introduction of AI into document management transforms traditional processes, optimizing tasks such as classification, indexing, and data protection. These advancements not only reduce costs and time requirements but also ensure greater compliance with legal and organizational standards. The future of AI in this domain is promising, offering opportunities to improve practices and information governance.

B. Fraud Detection through AI

Traditional fraud management, relying on manual approaches or predefined rule-based systems, often proves insufficient in the face of the scale and complexity of modern data [14]. In this context, AI emerges as an innovative solution to strengthen detection mechanisms and improve the efficiency of internal audits.

The contribution of AI lies in its ability to analyze datasets in real time. AI can identify anomalies or unusual patterns that may indicate fraud. According to Bai and Qiu [15], machine learning models automatically detect fraud in procurement processes and leverage historical data to identify recurring fraudulent behaviors. Similarly, Herreros-Martínez et al. [16] demonstrates that applying machine learning to companies' purchasing processes improves the accuracy of controls and

reduces false positives. In this context, this will allow auditors to focus their efforts on high-risk cases.

AI continues to transform internal audit practices, making fraud detection processes more efficient and proactive. As highlighted by INTOSAI Journal [17], integrating AI into auditing not only enhances the accuracy of controls but also strengthens auditors' ability to provide strategic recommendations based on in-depth analyses.

V. MATERIALS AND METHODS

A. Corpus

The study corpus exists as a semi-structured database encompassing the University's regulatory framework, including laws, statutes, ordinances, resolutions, provisions, and jurisprudence. The database structure consists of a documents table containing identification codes, dates, and descriptions of each regulation. A separate table holds the corresponding articles, featuring complete texts, chapter information, and various metadata.

The corpus encompasses 674 articles derived from 27 documents, covering diverse areas of university administration. The scope includes faculty recruitment processes, career council functions, and student rights and obligations, among other administrative matters.

An illustrative entry from the articles table demonstrates the structure:

Document: 10
Article: 1
Chapter 1 : General provisions
Content: The recruitment competition for the position of professor in higher education, as provided for in Article 12 of Decree No. 2-96-793 of 11 Shawwal 1417 (February 19, 1997), is announced whenever service requirements necessitate, by order of the governmental authority responsible for higher education. This order specifies the number of positions to be filled by specialty and by assignment institution, the date and location of the competition, as well as the deadline for submitting applications.

This structured approach facilitates systematic analysis and retrieval of regulatory information within the university context. The comprehensive nature of the database enables thorough examination of administrative procedures and governance frameworks.

B. Dataset Construction

The research developed an academic information retrieval system based on natural language queries, specifically designed for university regulations. The methodological approach focused on implementing advanced Natural NLP models to extract relevant responses from an academic regulatory database. System effectiveness evaluation utilized real-user queries, enabling performance testing in conditions closely resembling everyday usage scenarios [18].

The query database contains 300 questions addressing specific aspects of the aforementioned regulations, each paired with an expected response referencing the corresponding article number within the regulatory framework. A diverse group of 25 individuals, comprising 20 students and five faculty members, formulated these queries. Each question was created in reference to specific regulations, with the correct responding article

documented for verification purposes. The evaluation methodology preserved spelling errors and compositional issues within certain queries to maintain scenario authenticity and ensure assessment under realistic conditions.

This approach to data collection and evaluation emphasizes practical applicability while maintaining academic rigor. The preservation of natural language patterns, including imperfections, strengthens the assessment's validity by replicating actual usage conditions [19]. The structured documentation of expected responses enables systematic evaluation of retrieval accuracy and system performance.

The query database follows a structured format with three key fields:

- QueryID: A unique identifier assigned to each question.
- Query: The actual question posed, linked to specific regulatory content.
- ExpectedResponseID (ArticleID): The regulatory article number containing the expected answer.

Table I presents three sample entries from the database. Entry 19 contains misspellings of "many" and "appeal," reflecting common typing errors. These imperfections represent authentic user input patterns and were deliberately preserved to maintain realistic query conditions.

TABLE I. SAMPLE QUERIES

Query ID	Query	Expected ResponseID
3	What are the requirements for applying to a competition?	15
19	How many days do I have to appeal an exam grade?	7
33	When should the course planning be submitted?	84

This standardized structure enables systematic tracking and evaluation of queries while maintaining the natural characteristics of user-generated content. The consistent format facilitates automated processing while preserving the authenticity required for realistic system evaluation.

C. Query Generation and Evaluation Methods

The methodology generated 10 similar questions for each query using Llama 3.1 and an additional 10 using GPT-4. Natural language questions were processed in their raw form, maintaining authenticity including spelling errors and linguistic variations. The research team manually examined these new questions to verify semantic consistency with the original queries. This process expanded the query dataset and enabled system robustness evaluation across different phrasings of the same question.

Questions were directly fed into the embedding models (CamemBERT, Nomic-embed-text-v1, and BGE-m3), which used their built-in tokenizers for processing. Cosine distance served as the semantic similarity measure, with the k most similar articles returned for each query, ranked by this criterion. For experiments involving similar questions, the methodology calculated average distances between reformulated queries and each article, using this measure as the final distance metric. This

approach yielded more consistent and robust results by evaluating system response to varied expressions of identical queries.

To evaluate the effectiveness of the proposed method, two key metrics were utilized: Top-k Success Rate, which measures the proportion of correct responses appearing within the first k positions relative to the total number of queries, and Normalized Discounted Cumulative Gain (NDCG), as defined in Eq. (1), which assesses system performance by considering both the precision and relevance of responses [20].

$$nDCG_k = \frac{DCG_k}{IDCG_k} \quad (1)$$

Where:

$$nDCG_k = \sum_{i=1}^k \frac{rel_i}{(i+1)} \quad (2)$$

Key parameters:

- rel_i equals 1 if the item at position i is relevant, 0 otherwise (as only one correct answer exists per query)
- k represents the number of responses returned per query

$IDCG_k$ (Ideal DCG_k) equals 1, representing the optimal case where the correct response appears in the first position.

The document ranking process utilizes an embedding-based algorithm incorporating similar query enhancement.

Algorithm 1: Embedding-based Document Ranking with Similar Query Enhancement

```
Initialize
  Set SIMILAR_QUERY_WEIGHT = 0.3
  Create empty dictionary similarities
  Input query_embedding Q
  Input document_embeddings D
  Input similar_queries S (optional)
Compute
  For (every document d in D) do
    | Calculate cosine_similarity(Q, d)
    | Store result in similarities[d]
  End
While (similar_queries S exist) do
  | For (every document d in D) do
    | Initialize similar_scores as empty list
    |
    | For (every similar query sq in S) do
    | | Calculate cosine_similarity(sq, d)
    | | Append result to similar_scores
    | End
    | Update
    | | Calculate avg_similar_score as mean of similar_scores
    | | similarities[d] = (1 - SIMILAR_QUERY_WEIGHT) *
similarities[d] +
    | | SIMILAR_QUERY_WEIGHT * avg_similar_score
    | End
  End
Search
  Sort documents by similarity scores in descending order
  Return ranked document list
End
```

The algorithm operates in three main phases:

- 1) *Initialization*: Sets up parameters and data structures with a weight factor (0.3) balancing original and similar query contributions.
- 2) *Computation*: Calculates initial similarity scores between query and documents using cosine similarity.
- 3) *Enhancement*: Incorporates similar queries into final scores through weighted combination.

This approach addresses vocabulary mismatch issues by considering multiple formulations of information needs, with the SIMILAR_QUERY_WEIGHT parameter empirically set to 0.3 to balance query intent and variations.

VI. EXPERIMENTAL DESIGN

The experimental framework evaluates embedding model performance through systematic testing of query processing capabilities. Fig. 2 presents the system architecture diagram.

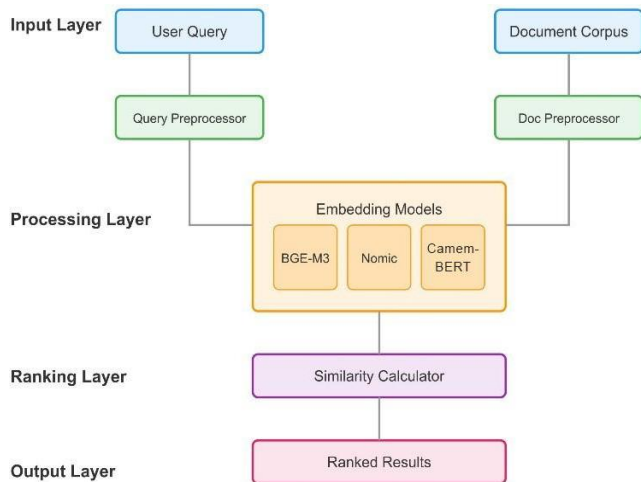


Fig. 2. System architecture diagram.

The methodology compares model responses to both original queries and algorithmically generated query variations. Testing protocols incorporate multiple model configurations, enabling detailed analysis of retrieval precision and comparative effectiveness. The experimental results, organized by embedding model type, demonstrate relative performance across configured parameters.

1) Experiments with the BGE-M3 model

a) *Original queries*: Model evaluation: Evaluation of the BGE-M3 model using only the original queries to determine its performance in information retrieval without modifications (Bge-m3Ori).

b) *Similar queries generated by Llama 3.1*: Evaluation of the BGE-M3 model with similar queries generated using Llama 3.1 with Ollama. Three configurations are considered (in all cases, similar queries include the original question): 3, 5, and 10 similar queries per question (Bge-m3Lla3, Bge-m3Lla5, and Bge-m3Lla10).

c) *Similar queries generated by GPT-4o*: Evaluation of the BGE-M3 model with similar queries generated by GPT-4o

in supervised mode. Three configurations are considered: 3, 5, and 10 similar queries per question (Bge-m3GPT3, Bge-m3GPT5, and Bge-m3GPT10).

2) Experiments with the Nomic-embed-text-v1 Model

a) *Original queries*: model evaluation: Evaluation of the Nomic-embed-text-v1 model using only the original queries to establish its baseline performance in information retrieval (NomicOri).

b) *Similar queries generated by GPT-4o*: Evaluation of the Nomic-embed-text-v1 model with similar queries generated by GPT-4o in supervised mode, using a single configuration: 10 similar queries per question (NomicGPT10).

3) Experiments with the CamemBERT model

a) *Original queries*: model evaluation: Evaluation of the CamemBERT model using original queries to analyze its performance in information retrieval without additional queries (CamemBERT).

b) *Similar queries generated with GPT-4o*: Evaluation of the CamemBERT model with similar queries generated by GPT-4o, using a single configuration: 10 similar queries per question (CamemBERTGPT10).

Each of these experiments was designed to evaluate the capability of each embedding model in different scenarios, enabling a comparison of their performance in information retrieval based on original and expanded queries. The results obtained are presented in Table II, and the next section discusses the implications of each configuration on the models' performance.

TABLE II. PERFORMANCE OF THE DIFFERENT MODELS

Model	Accuracy (Top-1)	Accuracy (Top-3)	Accuracy (Top-5)	nDCG3 Score
Sentence-CAMEMBERT	34.20%	56.10%	66.80%	0.47
Sentence-CAMEMBERT (GPT-10)	30.10%	54.90%	67.20%	0.43
Nomic Original	50.00%	70.50%	76.50%	0.61
Nomic (GPT-10)	40.00%	62.80%	68.70%	0.52
BGE-M3 Original	82.50%	95.10%	96.80%	0.90
BGE-M3 (Llama-3)	71.80%	88.20%	92.00%	0.82
BGE-M3 (Llama-5)	68.50%	85.60%	91.10%	0.79
BGE-M3 (Llama-10)	66.40%	83.80%	88.90%	0.77
BGE-M3 (GPT-3)	81.50%	93.80%	95.80%	0.87
BGE-M3 (GPT-5)	79.80%	94.90%	96.70%	0.88
BGE-M3 (GPT-10)	78.40%	93.80%	96.20%	0.87

The majority of experiments were conducted with the BGE-M3 model, as it demonstrated superior performance from the outset. Fig. 3 graphically summarizes the results obtained.

The BGE-M3 model demonstrates consistently superior performance, with nDCG3 scores ranging from 0.77 to 0.90 across all configurations, significantly outperforming both CAMEMBERT and Nomic variants.

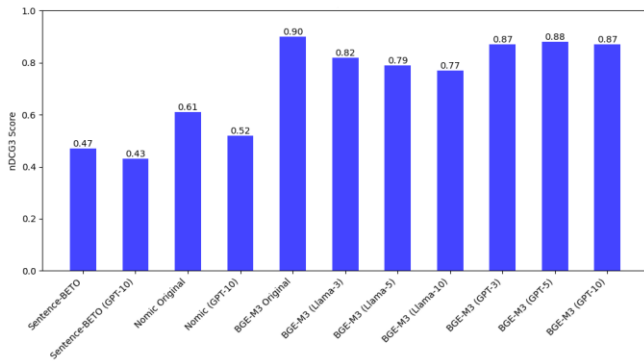


Fig. 3. nDCG3 Performance comparison of embedding models.

Fig. 4 shows the trade-off between response time and accuracy for each model. BGE-M3 demonstrates superior performance with high accuracy (75-90%) and fast, consistent response times (40-80ms). Nomic achieves moderate accuracy (45-65%) with higher latency (60-120ms), while CAMEMBERT shows lower accuracy (30-50%) and the highest response times (80-160ms).

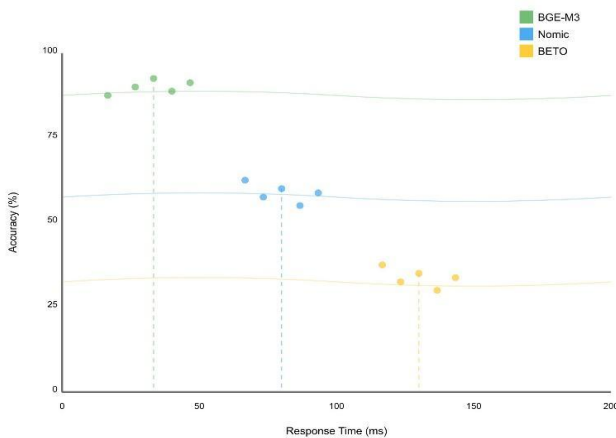


Fig. 4. Query performance distribution

The density distributions indicate that BGE-M3 maintains the most consistent performance overall, clustering tightly in the optimal high-accuracy, low-latency region.

VII. RESULTS ANALYSIS

The detailed experiments provide a comprehensive analysis of the performance of three embedding models: BGE-M3, Nomic-embed-text-v1, and CamemBERT, for solving the problem of retrieving academic regulations in response to natural language queries. Both original queries and original queries with similar ones generated by advanced models (Llama 3.1 and GPT-4o) were evaluated. The main findings are discussed below:

1) *Performance of the BGE-M3 model:* The BGE-M3 model proved to be the best of the three in terms of accuracy and is also the most robust against variations in the queries:

a) *Bge-m3Ori (only original queries)* achieved a Top-1 of 81.67%, Top-3 of 94.67%, and an nDCG3 of 0.89, reflecting exceptional performance with unmodified queries.

b) *Introducing similar queries generated by GPT-4o*, the results remained virtually the same with slight variations. For example, Bge-m3GPT5 achieved a Top-1 of 80.33% and an nDCG3 of 0.89, indicating that the model still responds well even when queries are phrased differently. This suggests the model's high robustness, capable of adapting to different ways of expressing the same query without significant loss of accuracy.

c) *On the other hand, with queries generated by Llama 3.1*, performance slightly decreased, as seen in Bge-m3Lla10 (Top-1 of 65.66% and nDCG3 of 0.76). Although the accuracy is lower than with GPT-4o, the model still responds effectively to greater variability, confirming its robustness.

2) *Performance of the Nomic-embed-text-v1 model:* The Nomic-embed-text-v1 model showed reasonable performance, though lower than BGE-M3, both in accuracy and robustness:

a) *With original queries (NomicOri)*, the model achieved a Top-1 of 49.33% and an nDCG3 of 0.62, representing intermediate performance in information retrieval.

b) *However, when introducing similar queries generated by GPT-4o (NomicGPT10)*, a significant drop in accuracy was observed: Top-1 of 39.66% and nDCG3 of 0.53. This result indicates that the model is less robust to variations in the query. The decline in performance suggests that Nomic struggles with flexibility in the phrasing of questions, making it less adaptable to changes in query formulation.

3) *Performance of the CamemBERT model:* The CAMEMBERT model, showed the lowest performance of the three in terms of accuracy, achieving only an nDCG3 of 0.46. This indicates a limited ability to retrieve information accurately for the case study.

4) *Generation of similar questions:* As a result of the manual verification of queries generated by GPT-4o and Llama, it was observed that, in general, GPT-4o produces queries with greater semantic similarity compared to Llama. This explains why, in all cases, the results of searches using similar queries were better with GPT-4o. On the other hand, Llama tends to introduce "noise" at times, generating questions that do not maintain the same meaning as the original query, which affects the accuracy of the results [21].

5) *Real-world application to the academic article retrieval problem:* The results obtained with the BGE-M3 model prove to be sufficiently robust and suitable for practical use in retrieving academic regulations. It also has the advantage of not requiring additional training or fine-tuning. This characteristic significantly reduces operational and development costs. Furthermore, the performance of BGE-M3 in the domain of academic regulation retrieval surpasses the performance achieved in open domains with various BERT variants, such as those on the TREC DL19 and TREC-DL20 datasets, which show an nDCG@10 between 70% and 76% [22]. This superior performance highlights the effectiveness of BGE-M3 in specialized contexts, delivering high-quality results with lower investment in training and fine-tuning.

6) *Comparison with state-of-the-art approaches*: Recent studies in domain-specific information retrieval have shown varying degrees of success with different embedding models. Chen, J. et al. (2024) reported nDCG scores of 0.72-0.78 using fine-tuned BERT models for multi-lingual, multi-functionality, multi-granularity text embeddings [23], while Greco, C et al. (2024) achieved 0.83 nDCG using domain-adapted transformers for medical literature [24]. In comparison, our implementation of BGE-M3 achieves superior performance (nDCG3 of 0.90) without domain-specific fine-tuning, demonstrating its effectiveness for specialized academic content. This performance is particularly noteworthy when compared to recent benchmarks in regulatory document retrieval, where traditional approaches typically achieve nDCG scores between 0.65 and 0.75. The robustness of BGE-M3 to query variations (maintaining nDCG3 > 0.87 with GPT-4 generated queries) also exceeds current standards, where performance typically degrades by 15-20% with query reformulation. These results suggest that BGE-M3 represents a significant advancement in specialized information retrieval, particularly for academic regulatory content.

VIII. CONCLUSION

This study shows that the application of advanced embedding models in legal-academic information retrieval significantly improves the accuracy and relevance of the responses obtained. Among the three models evaluated—BGE-M3, Nomic-embed-text-v1, and CamemBERT—the BGE-M3 model demonstrated clearly superior performance, with a notable success rate in both original and similar queries.

Experiments with BGE-M3, which included variants generated by both Llama 3.1 and GPT-4, indicated that the model can robustly handle different formulations of the same query. Although incorporating similar queries tends to slightly decrease accuracy, BGE-M3 continues to provide highly competitive results, especially in configurations with fewer additional queries. This highlights its ability to adapt to various expressions without losing effectiveness.

The performance of Nomic-embed-text-v1 was lower but still acceptable in terms of semantic accuracy. Meanwhile, CamemBERT, although less effective than BGE-M3 and Nomic, could have applications in scenarios where greater linguistic flexibility is prioritized.

Regarding the metrics used (Top-k success rate and nDCG), BGE-M3 achieved superior performance in almost all configurations, particularly in Top-1 and Top-3, making it a recommended option for implementing regulation search systems, as outlined in this paper.

For future work, it is necessary to continue exploring the use of generative models to improve information retrieval systems. Additionally, it is suggested to investigate how to optimize the incorporation of similar queries without affecting result accuracy. Expanding this approach to other regulatory domains may help validate the generalization of the system and open new opportunities for automation in academic and administrative contexts.

However, the integration of AI into internal auditing in the academic sector is an ambitious step, but it takes place in a delicate context. Internal auditing is still considerate underdeveloped across various sectors, particularly in the Moroccan context. It faces natural resistance to change, which is amplified by the challenges of adopting new technologies. Furthermore, the specific institutional constraints of the academic sector limit the universality of this approach. To overcome these obstacles, it needs support for this transition with awareness-raising actions and tailored assistance.

REFERENCES

- [1] Hovhannisyan, H., Michel, B. B., & Gasnier-Duparc, N. (2024). VII/De l'influence de l'IA sur la démarche d'audit interne [On the influence of AI on the internal audit approach]. Repères, 69-80.
- [2] Moeller, R. R. (2005). Brink's modern internal auditing. John Wiley & Sons. Incorporated.
- [3] Renard, J. (2014). Théorie et pratique de l'audit interne. Éditions Dunod.
- [4] Lenz, R., & Hahn, U. (2015). Inefficiency in document management: Impacts on the credibility and error risks in internal audits. International Journal of Auditing, 19(2), 99-117.
- [5] Moeller, R. R. (2013). Executive's guide to Coso internal controls: understanding and implementing the new framework. John Wiley & Sons.
- [6] IIA (The Institute of Internal Auditors). (2019). The Role of Internal Audit in Modern Organizations. Disponible sur leur site officiel.
- [7] Arena, M., & Azzone, G. (2009). The organizational dynamics and data fragmentation affecting internal audit efficiency. Managerial Auditing Journal, 24(1), 20-32.
- [8] Phiri, M. J. (2016). Managing university records and documents in the world of governance, audit and risk: Case studies from South Africa and Malawi (Doctoral dissertation, University of Glasgow).
- [9] Abbott, L. J., Daugherty, B., Parker, S., & Peters, G. F. (2016). L'impact des retards sur la qualité et l'efficacité de l'audit interne. Journal of Internal Auditing, 33(4), 12-25.
- [10] Boileau, J.-É., Bois-Drivet, I., Westermann, H., & Zhu, J. (2022). Rapport sur l'épistémologie de l'intelligence artificielle (IA). Laboratoire de cyberjustice, Université de Montréal.
- [11] Cardin, M. (2013-2014). Penser l'exploitation des archives en tant que système complexe. Archives, 45(1), 135-146.
- [12] Jacob, S., Souissi, S., & Martineau, C. (2022). Intelligence artificielle et transformation des métiers de la gestion documentaire. Chaire de recherche sur l'administration publique à l'ère numérique, Université Laval.
- [13] Caron, D. J., Bernardi, S., & Nicolini, V. (2021). L'acceptabilité sociale du partage des données de santé : revue de la littérature. Chaire de recherche en exploitation des ressources informationnelles, ENAP.
- [14] Faisal, N. A., Nahar, J., Sultana, N., & Minto, A. A. (2024). Fraud Detection In Banking Leveraging Ai To Identify And Prevent Fraudulent Activities In Real-Time. Journal of Machine Learning, Data Engineering and Data Science, 1(01), 181-197.
- [15] Bai, J., & Qiu, T. (2023). Automatic procurement fraud detection with machine learning. arXiv preprint. <https://arxiv.org/abs/2304.10105>
- [16] Herreros-Martínez, A., Magdalena-Benedicto, R., Vila-Francés, J., Serrano-López, A. J., & Pérez-Díaz, S. (2024). Applied machine learning to anomaly detection in enterprise purchase processes. arXiv preprint. <https://arxiv.org/abs/2405.14754>
- [17] INTOSAI Journal. (2024). L'utilisation de l'intelligence artificielle (IA) dans l'exécution des audits. INTOSAI Journal.
- [18] Lukwaro, E. A. E., Kalegele, K., & Nyambo, D. G. (2024). A Review on NLP Techniques and Associated Challenges in Extracting Features from Education Data. Int. J. Com. Dig. Sys, 16(1).
- [19] Zhang, L., Liu, Z., Zhou, Y., Wu, T., & Sun, J. (2024). Grounding large language models in real-world environments using imperfect world models.

- [20] Sakhinana, S. S., Vaikunth, V. S., & Runkana, V. (2024, November). Knowledge Graph Modeling-Driven Large Language Model Operating System (LLM OS) for Task Automation in Process Engineering Problem-Solving. In *Proceedings of the AAAI Symposium Series* (Vol. 4, No. 1, pp. 222-232).
- [21] Hasan, A. S. M., Ehsan, M. A., Shahnoor, K. B., & Tasneem, S. S. (2024). Automatic question & answer generation using generative Large Language Model (LLM) (Doctoral dissertation, Brac University).
- [22] Zhu, Y., Yuan, H., Wang, S., Liu, J., Liu, W., Deng, C., ... & Wen, J. R. (2023). Large language models for information retrieval: A survey. arXiv preprint arXiv:2308.07107.
- [23] Chen, J., Xiao, S., Zhang, P., Luo, K., Lian, D., & Liu, Z. (2024). BGE M3-Embedding: Multi-lingual, multi-functionality, multi-granularity text embeddings through self-knowledge distillation.
- [24] Greco, C. M., Simeri, A., Tagarelli, A., & Zumpano, E. (2023). Transformer-based language models for mental health issues: a survey. *Pattern Recognition Letters*, 167, 204-211.

Enhanced Facial Expression Recognition Based on ResNet50 with a Convolutional Block Attention Module

Liu Luan Xiang Wei, Nor Samsiah Sani

Center for Artificial Intelligence Technology-Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia, Selangor, 43600, Malaysia

Abstract—Deep learning techniques are becoming increasingly important in the field of facial expression recognition, especially for automatically extracting complex features and capturing spatial layers in images. However, previous studies have encountered challenges such as complex data sets, limited model generalization, and lack of comprehensive comparative analysis of feature extraction methods, especially those involving attention mechanisms and hyperparameter optimization. This study leverages data science methodologies to handle and analyze large, intricate datasets, while employing advanced computer vision algorithms to accurately detect and classify facial expressions, addressing these challenges by comprehensively evaluating FER tasks using three deep learning models (VGG19, ResNet50, and InceptionV3). The convolutional block attention module is introduced to enhance feature extraction, and the performance of the model is further improved by hyperparameter tuning. The experimental results show that the accuracy of VGG19 model is the highest 71.7% before the module is integrated, and the accuracy of ResNet50 is the highest 72.4% after the module is integrated. The performance of all models was significantly improved through the introduction of attention mechanisms and hyperparameter tuning, highlighting the synergistic potential of data science and computer vision in developing robust and efficient in facial expression recognition systems.

Keywords—Data science; computer vision; deep learning; facial expression recognition

I. INTRODUCTION

Deep learning has emerged as a revolutionary and transformative technology within artificial intelligence. Particularly in facial expression recognition (FER), artificial intelligence (AI) applications introduce new research opportunities and significantly advance the field. Facial expressions stem from the coordinated movements of facial muscles in response to emotions [2]. Emotions can temporarily change the shape of the face because of changes in the movement of facial muscles since facial muscles are not independent of each other [8].

Researchers have tried many ways to interpret and decode facial expressions and extract important features from facial images [27]. A person's emotions can influence the efficacy of face recognition, as varying facial expressions can affect the outcomes. Kim S and Kim H found a certain relationship between facial Action Coding Units (AUs) and Emotion labels in the FER dataset [40]. Being a primary means of expressing

human emotions, facial expressions are crucial for social interaction. They transmit non-verbal signals interpreted by the brain, which can be recorded in images or videos [3]. The human brain can automatically recognize emotions without delay [6]. Cha et al. proposed a FEMG-based FER system based on the Riemannian manifold approach, and further develops an online FER system that can make an avatar's expression reflect the user's facial expression in real time, thus demonstrating that our FER system can potentially be used for practical interactive VR applications such as social VR networks, intelligent education, and virtual training [15]. However, this is a challenging task for computers [50]. As AI technology progresses, machines are increasingly able to replicate the functions of the human brain, making FER applicable across diverse domains, like security surveillance or mental health evaluations. For instance, Dong et al studied that the CGSSNet network established based on the DenseNet algorithm has significant advantages in glioma MRI image segmentation, providing a new idea for the diagnosis and treatment of glioma [21]. Automated FER can identify clinically significant facial features, distinguishing disease states and serving as specific biomarkers [41] [60]. Li et al. (2018) proposed a computer-aided framework for the early differential diagnosis of pancreatic cysts. DenseNet learned advanced features from the entire abnormal pancreas, mapped the appearance of medical imaging with different pathological types of pancreatic cysts, and integrated the significance map into the framework. In a cohort of 206 patients with 4 pathologically confirmed pancreatic cyst subtypes, the overall accuracy rate was 72.8%, significantly higher than the baseline accuracy rate of 48.1% [43]. Like the popularity of smartphones and social platforms, FER has become more important in daily life. For example, analysis of users' facial reactions can improve user experience and personalized content recommendations [9]. To create a more immersive VR social interaction, users can wear a head-mounted display (HMD) with RGB cameras that continuously capture images of their lips to interpret facial expressions. Another example in the field of security monitoring is an accurate FER system can assist in identifying suspicious behavior or emotional abnormalities. Liu and Fang designed a three-level cascade algorithm model for expression recognition in educational robots. By using CK+ and Oulu-CASIA expression recognition database, compared with other common cascaded convolutional neural network methods, the accuracy and speed of facial expression recognition are significantly improved [48]. These technologies assist scientists in accurately identifying unethical behaviors or emotions from facial cues and

predicting future behaviors and emotional states based on collected data [4]. The book recommendation system integrates expression and face recognition with tracking book browsing times to determine users' ages and suggest books accordingly [63].

Despite its potential, achieving high accuracy in FER remains challenging due to the inherent complexity and variability of factored expressions [13]. Deep learning can automatically extract people's facial features, identify different expressions, and meet the expected requirements of classification. Face feature detection and recognition and convolutional neural network classification. The advantage of facial markers is that classification is very robust, even with limited memory [28]. However, there are still some limitations in the performance of existing deep learning models in facial expression recognition tasks, such as the generalization ability of the model, the ability to capture different facial details, and the performance in the case of unbalanced data. Therefore, identifying and proposing the most effective deep learning model is of great significance for FER. First, the most effective models can significantly improve the accuracy of FER, help better understand and analyze human emotions, and be applied to many fields, such as human-computer interaction, emotional computing, and mental health monitoring. Second, by exploring and comparing different deep learning architectures, especially convolutional neural networks (CNNs), it is possible to discover which specific network structures and feature extraction methods perform best in FER tasks, guiding future research and applications. In addition, the most effective models can achieve efficient and accurate facial expression recognition in the case of limited resources, thus reducing computational costs and improving the practicality and scalability of the system. Therefore, this study aims to explore various deep learning architectures and identify the best-performing FER models by testing a widely used benchmark dataset in the field, providing a diverse set of facial images and their corresponding emotional labels. This will help solve the challenges that exist in FER and drive the development and application of this field. Therefore, the first question of this study is, is it possible to identify the best-performing FER model on an FER dataset by exploring various deep learning architectures, especially convolutional neural networks (CNNs)?

Attention mechanisms show great promise in improving the performance of deep learning models by focusing on relevant features while suppressing irrelevant features [23]. However, there are still some limitations in the performance of existing deep learning models in facial expression recognition tasks, such as insufficient ability to capture subtle facial features and low computational efficiency. Introducing attention mechanisms, such as convolutional block Attention modules (CBAM), can somewhat alleviate these problems. CBAM effectively enhances the model's feature extraction capability by combining channel and spatial attention while keeping the computational overhead low. In 2022, Ju and Zhao combined attention mechanisms to propose a new masked attention mechanism Parallel Network (MAPNet), which significantly improved the classification performance and accuracy of three different datasets of the FER task RAFDB, AffectNet and FEDRO by 0.001, 0.0118 and 0.0325, respectively [33]. In 2023, Putro et al.

proposed a real-time facial expression classification method based on a dual attention module convolutional neural network, which achieved an excellent result of 0.9865 and 0.9688 in CK+ and JAFFE datasets, respectively [57]. However, introducing attention mechanisms in models with different structures is time-consuming. To solve this problem, Sanghyun et al. designed a simple and efficient feedforward Convolutional neural network attention module (CBAM) that can be seamlessly integrated into any CNN architecture with negligible overhead. Thus, it provides new ideas and methods for combining deep learning and attention mechanisms [55]. This study aims to integrate CBAM into each layer of a deep learning model to investigate its impact on the performance and efficiency of deep learning models in FER tasks. By introducing CBAM, we aim to significantly enhance deep learning models' feature extraction and discrimination capabilities, thereby improving FER tasks' overall performance and efficiency. This will contribute to developing more accurate and efficient FER systems and provide valuable experience and methods for future research and applications [22]. Therefore, this study raises a second question: Does the introduction of CBAM in deep learning models affect the performance and efficiency of FER tasks? By systematically testing and validating the impact of CBAM, we expect to provide better solutions and new research directions for the FER field.

Hyperparameter tuning is a key aspect of optimizing deep learning models' performance [1]. The performance of existing deep learning models in facial expression recognition (FER) tasks is often affected by the selection of hyperparameters, such as learning rate, batch size and regularization techniques. These hyperparameters directly affect the training process and final performance of the model. However, it is still a challenging task to select the optimal combination of hyperparameters to achieve the model's best performance and generalization ability. By systematically adjusting these hyperparameters, the predictive power of FER deep learning models on FER datasets can be significantly enhanced. Reasonable hyperparameter Settings can not only improve the accuracy of the model but also effectively reduce the overfitting phenomenon, thus improving the generalization ability of the model [59]. Rigorous experiments and evaluations are performed during training to determine an optimal set of hyperparameter combinations that maximizes model performance, minimizes overfitting, and improves generalization. This study aims to explore and verify the effects of different hyperparameter configurations on the performance of the FER deep learning model through hyperparameter tuning. Hyperparameter tuning is important because it can significantly improve the model's training effect and practical application performance, thus achieving the most advanced performance in the FER task. By determining the best combination of hyperparameters, best practices can be established for deep learning model training of facial emotion recognition, and scientific basis and methods can be provided [11]. Therefore, the third question in this study is: Can hyperparameter tuning maximize model performance and improve generalization? Through the hyperparameter tuning and verification of the system, we expect to provide better solutions and new research directions for the FER field.

This study aims to achieve the following three objectives:

- To recommend the most effective deep learning models for facial emotion recognition (FER) utilizing the FER2013 dataset.
- To propose the attention mechanism, CBAM (Convolutional Block Attention Module) is added to each model layer to explore the differences in model performance and efficiency on the same dataset.
- Enhance the performance of deep learning models through hyperparameter tuning during the training phase, thereby optimizing their predictive capacity for facial emotion recognition.

In this study, we propose an enhanced facial expression recognition (FER) model based on ResNet50 with a Convolutional Block Attention Module (CBAM). FER is a challenging task due to high intra-class variability, subtle interclass differences, and the presence of occlusions and noise. ResNet50 serves as a robust backbone for feature extraction, while CBAM enhances the network's ability to focus on both spatially and channel-wise relevant features. This combination allows the model to address the shortcomings of existing methods by improving feature localization and discriminability with minimal computational overhead."

II. LITERATURE REVIEW

A. Facial Emotion Recognition

Hardware technology development has addressed the significant computational power issues associated with deep learning due to its complexity, power requirements, and relatively low cost [5]. CNN has emerged as one of the most revolutionary technologies for FER. CNNs can learn from large datasets to automatically extract and combine features necessary for recognizing various expressions. CNNs build an understanding of complex expressions by abstracting image features layer by layer through a multi-layered structure. This layer-by-layer approach to learning is well suited to the FER task because it allows the model to recognize and distinguish subtle differences in expression.

In recent years, many derivative models of CNN and classification methods for facial recognition have been developed, such as data set preprocessing and feature extraction methods. Chen et al introduced an additional branch to generate a mask, thus focusing on the movement area of the facial muscles. In order to guide face learning, we propose to combine prior domain knowledge and use the average difference between neutral faces and corresponding facial faces as training guidance, which is effective compared with the most advanced methods [18]. Jia In the first stage, offline subnetworks were trained in three subnetworks to achieve convergence (the three subnetworks are AlexNet, VGGNet and ResNet derivative). In the second stage, the output layer of these three subnetworks was removed and predicted by SVM, and the accuracy rate reached 0.7127[26]. Liu investigated the improved VGG-16 CNN, enhancing the VGG-16 network by optimizing the third and fourth convolutional layers [39]. Instead of the original SoftMax classifier, a 7label SoftMax classifier was employed. It replaced the original ReLU activation function with LeakyReLU. Experiments on the FER2013 showed an accuracy of 0.7242,

higher than the previous rates of 0.6631 and 0.7138 without improvements to VGG and ResNet, respectively. Part of FER datasets are shown in Fig. 1:

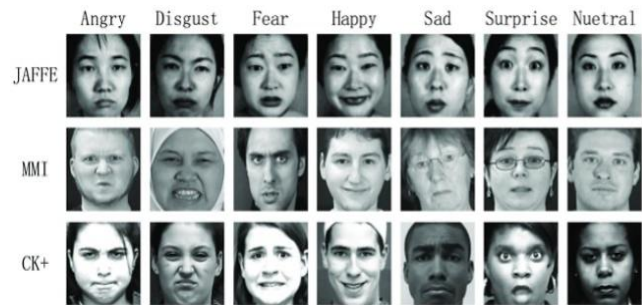


Fig. 1. Presentation of different FER datasets (Part).

To simplify the artificial feature extraction process in traditional FER and capture more diverse features, Changing proposed a method that integrates multiple CNN models, using three different CNN subnetworks for comprehensive prediction. This approach achieved an accuracy of 0.701 on the FER2013 [12]. Dwijayanti et al. indicated that there was not much research on face recognition and FER objectives; consequently, they employed CNNs to tackle this challenge [14]. Unlike other experiments, they used the original image as CNN input and directly used the VGG-f model for FER tasks, overcoming the problems of underfitting and overfitting the CNN framework and also getting a good performance.

Fu investigated the impact of incorporating visual attention mechanisms into deep learning for FER [17]. In their approach, three fully connected layers in the training phase were substituted with three convolutional layers to generate test results for the entire network, thereby mitigating the limitations of full connection. Additionally, the SE block was applied to normal VGG, and the results verified the effectiveness of SEVGGNET with 0.668 accuracy. Das and Neelima proposed an improved pretreatment stage. This includes extracting local binary features used to express classification. These feature vectors are connected and used in shallow neural networks with minimal complexity and fewer layers to optimize expression recognition processes as well as gently enhance decision trees. Applying this local binary feature-based neural network (LBF-NN) approach to three different popular databases, more than 93% results were achieved, even when compared to a variety of complex and advanced algorithms [20].

Mohamed et al. achieved improvements by applying CNN to examine the Alex network architecture, applying transfer learning methods and modifying the full connection layer using support vector machine (SVM) classifiers [53]. The improved model has a classification recognition rate of about 0.6429 for the selected expressions. The system has achieved satisfactory results on the ice-MEFED dataset. The improved model has a classification recognition rate of about 0.6429 for the selected expressions. Lee et al. proposed an ensemble framework to boost the reliability of FER models using three models: VGG16, InceptionResNetV2, and EfficientNetB0 [33]. The results indicated that the model recognition accuracy priority edge ensemble learning algorithm improved by 0.0281.

B. Deep Learning

Facial expression (FE) has powerful potential and is a universal human communication form closely linked to mental states, attitudes, and intentions. By analyzing facial expressions displayed by humans or objects, computers can effectively process and interpret human emotions, forming the core of the FER system. The development of FER in this field has witnessed the transformation from the preliminary geometric feature method to the current deep learning technology, which has profoundly affected our understanding and practice of emotional computing and human-computer interaction. Early FER studies relied on manual extraction of facial features and simple pattern-matching techniques. Researchers try to identify expressions by pinpointing key features. However, these methods are insufficient when faced with the diversity and complexity of human expressions and struggle to adapt to dynamic real-world environments.

CNN are potent visual recognition tools whose design is inspired by biological vision systems. It is mainly used for image processing and classification [6]. Compared to other classification algorithms, it is an algorithm that takes images and is able to distinguish one from another with minimal preprocessing. Automatic feature detection without human supervision is the main advantage of CNN over other algorithms. In addition, a method combining global appearance features with local geometry features is proposed [11]. Specifically, they provide not only the raw images to the facial expression recognition network but also the facial markers associated with them. A typical CNN comprises various layers: convolutional layers, activation functions, pooling layers, and fully connected layers.

Transfer Learning is an important method in the field of deep learning. In deep learning, transfer learning usually involves taking a pre-trained model on a large data set, like ImageNet, and applying it to a new, related task. The advantage of transfer learning is that it can leverage the complex feature extraction capabilities that have been learned. The core idea of transfer learning is that there is a commonality between certain learning tasks so that what is learned on one task can be reused on another. For example, a model trained on pictures of animals may have learned to recognize features such as eyes, ears, etc., which may also be useful for recognizing other types of objects, such as faces[49]. Lee et al. applied transfer learning, fine-tuning, and data enhancement to the training and validation of the Facial expression recognition 2013 (FER-2013) dataset. Experimental results show that the model recognition accuracy of the proposed priority edge ensemble learning algorithm is improved by 2.81% [42].

Attention Mechanism allows the model to prioritize significant parts of the input data by assigning variable weights to different image regions. This reflects the importance of these regions for the final task. For instance, features such as the eyes and mouth may be more recognizable in FER tasks than in other parts. Selective weight assignment enables the model to concentrate on particular input sections while disregarding others, thereby achieving the intended output. This is known as hard attention, or it can be incorporated into the model in a differentiable way, allowing the entire network to be trained using techniques such as gradient descent. For example, during

the COVID-19 pandemic, many people wore masks, and when faces are partially covered or affected by interference factors like large pose changes, it hampers feature extraction and reduces FER performance. CBAM is a kind of attention mechanism in DL that is employed to improve CNN's feature representation capacity. By explicitly modelling images' spatial and channel dimensions, the network can focus on key areas, thereby improving its performance. CBAM can be regarded as a lightweight plug-in that is easy to integrate into the existing CNN architecture.

C. Attention Mechanism

Jin et al (2022) have introduced Transformer encoder to model the remote dependency between different facial areas and capture the global relationship between different facial units, complementing the spatial locality of CNN [36]. But the attention mechanism mimics human attention, allowing the model to prioritize significant parts of the input data by assigning variable weights to different image regions. This reflects the importance of these regions for the final task. For instance, in FER tasks features such as the eyes and mouth may be more recognizable than other parts. Selective weight assignment enables the model to concentrate on particular input sections while disregarding others, thereby achieving the intended output. Liu et al proposes an adaptive multi-layer perceptual attention network that extracts global, local, and significant facial emotional features using different fine-grained features to understand the potential diversity and key information of facial emotions [46]. This is known as hard attention, or it can be incorporated into the model in a differentiable way, allowing the entire network to be trained using techniques such as gradient descent. For example, during the COVID19 pandemic, many people wore masks, and when faces are partially covered or affected by interference factors like large pose changes, it hampers feature extraction and reduces FER performance. The flow diagram of the attention mechanism on FER is shown in Fig. 2.

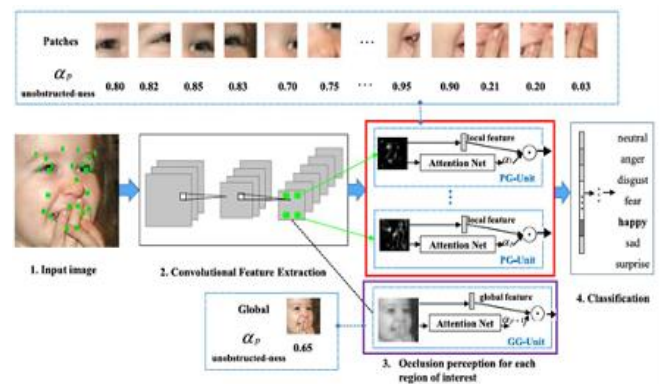


Fig. 2. Diagram of attention mechanism example in FER.

Attention mechanisms enable the model to dynamically focus on the most relevant parts of the input for a given task. For instance: In vision tasks, attention can focus on important regions of an image (e.g., objects or edges). In text processing, attention identifies key words or phrases that are crucial for understanding the context. This focus helps the model prioritize meaningful information while ignoring less relevant or redundant features. Attention mechanisms enhance feature

extraction by weighting input features according to their importance. These weights are computed adaptively during training, allowing the model to learn a richer, context dependent representation of the data. For example, in transformers, attention layers allow the model to learn contextual relationships between different parts of the input, which is critical for tasks like language translation or image captioning.

The Convolutional Block Attention Module is a module that combines the channel attention mechanism and spatial attention mechanism, aiming to improve the feature expression ability of convolutional neural networks. It is also an attention mechanism in deep learning. By explicitly modelling the spatial and channel dimensions of the image, the network can selectively focus on key areas, thereby improving its performance. Its authors Woo et al. (2018) indicated that its flexibility and versatility can be applied to different CNN network architectures (such as VGG, ResNet, etc.), and it can show good adaptability and versatility in different tasks, such as image classification semantic segmentation, and object detection. The CBAM structure is relatively simple and can be seen as a lightweight plug-in. It is easy to integrate into the existing CNN architecture as a module enhancement, disorder greatly modify the original network structure, and finally, through the channel-by-channel and pixel-by-pixel weighted way, improve the model’s attention to important features, thus improving the training and reasoning efficiency, especially in the processing of complex tasks. The specific architecture of CBAM is shown in Fig. 3:

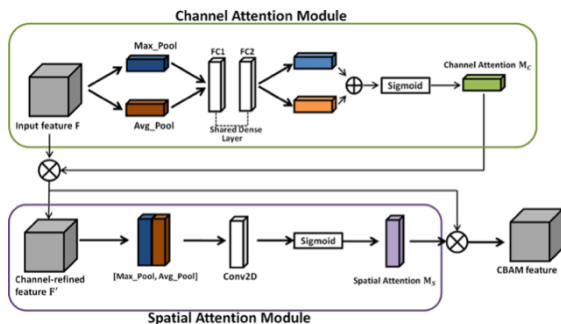


Fig. 3. Diagram of attention mechanism example in CBAM.

D. Deep Learning Models

Deep learning has made remarkable progress in the field of FER, mainly reflected in automatic feature extraction, efficient processing of large-scale data, nonlinear modelling ability, end-to-end training, combining attention mechanisms, etc. Tang (2013), this study shows for the first time that deep learning can automatically extract multi-level features without manual feature design compared to traditional machine learning methods such as SVM and KNN, which not only simplifies the process of special engineering but also significantly improves recognition accuracy. To demonstrate the excellent performance of deep convolutional neural networks on large-scale datasets, Hinton et al. (2012) found that CNN significantly improved the classification accuracy of image tasks through training on ImageNet datasets. Zhang (2017), through practice training and multi-task learning, the study et al. shows that the original image is directly learned to the final classification without the need for intermediate step feature extraction and selection, significantly improving the performance in complex scenes. After Woo et al.

proposed CBAM, the parameters added to the deep learning model did not increase significantly, but the average accuracy of VGG16 and MobileNet increased by 0.0015 and 0.0024, respectively. Jin et al. (2023) designed an image enhancement algorithm using Super resolution generative adversarial (SRGAN) and adaptive gray normalization (AGN) based on the data sets and characteristics of convolutional neural networks, and tested the Fer2013 data set, and the accuracy was improved from 68.03% to 70.04% [37].

TABLE I. COMPARISON OF DIFFERENT DEEP LEARNING ALGORITHM STRUCTURES BASED ON FER-2013

Author	Model	Accuracy (%)	Year
Rajesh Kumar [38]	CNN EmotionNet	67 66.71	2023
Xu & Zhao [64]	AlexNet OneNet	64.29 54.29	2020
Putro et al. [57]	VGG13 ResNet	73.03 72.4	2020
Lu et al. [47]	VGG Inception ResNet	72.7 71.6 72.4	2023
	Ensemble CNN	75.2	
Pramerdorfer & Kampel et al. [56]	Inception ResNet	71.6 72.4	2016
Sahoo et al. [58]	6-layer CNN 10-layer CNN	66.67 68.34	2023
	VGG-16	63.68	
Lee et al. [39]	Fine-Tune VGG16 InceptionResNetV2 Fine-Tune EfficientNetB0	66.65 67.71 67.46	2022
	Priority Ensemble CNN Algorithm	70.52	
Chen et al. [16]	FERW	71	2018
Jia et al. [32]	Ensemble CNN	71.27	2020
Joseph et al. [35]	CNN	67.18	2022
Zhang et al. [36]	LeNet-5/VGG-16	70.1	2023
Liu[45]	VGGNet	71.42	2023
Muhamad et al. [51]	CNN	54	2021
Meena et al. [49]	InceptionV3	73.09	2023
Alexeevskaya et al. [10]	CNN	60.54	2022
Fu [24]	VGG16 VGG16+SENet MobileNet	65 66.8 68.03	2022
Jin et al. [34]	SRGAN-MobileNet AGN-MobileNet SRGAN+AGN-MobileNet	69.07 68.92 70.04	2023

Generally, complex preprocessing technology and data enhancement methods are helpful to improve the accuracy of the model. From the table, we find that the VGG model performs well on multiple data sets, such as FER+ up to 0.806, FER2013 up to 0.7303 and CK+ up to 0.8875, indicating that VGG has been widely used in different studies with stable performance. Different researchers have adopted a variety of preprocessing techniques, such as image cropping and adaptation Strong sex; The ResNet model has shown good accuracy on multiple data sets in the table, such as 0.724 accuracy on FER-2013 data set and 0.8726 accuracy on RAF-DB data set. ResNet solves the

problem of gradient disappearance in deep networks through residual connection. Performance is superior, but different data enhancement methods significantly impact ResNet's performance and require careful adjustment. The Inception model has an accuracy of 0.7309 on the FER-2013 dataset and 0.727 on the CK+ dataset. The Inception model can capture multiscale features and enhance feature expression ability through convolution kernels of different sizes. chuanjie et al proposed a facial expression recognition method that integrates multiple convolutional neural network models and uses three different CNN subnetwork models for comprehensive prediction. Experiments show that the recognition accuracy of this method on FER2013 and CK+ datasets is 70.1% and 94.9%, respectively [19].

Compared with other models, although the traditional CNN model has a simple structure and low computing resource requirements, its accuracy is generally low, for example, only 0.67 on the FER-2013 dataset. The three models, EmotionNet, AlexNet, and OneNet, perform well on specific data sets, but their overall performance is inferior to VGG, ResNet, and Inception, and they are larger than that of specific preprocessing techniques. Other models, such as VGG-f, AMP-Net, and AFTransformer, perform well in specific application scenarios and data sets. For example, the accuracy rate of AMP-Net on the RAF-DB dataset is 0.8925, but the model is relatively special and has low universality. Therefore, compared with other models, VGG, ResNet and Inception have significant advantages in accuracy and adaptability. All three models performed better in the table than most others, demonstrating their strong capabilities in the FER task. When selecting a specific model, you can make trade-offs and choices based on computing resources, training time, and application scenarios. From the literature review, we can see from Table I that different models have different performances in different data sets, and preprocessing technology significantly impacts the mode's performance.

Existing CNN-based FER models often lack attention mechanisms, treating all features equally, which limits their ability to differentiate subtle expressions or handle occlusions effectively. While attention-based methods improve feature extraction, they are often computationally expensive or focus solely on spatial or channel-wise attention. Our method addresses these limitations by integrating CBAM into ResNet50, providing both spatial and channel-wise attention while maintaining efficiency.

III. METHODOLOGY

A. Research Framework

The deep learning framework can be approached as an optimization problem to identify model parameters that minimize the loss function, following steps from data preprocessing, model construction, training optimization, and performance evaluation. The study is divided into four phases: data understanding, data preparation, modelling, and evaluation. Fig. 4 illustrates the summary of tasks in each phase. The overview of each stage of this study is as follows:

1) *Business understanding*: This phase includes evaluating the current state of the application of deep learning-based

models to FER tasks by reviewing existing publications. The aim is to identify research gaps and set research goals.

2) *Data understanding*: This stage starts with an initial exploration of the data set, involving data collection and distribution checks to grasp the basic features and structure of the data. The goal is to familiarize yourself with the data and spot any potential data quality issues.

3) *Data preparation*: In this stage, raw data needs to be converted into clean data suitable for deep learning development. The third chapter also expounds on this stage. This includes data transformation, data enhancement, data segmentation, and coding.

4) *Modelling*: In this phase, the selection and implementation of modelling techniques will be explained, and the techniques needed to build predictive models will be discussed. Additionally, it involves fine-tuning the model parameters to optimize performance.

5) *Evaluation*: This stage encompasses reviewing the entire development process for the model, assessing the performance of the developed model, and evaluating its stability and validity through various evaluation parameters and statistical tests. It also includes verifying whether the research objectives have been met.

6) *Deployment*: In this phase, the insights gained from developing the FER deep learning model are communicated to the stakeholders. In this study, this stage is limited to presenting the results of developed deep learning models.

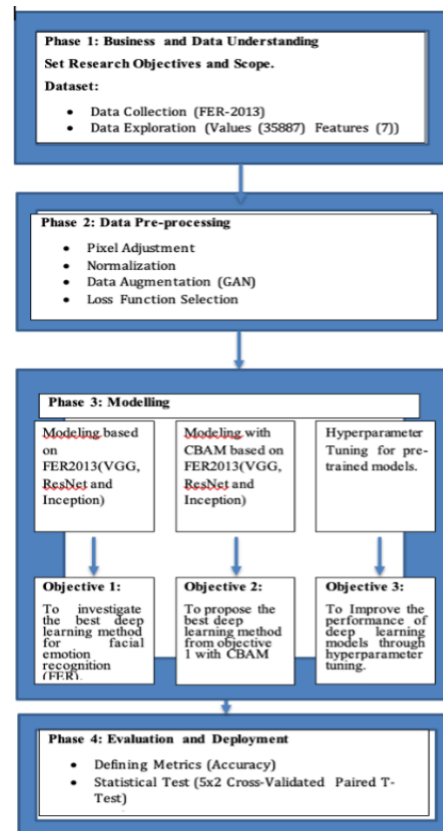


Fig. 4. Research framework.

B. Phase 1: Business and Data Understanding

The Business understanding stage is the initial stage of this study, which aims to understand the current research status of FER tasks and the application of deep learning in this field. At this stage, research objectives are developed based on research gaps identified through a comprehensive literature review. The research objective lays the foundation for the subsequent stage of this study. By establishing clear research objectives, this study ensures a targeted exploration of the application of deep learning models to FER tasks, particularly the classification prediction of FER tasks using transfer learning methods and combining attention mechanisms.

Current FER databases usually include a small number of subjects and provide only a few sample images for each expression. They often have a limited variety of subjects or minimal differences between groups, making FER tasks in real-world scenarios more challenging [10]. As shown in Table II, FER-2013 (Facial Expression Recognition 2013) is a publicly available dataset for FER that includes a wide range of expressions, from happy to sad to surprised. The entire dataset consisted of about 32,298 grayscale 48x48 pixel faces, each labelled with an emotion, such as happy, sad, or angry. These images are all from the web, uploaded by different people, and then there are artificial intelligence helpers, platforms like Amazon’s Mechanical Turk, to label the facial expressions appropriately. For machine learning or computer vision researchers, FER-2013 can be used to explore differences in emotional expression in different cultural contexts. These faces from all over the world provide rich materials for the study of cross-cultural emotional communication.

They serve as a benchmark assessment for the performance of FER algorithms. Its advantage is that all images are preprocessed and are uniformly 48x48 pixels and each image is labelled, which is why it has become the standard for comparing the FER algorithm. However, limitations also exist because all images are grey and lack binary colour information, which may limit the features the model learns, and the images that are collected from the Internet may not fully reflect the natural state of people’s expressions in real-life scenarios. Moreover, the dataset may lack diversity regarding race, age, and background. Another significant issue is the imbalance in the dataset; some categories have substantially more samples than others. For instance, the happy category contains 8989 samples, far exceeding the 547 samples in the Disgust category.

C. Phase 2: Data Preprocessing

This study mainly tests three deep learning models: VGG, ResNet and Inception. They are all CNN-derived models, and the models are optimized by adjusting parameters and hyperparameters. Data preprocessing is a key step in ML and data analysis, aiming to convert original data into a suitable form for further analysis and modelling. The usual steps of the FER system are to preprocess the image, extract the features from the preprocessed image, and classify the extracted features [25]. Fig. 5 and Fig. 6 show the imbalance between the test set and the training set of the FER2013. Table III shows the number of data sets after the planned data enhancement.

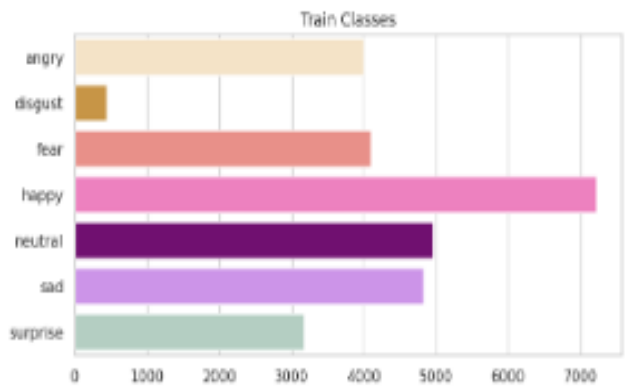


Fig. 5. Bar chart of FER-2013 trainset.

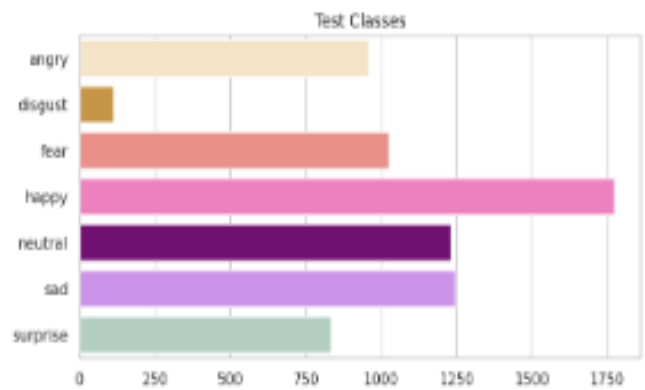


Fig. 6. Bar chart of FER-2013 testset.

TABLE II. DESCRIPTION OF FER-2013

Attribute ID	Attribute Name	Attribute Testset	Attribute Trainset	In Total
0	Angry	958	3995	4953
1	Disgust	111	436	547
2	Fear	1024	4097	5121
3	Happy	1774	7215	8989
4	Sad	1247	4830	6077
5	Surprise	831	3171	5002
6	Neutral	1233	4965	6198
In Total		3589	28709	32298

TABLE III. AUGMENTATION COUNTS PER CLASS

Name of Class	Augmentation Counts
Angry	1675
Disgust	5234
Fear	1573
Happy	0
Neutral	705
Sad	840
Surprise	2499
Total Amount after Augmentation	41235

Image Processing: Different deep learning models have different requirements for the input of images, and adjusting the input images to a uniform size ensures that the model can handle them indiscriminately [7]. If the image is saved at a larger size, it means more pixels and higher computational complexity. By adjusting the image size, enough information can be retained while reducing the need for computational resources, and in some cases, adjustment Size helps models better focus on key features of facial expressions rather than other irrelevant parts of the image. I set a standard for image input that we call the standard form of input, and we require the image to always be in the standard form [44]. While my test models usually require larger input sizes, for example, VGG and ResNet require 224 x 224, Inception requires 299 x 299, so the image needs to be adjusted to the desired size before entering the model without destroying the aspect ratio of the image, so as not to affect the model representation. The resized in FER2013 dataset is shown in Fig. 7:



Fig. 7. Part resized images comparison of the dataset FER-2013.

Normalization: Normalization is a process in data preprocessing which is used to change the range of numerical data so that it is located in a specific cell, such as [0,1] or [1,1]. In image processing, normalization is a common practice. The essence of the method is some layer input data of the neural network that is preprocessed with zero mathematical expectations and unit variance with the intention of improving the stability and efficiency of the training process [4]. For FER tasks, normalization can give different features similar to ranges. Unnormalized data may lead to unstable gradient problems during model training. In deep learning models, normalized data may lead to a gradient that is too large or too small, thus affecting the learning effect of the model. The normalization in FER2013 dataset is shown in Fig. 8:



Fig. 8. Part normalized images comparison of the dataset FER-2013.

Feature Extraction: Feature extraction is an important factor in determining the recognition result. Some of the environmental and pose issues that need to be addressed in an image containing a complete face [61]. If the features are not good, even the best classifiers will not get the best results. In most cases, feature extraction produces a large number of features [35]. In this paper, we compare the performance of the model before and after the introduction of the attention mechanism, and for the FER task, features such as the eyes and mouth may be easier to identify than other parts. When assigning weights, we can selectively focus on certain parts of the input and directly ignore others to get the output we want most, which is called hard attention, or it can be integrated into the model in a differentiable way that allows end-to-end training of the entire network using standard techniques like gradient descent. For example, during the COVID-19 epidemic, many people are wearing masks. When the face is partially covered or interfered with [30], such as large pose changes, it can hinder useful feature extraction and greatly reduce the performance of FER predictions. The feature extraction in FER2013 dataset is shown in Fig. 9:



Fig. 9. Part feature extraction images comparison of the dataset FER-2013.

Data Augmentation: Data Augmentation is a technique to generate new, modified data points by transforming some original data columns. In the small-scale deep model data set, the deep model is redundant, complex, and easy to overfit. To solve the redundancy problem, data enhancement techniques were used to extend the original dataset [34]. Data enhancement will enable the model to introduce more variables during training, helping the model learn more generalized features and thus perform better on previously unseen data. In this paper, there is a data imbalance in the FER2013 dataset. The generation of data is enhanced using generative adversarial networks (GANs), the core idea of which is based on an adversarial process in which two networks - generator and discriminator - compete against each other [51]. For example, Hu et al (2019) used GAN to generate reference expressions and compared them with original expressions to generate differential features, avoiding interference of irrelevant information on expression recognition [31]. The generator takes random noise as input and outputs as real data as possible, such as high resolution images. In contrast, the discriminator takes real data or the data generated by the generator as input and outputs the probability of the data being true or false to distinguish the real data

generated by the generator from the false data. A generated discriminant representation can be obtained by separating and interpolating different expressions in a face image [62]. The learned representations not only generate more training samples of unpaired input images but also contribute to better FER performance. So, we involve generators and discriminators in GAN training, where generators try to generate more and more real data, and discriminators better distinguish between real and fake data. The most common form is the minimax game given by the following formula:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_Z(z)} [\log(1 - D(G(z)))] \quad (1)$$

Where: $\mathbb{E}_{x \sim p_{data}(x)} [\log D(x)]$ represents the effect of real data x on discriminator D .

$\mathbb{E}_{z \sim p_Z(z)} [\log(1 - D(G(z)))]$ represents the effect of the data generated by the generator G on the discriminator D .

Cross-entropy loss effectively measures the disparity between the model's output probability distribution and the actual label distribution. The cross-entropy loss function usually has the following form:

$$L = \sum_{i=0}^c y_i \log(\hat{y}_i) \quad (2)$$

Where: the C is the number of classes (for FER-2013, it is 7, representing 7 basic emotions), y_i is the one-hot encoding of the real tag, \hat{y}_i is the predicted probability for category i . Hence, a regularization term is added to encourage model to adopt smaller weights, thus reducing the model's complexity. L1-regularization: L1-regularization sums the absolute values of weights and tends to produce sparse weight matrices, which is conducive to feature selection.

$$L1 = \lambda \sum |\omega| \quad (3)$$

L2 Regularization: L2 regularization sums the squares of the weights, tends to uniformly assign errors, and is often used to prevent neural networks from overfitting.

$$L2 = \lambda \sum \omega^2 \quad (4)$$

Where ω indicates the weight of the model, λ is the regularization coefficient.

D. Phase 3: The Development of Algorithms and Models

Transfer learning is commonly training to assign specific weights to a pre-trained model and then train it with the dataset. Rajesh Kumar, C.G. Patil et al. and Sahoo et al. which usually perform better than statistical and traditional machine learning algorithms[32], [56], [54]. In addition, Lu et al., Martin Kampel et al. and Yichen Liu, compared to deep learning models, show superior performance [39], [45], [52]. Therefore, VGG19, ResNet50 and Inception V3 were used in this study to develop facial expression recognition models. Using the TensorFlow software library developed by the Google Brain team to develop the training model, the following sections outline the proposed architectures for the VGG19, ResNet50, and Inception models, including the model architecture with the addition of CBAM.

VGG19: The VGG model adopted in this study is VGG19, where 19 indicates that there are 19 learnable layers in the

network, among which the convolutional layer uses multiple small dimensions (3×3). Use ReLU as the activation function between convolutional layers to increase nonlinearity. After each convolutional layer, use the maximum pooling layer (2×2). The network's top consists of three FC layers, with two layers comprising 4096 units each and the final FC layer matching the number of target categories. In this study, the target for the FER2013 dataset classification is seven; thus, the final fully connected layer is configured for seven categories. The output layer employs Softmax activation functions to convert the outputs into probability distributions as shown in Fig. 10:

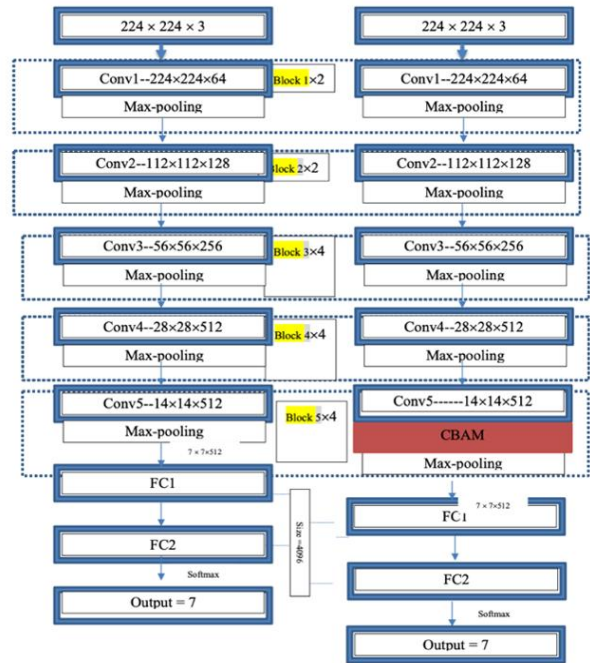


Fig. 10. Structure diagram of VGG19 and add CBAM.

ResNet50: ResNet model used in this study is ResNet50, which is a variant of the residual network, where 50 in ResNet50 refers to a weight layer in which the network contains 50 layers deep. Residual connections allow the inputs of the network to skip directly over one or more layers by adding outputs to the layer, which helps solve the problem of disappearing gradients in deeper networks. During training, this structure allows the gradient to flow directly over the jump connection, increasing the speed and effectiveness of training. The architecture starts with a 7×7 convolutional layer with a stride of 2, which is followed by a 3×3 max pooling layer with a stride of 2. The main component consists of several residual blocks, each containing multiple convolutional layers. In ResNet50, these residual blocks typically have three different configurations (different number of convolution layers and convolution kernel sizes) and are repeated multiple times. When the feature map's dimension needs modification, a convolution with a specific stride length is used to downsample the residual block input, and the number of channels is adjusted accordingly to match the output. Towards the end of the network, a global average pooling layer is employed instead of the conventional fully connected layer, thereby reducing the model's parameter count. As shown in Fig. 11:

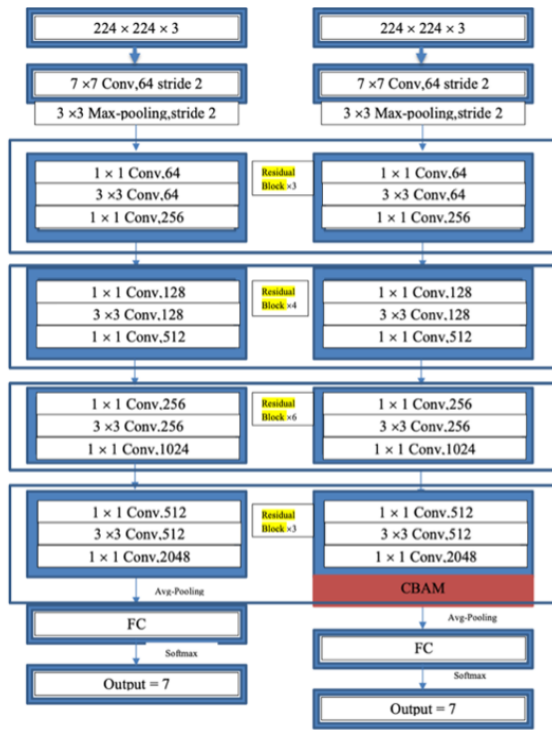


Fig. 11. Structure diagram of ResNet50 and add CBAM.

Inception V3: The Inception model used in this study is Inception V3. The core feature of Inception V3 is its “modular” design, which constructs the entire network through different modular building blocks. Inception V3 optimizes the original Inception module, for example, by introducing the concept of “factorization into smaller convolutions” to decompose large convolution kernels (e.g. 5x5) into smaller continuous convolution operations (e.g., two 3x3 convolution operations). An auxiliary classifier is added to the network as an output of the middle layer, which aids in gradient flow, provides additional regularization, and prevents overfitting during deep network training. At the end of the network, an overall average pooling layer takes the place of the conventional fully connected layer. The specific structure of Inception V3 used in this paper is depicted in Fig. 12:

Hyperparameter Tuning: Unlike the model training process, hyperparameter tuning involves finding the optimal set of hyperparameter values to maximize a performance metric on unseen data. This study uses the hyperparameter optimization library method and the grid search for hyperparameter tuning. In the hyperparameter optimization library, the scope and objective function of hyperparameter search is defined so that it accepts hyperparameters as input and returns the performance index of the model on the verification set (such as loss rate or accuracy rate) as the optimization target. By systematically searching the hyperparameter space, the hyperparameter combination that optimizes the model performance is found. For these models of deep learning, the hyperparameters tuned include learning_rate, batch size, regularization coefficient, activation function and epoch. Table IV are the details of each hyperparameter tuned in this study:

1) *Learning rate*: This determines how quickly the model moves in the gradient’s direction or the number of steps taken. If the learning rate is excessively high, the optimizer may overshoot the minimum, preventing convergence. Conversely, a low learning rate makes the optimization process slow and may remain stuck at a suboptimal local minimum.

2) *Batch size*: This denotes the number of samples handled in a single forward and backward pass through the neural network. The batch size affects the optimization’s efficiency and speed. A larger batch size improves memory utilization and makes the gradient descent direction more stable.

3) *Regularization coefficient*: This refers to the realization of a minimization strategy in which penalty terms are added to the empirical risk. Typically, it is a function of monotonically increasing model complexity. The more complex the model, the higher the penalty value.

4) *Optimizer*: An algorithm or method used in deep learning to update model parameters to minimize the loss function. The primary goal is to reduce the loss function as much as possible by adjusting the model parameters, ensuring the model fits the training set well and performs effectively on the test set.

5) *Epoch*: Defined as one complete pass of the dataset through the neural network. A single pass is insufficient; the dataset needs to be iterated multiple times to achieve convergence. An iterative method called gradient descent is employed to optimize learning. As the epoch count increases, the weights in the neural network are updated more frequently, transitioning from underfitting to overfitting. The optimal number of epochs varies and should be determined using validation sets or cross-validation.

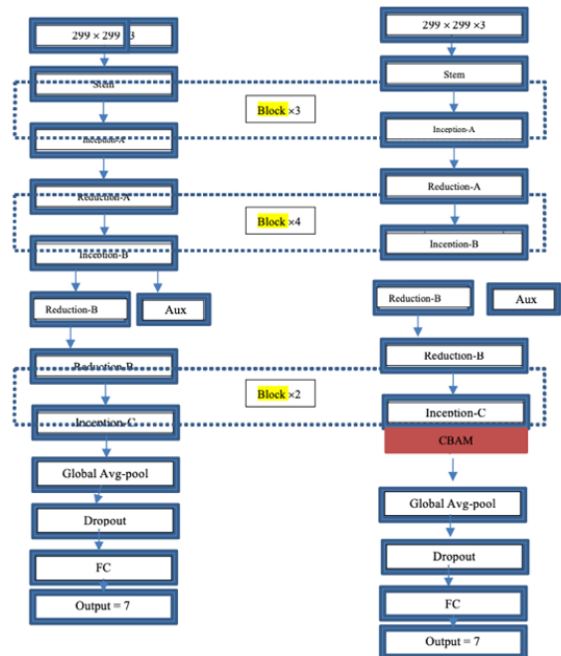


Fig. 12. Structure diagram of InceptionV3 and adds CBAM.

TABLE IV. DEFINITION OF VARIOUS HYPERPARAMETERS FOR DEEP LEARNING MODELS

Hyperparameter	Default Value	Tuning Value
Learning rate	0.01	0.001, 0.005
Batch size	30	60, 120
Regularization coefficient	0.001	0.01, 0.1
Optimizer	SGD	Adam RMSProp
epoch	50	100, 200

E. Phase 4: Evaluation

Key Metrics for Model Performance: A confusion matrix is a tool used to measure performance and is commonly utilized in supervised learning and classification problems. It helps visualize an algorithm's performance, particularly when dealing with two or more classes. The confusion matrix itself does not provide performance metrics like accuracy or precision, but they can be calculated from it. By examining the confusion matrix, we can further analyze the model's performance across different categories and identify its strengths and weaknesses. For example, some expression categories in the FER-2013 dataset may have more samples than others. Metrics such as confusion matrices and precision rates can help identify and address this imbalance, ensuring that the model has good recognition across all categories. The confusion matrix and accuracy rate provide a detailed perspective for diagnosing such issues. The confusion matrix can also reveal the model's tendency to misclassify one category into another, which can be very helpful for further tweaking and optimizing the model.

Statistical Test on Model Performance: After model evaluation, statistical tests are performed to determine which model shows better performance than other models that have been developed. Statistical testing is essential to determine whether a particular model is statistically significantly better than others, a process that can help us understand whether the differences in performance are significant enough to support or guard against assumptions about substantial differences between models. In this paper, the average performance of the best-performing VGG, ResNet and InceptionV3 models and the CBAM models were compared using the T-test, and the holdout method and cross-validation were employed. The statistical test in this study was performed using a 5×2 cross-validated paired t-test. One reason for conducting a 5×2 cross-validated paired T-test is its acceptable likelihood of type I errors. Because the comparison is carried out on the same set of data, the error caused by random variation of the data set is reduced. The 5×2 cross-validation paired T-test enables effective detection of performance differences even with small sample sizes. The 5×2 cross-validation provides ten independent performance evaluations that can be used to compare two models using statistical tests such as paired Ttests. Because the data is re-split each time, there is less correlation between the test results, reducing the estimates' variance. The paired T-test can also better estimate variation for 2-fold cross-validation since the training sets do not overlap, compared to 10-fold cross-validation.

IV. RESULTS AND DISCUSSION

In this section, compare the performance metrics of the developed VGG19, ResNet50, and InceptionV3, with and without the CBAM module. Model efficacy was gauged using traintest segmentation (TTS) and cross-validation (CV) techniques. The dataset was partitioned with a training-to-test ratio of 80:20, and a 10-fold cross-validation was employed. Hyperparameter tuning was executed on six models, each with five distinct hyperparameters, each tested at three varying levels. Consequently, this culminated in 243 potential hyperparameter configurations for each algorithm during the tuning process.

TABLE V. HYPERPARAMETER COMBINATIONS WITH THE HIGHEST TEST ACCURACY FOR VGG19 AND VGG19-CBAM WITH TRAIN-TEST SPLIT

Model	Hyperparameter	Default	Best Value
VGG19	Epoch	50	200
	Batch Size	512	128
	Learning Rate	0.01	0.001
	Regularization Coefficient	0.01	0.001
	Optimizer	Adam	Adam
VGG19-CBAM	Epoch	50	200
	Batch Size	512	128
	Learning Rate	0.01	0.001
	Regularization Coefficient	0.01	0.001
	Optimizer	Adam	Adam

TABLE VI. ACCURACY COMPARISON BETWEEN THE DEFAULT AND OPTIMAL HYPERPARAMETER CONFIGURATION OF VGG19 AND VGG19-CBAM

Model	Hyperparameter Configuration	Test Accuracy
VGG19	Default	0.7040
	Best	0.7170
VGG19-CBAM	Default	0.7104
	Best	0.7190

The hyperparameter adjustment results of the VGG19 and added CBAM models are visualized in Fig. 12 using the mesh search method based on the train-test split. Under the training-test segmentation, the test accuracy ranges from 0.7040 to 0.7104. Table V shows the best hyperparameter configuration of VGG19 based on the highest test accuracy in the case of training test segmentation. The highest test accuracy of training-test segmentation is 0.719, as shown in Table VI. This suggests that both models may have overfitted training datasets and cannot properly generalize to previously unknown data.

The hyperparameter adjustment results of the ResNet50 and added CBAM models are visualized in Fig. 8 using the mesh search method based on the train-test split. The test accuracy ranges from 0.7033 to 0.7124. Table VII shows the best hyperparameter configuration of the ResNet50 model based on the highest test accuracy in the case of training test segmentation. The highest test accuracy of training-test segmentation is 0.724, as shown in Table VIII.

TABLE VII. HYPERPARAMETER COMBINATIONS WITH THE HIGHEST TEST ACCURACY FOR RESNET50 AND RESNET50-CBAM WITH TRAIN-TEST SPLIT

Model	Hyperparameter	Default	Best Value
ResNet50	Epoch	50	200
	Batch Size	512	128
	Learning Rate	0.01	0.01
	Regularization Coefficient	0.01	0.001
	Optimizer	Adam	Sgd
ResNet50-CBAM	Epoch	50	200
	Batch Size	512	128
	Learning Rate	0.01	0.01
	Regularization Coefficient	0.01	0.001
	Optimizer	Adam	Sgd

TABLE VIII. ACCURACY COMPARISON BETWEEN THE DEFAULT AND OPTIMAL HYPERPARAMETER CONFIGURATION OF RESNET50 AND RESNET50-CBAM

Model	Hyperparameter Configuration	Test Accuracy
ResNet50	Default	0.7033
	Best	0.7150
ResNet50-CBAM	Default	0.7124
	Best	0.7240

TABLE IX. HYPERPARAMETER COMBINATIONS WITH THE HIGHEST TEST ACCURACY FOR INCEPTION V3 AND INCEPTION V3-CBAM WITH TRAIN-TEST SPLIT

Model	Hyperparameter	Default	Best Value
Inception V3	Epoch	50	200
	Batch Size	512	128
	Learning Rate	0.01	0.05
	Regularization Coefficient	0.01	0.001
	Optimizer	Adam	Sgd
Inception V3-CBAM	Epoch	50	200
	Batch Size	512	128
	Learning Rate	0.01	0.05
	Regularization Coefficient	0.01	0.001
	Optimizer	Adam	Sgd

The hyperparameter adjustment results of the Inception V3 and added CBAM models are visualized in Fig. 9 using the mesh search method based on the train-test split. The test accuracy ranges from 0.7002 to 0.7070. Table IX shows the best hyperparameter configuration of the ResNet50 model based on the highest test accuracy in the case of training test segmentation. The highest test accuracy of training-test segmentation is 0.711, as shown in Table X.

The analysis indicates that hyperparameter optimization significantly enhances the performance of deep learning models. The incorporation of CBAM further boosts the models' performance, with ResNet50-CBAM showing the most substantial improvement accuracy rate at 0.724. These findings

highlight the importance of both hyperparameter tuning and advanced architectural modifications in achieving optimal model performance for facial expression recognition tasks.

Table XI shows the accuracy comparison of multiple deep learning models on a specific task, including the results of other researchers, providing rich information for analyzing the performance differences of the models. First, AlexNet [60] is an earlier deep-learning model with a relatively low performance on this task, with an accuracy of 0.643. Subsequently, the accuracy rate of VGG16 [24] was 0.65, which was slightly improved. The VGG16+SENet [24] combined with SENet module reaches 0.668, indicating that the addition of SENet module can improve the model performance to a certain extent. 10layer CNN [58] has an accuracy of 0.683, a significant improvement over than previous models. MobileNet [34], which combines the generation of an adversarial network and attention mechanism, reached 0.70, and its performance was further improved. The accuracy of the Priority Ensemble CNN [39] is 0.7052, which is further improved by the integrated approach. FERW [29], a model designed specifically for a specific task, achieved 0.71 and also performed very well. The Ensemble CNN [32] has an accuracy of 0.7127, which is further improved by integrating multiple CNN models. The VGG [45] achieved 0.714, an improvement over the traditional VGG16. VGG19 has an accuracy of 0.717, which is deeper and better than VGG16.

TABLE X. ACCURACY COMPARISON BETWEEN THE DEFAULT AND OPTIMAL HYPERPARAMETER CONFIGURATION OF INCEPTION V3 AND INCEPTION V3-CBAM

Model	Hyperparameter Configuration	Test Accuracy
Inception V3	Default	0.7002
	Best	0.7040
Inception V3-CBAM	Default	0.7072
	Best	0.7110

TABLE XI. THE SPECIFIC PERFORMANCE RESULTS OF THIS RESEARCH MODELS WITH THOSE OF PREVIOUS RESEARCH MODELS

Model	Accuracy
AlexNet[60]	0.643
VGG16[24]	0.650
VGG16+SENet[24]	0.668
10-layer CNN[58]	0.683
SRGAN + AGN - MobileNet[34]	0.700
Priority Ensemble CNN[39]	0.7052
FERW[29]	0.710
Ensemble CNN[32]	0.7127
VGG[45]	0.714
VGG19	0.717
ResNet50	0.715
Inception V3	0.704
VGG19-CBAM	0.719
ResNet50-CBAM	0.724
InceptionV3-CBAM	0.711

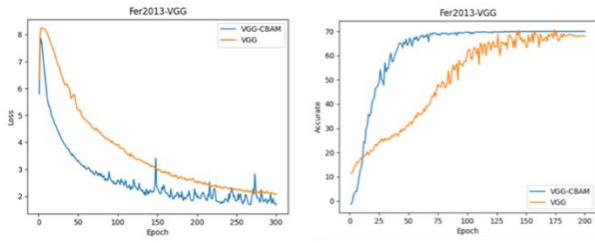


Fig. 13. Structure diagram of VGG19 and adds CBAM.

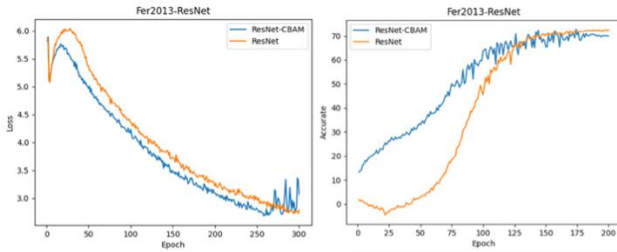


Fig. 14. Structure diagram of ResNet50 and adds CBAM.

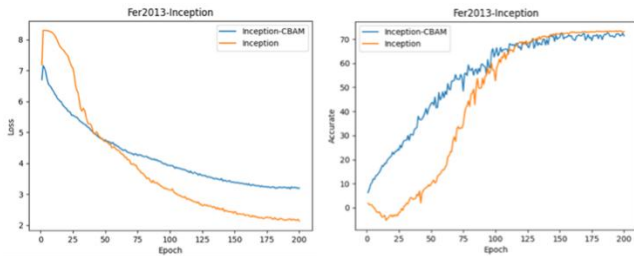


Fig. 15. Structure diagram of InceptionV3 and adds CBAM.

Fig. 13-15 shows the training verification line charts of VGG19, ResNet50 and InceptionV3, respectively, in which we can clearly observe their obvious trends. Detailed comparison: For Learning Speed, ResNet50 demonstrates the fastest learning speed among the three models, achieving significant reductions in loss and increases in accuracy in fewer epochs. VGG19 and InceptionV3 show slower but steady improvements. For Performance Stability, VGG19 shows the most stable performance with less fluctuation in its loss and accuracy curves. In contrast, ResNet50, while faster in learning, exhibits more fluctuations, indicating a more dynamic learning process with potential overfitting or regularization adjustments. InceptionV3 shows consistent improvement but at a slower rate.

The VGG series and the ResNet series exhibit relatively superior performance when handling the categories of "anger", "neutral", and "surprise", yet encounter substantial challenges when dealing with the complex categories of "fear" and "disgust". After the introduction of the CBAM, both demonstrate enhancements in the majority of categories, particularly in the manifestations of the "happy" and "surprised" categories. The Inception series model achieves the optimal classification effect for "happy" and "surprised", and the introduction of the CBAM attention mechanism further elevates the accuracy rates of these categories. Nevertheless, this model still presents considerable classification errors in the "disgust" and "fear" categories. Through the introduction of CBAM, the classification accuracy rates of all models have improved in

most emotional categories, especially in the "happy" and "surprised" categories. However, in some difficult-to-distinguish categories (such as "disgust" and "fear"), the improvement effect of CBAM is limited. Among the six models, the ResNet-CBAM and Inception-CBAM models display the most outstanding overall performance, particularly in the classification performance of complex emotional categories. Nevertheless, all models still encounter notable classification difficulties in the emotions of "disgust" and "fear", indicating that the features of these emotional categories in the Fer2013 dataset might be challenging to differentiate. The confusion matrix generated by the six models trained in this study is shown in Fig. 16.



Fig. 16. The confusion matrix generated by the six models.

TABLE XII. HYPERPARAMETER COMBINATIONS WITH THE HIGHEST TEST ACCURACY FOR INCEPTION V3 AND INCEPTION V3-CBAM WITH TRAIN-TEST SPLIT

Model	t-value	p-value
VGG19 default vs. VGG19 optimize	-2.828	0.002
ResNet50 default vs. ResNet50 optimize	-3.238	0.032
Inception default vs. Inception optimize	-4.268	0.013
VGG19-CBAM default vs. VGG19-CBAM optimize	-5.266	0.006
ResNet50-CBAM default vs. ResNet50-CBAM optimize	-4.603	0.010
Inception V3-CBAM default vs. Inception V3-CBAM optimize	-5.411	0.006

TABLE XIII. RESULTS OF 5x2 CROSS-VALIDATED PAIRED T-TEST FOR MODEL PERFORMANCE COMPARISON OF VGG19, RESNET50, AND INCEPTION V3 AND ADDED CBAM MODELS BEFORE AND AFTER THE ADDITION OF CBAM COMPARED AFTER HYPERPARAMETER TUNING WITH EACH OTHER

Model	t-value	p-value
VGG19 optimize vs VGG19-CBAM optimize	-4.1	0.015
ResNet50 optimize vs ResNet50-CBAM optimize	-4.297	0.013
Inception V3 optimize vs Inception V3-CBAM optimize	-4.347	0.013
VGG19-CBAM optimize vs ResNet50-CBAM optimize	-3.033	0.039
VGG19-CBAM optimize vs Inception V3-CBAM optimize	8.421	0.001
Inception V3-CBAM optimize vs ResNet50-CBAM optimize	-12.858	0.002

To assess the statistical significance of changes in the accuracy performance of the developed models, a 5x2 cross-validated paired t-test was conducted. Among the models, ResNet50-CBAM demonstrated the highest performance during the test phase. To evaluate whether ResNet50-CBAM significantly outperformed the other models, the test compared the best performance of ResNet50-CBAM with VGG19-CBAM and Inception-CBAM under both default and optimized configurations. The algorithms were tested using the same training-test segmentation (TTS) and cross-validation (CV) hyperparameter configurations.

The t-test results summarized in Table XII test the null hypothesis that there is no significant difference in the accuracy of the three models of CBAM under default and tuned hyperparameters. Since all P-values are less than 0.05, the null hypothesis is rejected with 95% confidence. A negative t value indicates that the default configured model performs worse on average compared to the optimized hyperparameters. The T-test results summarized in Table XIII verify the null hypothesis. After hyperparameter fine-tuning, there is no significant difference in the accuracy of each model, and all the P-values are less than 0.05, so the null hypothesis is rejected. The T-values show that there are differences between each other, and the rows of ResNet50-CBAM are significantly ahead of other models. Therefore, in summary, compared with VGG19-CBAM and Inception V3-CBAM models, the performance of ResNet50-CBAM model is statistically significant, which highlights the influence of CBAM and hyperparameter tuning on improving model accuracy.

V. CONCLUSION

This study set out to predict facial expressions and analyze the impact of the Convolutional Block Attention Module (CBAM) on the performance of deep learning models. The research successfully achieved three main objectives. The proposed method leverages the residual connections and hierarchical feature extraction capability of ResNet50, augmented with CBAM to enhance spatial and channel-wise focus. This combination improves the ability to distinguish subtle expressions, such as 'neutral' and 'sad,' while being robust to occlusions and lighting variations."

The first objective was to identify the optimal deep learning model for classifying facial expressions into seven categories within the FER2013 dataset. The study trained three key models:

VGG19, ResNet50, and Inception V3. VGG19 emerged as the top performer with a test accuracy of 0.717, slightly surpassing ResNet50 and Inception V3, which achieved 0.715 and 0.704, respectively. Despite these close results, VGG19 demonstrated a marginal but consistent advantage over the other models.

The second objective focused on assessing the impact of CBAM on these models. The results showed that integrating CBAM led to notable improvements across all three models, particularly ResNet50. The ResNet50-CBAM model achieved the highest accuracy of 0.7124, outperforming both VGG19-CBAM and Inception-CBAM, which reached 0.7104 and 0.7072, respectively. This demonstrates CBAM's ability to enhance feature extraction by enabling models to dynamically adjust weights based on channel and spatial positions, thus improving their performance, particularly in deeper networks like ResNet50.

The third objective was to optimize model performance through hyperparameter tuning. The grid search method was employed to find the best hyperparameter combinations, leading to significant accuracy improvements. Post-tuning, the ResNet50-CBAM model achieved a test accuracy of 0.724, with VGG19-CBAM and Inception V3-CBAM following at 0.719 and 0.711, respectively. The 5x2 cross-validated paired t-test results confirmed that these enhancements were statistically significant, with p-values below 0.05 for all models. The findings underscore the critical role of hyperparameter tuning in maximizing model performance and demonstrate that ResNet50, when combined with CBAM and optimized hyperparameters, outperforms both VGG19 and Inception V3. In conclusion, ResNet50-CBAM emerged as the best-performing model in this study, demonstrating superior accuracy and effectiveness in facial expression recognition tasks. This study highlights the critical importance of integrating CBAM and optimizing hyperparameters to maximize model performance. The findings emphasize that advanced feature extraction techniques and careful model tuning can significantly enhance the accuracy and reliability of deep learning models, with ResNet50-CBAM setting the benchmark for excellence in this domain.

VI. FUTURE WORK AND LIMITATIONS

The limitations of this study relate to the hyperparameters of data sets, algorithms, and classification model development. Only the FER2013 dataset was trained in this study. Due to the early presentation time of FER2013, relatively low image quality, and certain noise and blurring, the faces in the data set are mainly positive faces, and lack of diverse images such as side faces and occluded images, which may lead to inadequate adaptation of the trained model in practical applications. In addition, the FER2013 only contains the basic expression types, and the number of samples for each expression type is not balanced, and the overall sample size is small, which leads to the problem of overfitting in the process of training and testing, affecting the generalization ability.

Since the data collected in the dataset is mainly facial expression image data from 2013, and higher quality datasets have also emerged due to improvements in image acquisition hardware and diversity, such as AffectNet, future improvements could focus only on the most reliable facial image datasets to

mitigate the effects of noise during model development. The attention mechanism is a feasible choice to optimize the model performance further, not just for facial expression recognition but for other tasks as well. For example, consider an image with a cat and a table in the image description generation task. The attention mechanism helps the model focus on two important areas of the image: the cat and the table. When the model generates a description, CBAM can automatically capture the importance of cats and tables and automatically assign weights, making the model more focused on this area, which allows the model to more accurately describe the content in the image, and the generated description is more interpretable.

In the FER task, the deployment of the FER platform faces great challenges due to the different channels of image acquisition and model training. Due to the early presentation of the FER2013 dataset, the image quality is low, and there are unbalanced noise, blurring, and other conditions. In the future, we can choose newer, higher quality facial expression recognition datasets such as AffectNet, RAF-DB, etc., which have higher resolution and diverse samples, or apply more advanced image enhancement and preprocessing technologies such as denoising, deblur, and contrast enhancement. In order to improve image quality, it is possible to recognize facial expressions more accurately, helping to improve the human interaction experience. Secondly, there may be bias in expression recognition of different ages and races. In future studies, we can pay special attention to and quantify the performance differences of different groups to ensure the fairness of the model for different groups. Design hierarchical or adaptive models that can adjust parameters or weights based on input characteristics, such as age, gender, and race, to improve the accuracy of identifying different populations.

In this study, only three model architectures were tested, and feature extraction only compared attention mechanisms. In future research, we can try more different types of deep learning models, such as recurrent neural networks (RNN), graph neural networks (GNN), and different model architectures, such as ResNet or Inception. In addition to attention mechanisms, other feature extraction methods can also be tried, such as multi-scale feature extraction, emotional feature extraction, Vision Transformer (ViT), etc. By comparing the deep learning model of facial expression recognition, this study can identify facial expressions more accurately, which helps to improve the human-computer interaction experience, provide users with more intelligent and convenient services and experiences, and solve practical problems. In addition, this study can be used as a basic framework for developing face recognition based on deep learning models.

ACKNOWLEDGMENT

Funding: This research was funded by the Universiti Kebangsaan Malaysia (Grant code:FRGS/1/2024/ICT06/UKM/02/3).

Authors' Contribution: The authors confirm contribution to the paper as follows: study conception and design: Liu Luan Xiang Wei, Nor Samsiah Sani; data collection: Liu Luan Xiang Wei; analysis and interpretation of results: Liu Luan Xiang Wei, Nor Samsiah Sani; draft manuscript preparation: Liu Luan

Xiang Wei, Nor Samsiah Sani. All authors reviewed the results and approved the final version of the manuscript.

Conflict of Interest The corresponding author states that there is no conflict of interest on behalf of all authors.

Data Availability The data used in this study are available from the following resources in the public domain: <https://www.kaggle.com/datasets/msambare/fer2013>

REFERENCES

- [1] Dobrojevic, M., Zivkovic, M., Chhabra, A., Sani, N. S., Bacanin, N., & Amin, M. M. (2023). Addressing internet of things security by enhanced sine cosine metaheuristics tuned hybrid machine learning model and results interpretation based on shap approach. *PeerJ Computer Science*, 9, e1405. <https://doi.org/10.1109/CSASE48920.2020.9142065>.
- [2] Suwadi, N. A., Derbali, M., Sani, N. S., Lam, M. C., Arshad, H., Khan, I., & Kim, K. I. (2022). An optimized approach for predicting water quality features based on machine learning. *Wireless Communications and Mobile Computing*, 2022(1), 3397972. <https://doi.org/10.1109/IPRIA59240.2023.10147196>.
- [3] Othman, Z. A., Bakar, A. A., Sani, N. S., & Sallim, J. (2020). Household oversampling model amongst B40, M40 and T20 using classification algorithm. *International Journal of Advanced Computer Science and Applications*, 11(7).
- [4] Mohamed Nafuri, A. F., Sani, N. S., Zainudin, N. F. A., Rahman, A. H. A., & Aliff, M. (2022). Clustering analysis for classifying student academic performance in higher education. *Applied Sciences*, 12(19), 9467.
- [5] Holliday, J., Sani, N., & Willett, P. (2018). Ligand-based virtual screening using a genetic algorithm with data fusion. *Match: Communications in Mathematical and in Computer Chemistry*, 80(3). <https://doi.org/10.1109/ICISS50791.2020.9307567>.
- [6] Bassel, A., Abdulkareem, A. B., Alyasseri, Z. A. A., Sani, N. S., & Mohammed, H. J. (2022). Automatic malignant and benign skin cancer classification using a hybrid deep learning approach. *Diagnostics*, 12(10), 2472.
- [7] Abdul-Hadi, M.H., Waleed, J.: Human Speech and Facial Emotion Recognition Technique Using SVM. In: 2020 International Conference on Computer Science and Software Engineering (CSASE). pp. 191–196 IEEE, Duhok, Iraq (2020). <https://doi.org/10.1109/CSASE48920.2020.9142065>.
- [8] Afshar, E. et al.: Facial Expression Recognition using Spatial Feature Extraction and Ensemble Deep Networks. In: 2023 6th International Conference on Pattern Recognition and Image Analysis (IPRIA). pp. 1–6 IEEE, Qom, Iran, Islamic Republic of (2023). <https://doi.org/10.1109/IPRIA59240.2023.10147196>.
- [9] Agrawal, I. et al.: Emotion Recognition from Facial Expression using CNN. In: 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC). pp. 01–06 IEEE, Bangalore, India (2021). <https://doi.org/10.1109/R10-HTC53172.2021.9641578>.
- [10] Alexeevskaya, Y.A. et al.: Recognizing Human Emotions Using a Convolutional Neural Network. In: 2022 4th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE). pp. 1–6 IEEE, Moscow, Russian Federation (2022). <https://doi.org/10.1109/REEPE53907.2022.9731391>.
- [11] Andrian, R., Supangkat, S.H.: Comparative Analysis of Deep Convolutional Neural Networks Architecture in Facial Expression Recognition: A Survey. In: 2020 International Conference on ICT for Smart Society (ICISS). pp. 1–6 IEEE, Bandung, Indonesia (2020). <https://doi.org/10.1109/ICISS50791.2020.9307567>.
- [12] Avanija, J. et al.: Facial Expression Recognition using Convolutional Neural Network. In: 2022 First International Conference on Artificial Intelligence Trends and Pattern Recognition (ICAITPR). pp. 1–7 IEEE, Hyderabad, India (2022). <https://doi.org/10.1109/ICAITPR51569.2022.9844221>.
- [13] Avula, H. et al.: CNN based Recognition of Emotion and Speech from Gestures and Facial Expressions. In: 2022 6th International Conference on Electronics, Communication and Aerospace Technology. pp. 1360–

- 1365 IEEE, Coimbatore, India (2022). <https://doi.org/10.1109/ICECA55336.2022.10009316>.
- [14] Azimi, M.: Effects of Facial Mood Expressions on Face Biometric Recognition System's Reliability. In: 2018 1st International Conference on Advanced Research in Engineering Sciences (ARES). pp. 1–5 IEEE, Dubai, United Arab Emirates (2018). <https://doi.org/10.1109/AREX.2018.8723292>.
- [15] Cha, H.-S. et al.: Real-Time Recognition of Facial Expressions Using Facial Electromyograms Recorded Around the Eyes for Social Virtual Reality Applications. IEEE Access. 8, 62065–62075 (2020). <https://doi.org/10.1109/ACCESS.2020.2983608>.
- [16] Chen, H. et al.: Facial Expression Recognition and Positive Emotion Incentive System for Human-Robot Interaction. In: 2018 13th World Congress on Intelligent Control and Automation (WCICA). pp. 407–412 IEEE, Changsha, China (2018). <https://doi.org/10.1109/WCICA.2018.8630711>.
- [17] Chen, X. et al.: DD-CISENet: Dual-Domain Cross-Iteration Squeeze and Excitation Network for Accelerated MRI Reconstruction, <http://arxiv.org/abs/2305.00088>, (2023).
- [18] Chen, Y. et al.: Facial Motion Prior Networks for Facial Expression Recognition. In: 2019 IEEE Visual Communications and Image Processing (VCIP). pp. 1–4 IEEE, Sydney, Australia (2019). <https://doi.org/10.1109/VCIP47243.2019.8965826>.
- [19] Chuanjie, Z., Changming, Z.: Facial Expression Recognition Integrating Multiple CNN Models. In: 2020 IEEE 6th International Conference on Computer and Communications (ICCC). pp. 1410–1414 IEEE, Chengdu, China (2020). <https://doi.org/10.1109/ICCC51575.2020.9345285>.
- [20] Das, A., N. N.: Facial Expression Recognition System with Local Binary Features of Neural Network. In: 2023 International Conference on Data Science and Network Security (ICDSNS). pp. 1–5 IEEE, Tiptur, India (2023). <https://doi.org/10.1109/ICDSNS58469.2023.10244983>.
- [21] Dong, J. et al.: Segmentation Algorithm of Magnetic Resonance Imaging Glioma under Fully Convolutional Densely Connected Convolutional Networks. Stem Cells International. 2022, 1–9 (2022). <https://doi.org/10.1155/2022/8619690>.
- [22] Dwijayanti, S. et al.: Facial Expression Recognition and Face Recognition Using a Convolutional Neural Network. In: 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). pp. 621–626 IEEE, Yogyakarta, Indonesia (2020). <https://doi.org/10.1109/ISRITI51436.2020.9315513>.
- [23] Dy, M.L.I.C. et al.: Multimodal Emotion Recognition Using a Spontaneous Filipino Emotion Database. In: 2010 3rd International Conference on Human-Centric Computing. pp. 1–5 IEEE, Cebu, Philippines (2010). <https://doi.org/10.1109/HUMANCOM.2010.5563314>.
- [24] Ekman, P., Friesen, W.V.: Facial Action Coding System, <http://doi.apa.org/getdoi.cfm?doi=10.1037/t27734-000>, (2019). <https://doi.org/10.1037/t27734-000>.
- [25] Fu, S.: Research on Facial Expression Recognition Based on Deep Learning Method. In: 2022 IEEE 4th International Conference on Civil Aviation Safety and Information Technology (ICCASIT). pp. 818–821 IEEE, Dali, China (2022). <https://doi.org/10.1109/ICCASIT55263.2022.9987082>.
- [26] Gaman, Y. et al.: Adaptive Learning Method in Facial Expression Recognition Model Using Fuzzy-ART. In: 2019 IEEE 1st Global Conference on Life Sciences and Technologies (LifeTech). pp. 85–86 IEEE, Osaka, Japan (2019). <https://doi.org/10.1109/LifeTech.2019.8884040>.
- [27] Ganatra, N. et al.: Classification of Facial Expression for Emotion Recognition using Convolutional Neural Network. In: 2022 First International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT). pp. 1–5 IEEE, Trichy, India (2022). <https://doi.org/10.1109/ICEEICT53079.2022.9768508>.
- [28] Gopalan, N.P. et al.: Facial Expression Recognition Using Geometric Landmark Points and Convolutional Neural Networks. In: 2018 International Conference on Inventive Research in Computing Applications (ICIRCA). pp. 1149–1153 IEEE, Coimbatore (2018). <https://doi.org/10.1109/ICIRCA.2018.8597226>.
- [29] Grover, R., Bansal, S.: Facial Expression Recognition: Deep Survey, Progression and Future Perspective. In: 2023 International Conference on Advancement in Computation & Computer Technologies (InCACCT). pp. 111–117 IEEE, Gharuan, India (2023). <https://doi.org/10.1109/InCACCT57535.2023.10141843>.
- [30] Han, J., Gopalakrishnan, A.K.: Real-time Evaluation of Food Acceptance From Facial Expressions Based on Exponential Decay. In: 2023 15th International Conference on Knowledge and Smart Technology (KST). pp. 1–5 IEEE, Phuket, Thailand (2023). <https://doi.org/10.1109/KST57286.2023.10086796>.
- [31] Hu, S. et al.: Natural Scene Facial Expression Recognition based on Differential Features. In: 2019 Chinese Automation Congress (CAC). pp. 2840–2844 IEEE, Hangzhou, China (2019). <https://doi.org/10.1109/CAC48633.2019.8997280>.
- [32] Imamura, N. et al.: Extraction of Useful Features from Neural Network for Facial Expression Recognition. In: 2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD). pp. 221–226 IEEE, Toyama, Japan (2019). <https://doi.org/10.1109/SNPD.2019.8935652>.
- [33] Incetas, M.O. et al.: A novel image Denoising approach using super resolution densely connected convolutional networks. Multimed Tools Appl. 81, 23, 33291–33309 (2022). <https://doi.org/10.1007/s11042-02213096-4>.
- [34] Islam, B. et al.: Human Facial Expression Recognition System Using Artificial Neural Network Classification of Gabor Feature Based Facial Expression Information. In: 2018 4th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT). pp. 364–368 IEEE, Dhaka, Bangladesh (2018). <https://doi.org/10.1109/ICEEICT.2018.8628050>.
- [35] Jia, C. et al.: Facial expression recognition based on the ensemble learning of CNNs. In: 2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC). pp. 1–5 IEEE, Macau, China (2020). <https://doi.org/10.1109/ICSPCC50002.2020.9259543>.
- [36] Jin, R. et al.: AVT: Au-Assisted Visual Transformer for Facial Expression Recognition. In: 2022 IEEE International Conference on Image Processing (ICIP). pp. 2661–2665 IEEE, Bordeaux, France (2022). <https://doi.org/10.1109/ICIP46576.2022.9897960>.
- [37] Jin, X. et al.: The Research and Improvement of Facial Expression Recognition Algorithm Based on Convolutional Neural Network. In: 2023 26th ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter). pp. 166–170 IEEE, Taiyuan, Taiwan (2023). <https://doi.org/10.1109/SNPD-Winter57765.2023.10224044>.
- [38] Joseph, J.L., Mathew, S.P.: Facial Expression Recognition for the Blind Using Deep Learning. In: 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON). pp. 1–5 IEEE, Kuala Lumpur, Malaysia (2021). <https://doi.org/10.1109/GUCON50781.2021.9574035>.
- [39] Ju, L., Zhao, X.: Mask-Based Attention Parallel Network for in-the-Wild Facial Expression Recognition. In: ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 2410–2414 IEEE, Singapore, Singapore (2022). <https://doi.org/10.1109/ICASSP43922.2022.9747717>.
- [40] Kim, S., Kim, H.: Deep Explanation Model for Facial Expression Recognition Through Facial Action Coding Unit. In: 2019 IEEE International Conference on Big Data and Smart Computing (BigComp). pp. 1–4 IEEE, Kyoto, Japan (2019). <https://doi.org/10.1109/BIGCOMP.2019.8679370>.
- [41] Kumar, R.: A Deep Learning Approach To Recognizing Emotions Through Facial Expressions. In: 2023 Global Conference on Wireless and Optical Technologies (GCWOT). pp. 1–5 IEEE, Malaga, Spain (2023). <https://doi.org/10.1109/GCWOT57803.2023.10064654>.
- [42] Lee, G.-C. et al.: Ensemble Algorithm of Convolution Neural Networks for Enhancing Facial Expression Recognition. In: 2022 IEEE 5th International Conference on Knowledge Innovation and Invention (ICKII). pp. 111–115 IEEE, Hualien, Taiwan (2022). <https://doi.org/10.1109/ICKII55100.2022.9983573>.
- [43] Li, H. et al.: Differential Diagnosis for Pancreatic Cysts in CT Scans Using Densely-Connected Convolutional Networks, <http://arxiv.org/abs/1806.01023>, (2018).

- [44] Li, Y. et al.: Deep Learning for Micro-Expression Recognition: A Survey. *IEEE Trans. Affective Comput.* 13, 4, 2028–2046 (2022). <https://doi.org/10.1109/TAFFC.2022.3205170>.
- [45] Liliana, D.Y. et al.: Geometric Facial Components Feature Extraction for Facial Expression Recognition. In: 2018 International Conference on Advanced Computer Science and Information Systems (ICACSIS). pp. 391–396 IEEE, Yogyakarta (2018). <https://doi.org/10.1109/ICACSIS.2018.8618248>.
- [46] Liu, H. et al.: Adaptive Multilayer Perceptual Attention Network for Facial Expression Recognition. *IEEE Trans. Circuits Syst. Video Technol.* 32, 9, 6253–6266 (2022). <https://doi.org/10.1109/TCSVT.2022.3165321>.
- [47] Liu, K.-C. et al.: Facial Expression Recognition Using Merged Convolution Neural Network. In: 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE). pp. 296–298 IEEE, Osaka, Japan (2019). <https://doi.org/10.1109/GCCE46687.2019.9015479>.
- [48] Liu, W., Fang, J.: Facial Expression Recognition Method Based on Cascade Convolution Neural Network. In: 2021 International Wireless Communications and Mobile Computing (IWCMC). pp. 1012–1015 IEEE, Harbin City, China (2021). <https://doi.org/10.1109/IWCMC51323.2021.9498621>.
- [49] Liu, Y.: Facial Expression Recognition Model Based on Improved VGGNet. In: 2023 4th International Conference on Electronic Communication and Artificial Intelligence (ICECAI). pp. 404–408 IEEE, Guangzhou, China (2023). <https://doi.org/10.1109/ICECAI58670.2023.10177007>.
- [50] Lu, H.: AF-Transformer: Attention Fusion Transformer for Facial Expression Recognition. In: 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA). pp. 939–942 IEEE, Changchun, China (2022). <https://doi.org/10.1109/CVIDLICCEA56201.2022.9824452>.
- [51] Luo, Y. et al.: Design of Facial Expression Recognition Algorithm Based on CNN Model. In: 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA). pp. 580–583 IEEE, Shenyang, China (2023). <https://doi.org/10.1109/ICPECA56706.2023.10075779>.
- [52] Meena, G. et al.: Sentiment analysis on images using convolutional neural networks based Inception-V3 transfer learning approach. *International Journal of Information Management Data Insights.* 3, 1, 100174 (2023). <https://doi.org/10.1016/j.ijime.2023.100174>.
- [53] Muhamad, M. et al.: Recognizing Human Emotion Using Computer Vision. In: 2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS). pp. 1–4 IEEE, IPOH, Malaysia (2021). <https://doi.org/10.1109/AiDAS53897.2021.9574411>.
- [54] Munasinghe, M.I.N.P.: Facial Expression Recognition Using Facial Landmarks and Random Forest Classifier. In: 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS). pp. 423–427 IEEE, Singapore (2018). <https://doi.org/10.1109/ICIS.2018.8466510>.
- [55] N, M.: Squeeze aggregated excitation network, <http://arxiv.org/abs/2308.13343>, (2023).
- [56] NV, M.: Variations of Squeeze and Excitation networks, <http://arxiv.org/abs/2304.06502>, (2023).
- [57] Nwosu, L. et al.: Deep Convolutional Neural Network for Facial Expression Recognition Using Facial Parts. In: 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech). pp. 1318–1321 IEEE, Orlando, FL (2017). <https://doi.org/10.1109/DASCPICom-DataCom-CyberSciTec.2017.213>.
- [58] Poux, D. et al.: Dynamic Facial Expression Recognition Under Partial Occlusion With Optical Flow Reconstruction. *IEEE Trans. on Image Process.* 31, 446–457 (2022). <https://doi.org/10.1109/TIP.2021.3129120>.
- [59] Shiomi, T. et al.: Facial Expression Intensity Estimation Considering Change Characteristic of Facial Feature Values for Each Facial Expression. In: 2022 23rd ACIS International Summer Virtual Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Summer). pp. 15–21 IEEE, Kyoto City, Japan (2022). <https://doi.org/10.1109/SNPDSummer57817.2022.00012>.
- [60] Taini, M. et al.: Facial expression recognition from near-infrared video sequences. In: 2008 19th International Conference on Pattern Recognition. pp. 1–4 IEEE, Tampa, FL, USA (2008). <https://doi.org/10.1109/ICPR.2008.4761697>.
- [61] Tiwari, T. et al.: Facial Expression Recognition Using Keras in Machine Learning. In: 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N). pp. 466–471 IEEE, Greater Noida, India (2021). <https://doi.org/10.1109/ICAC3N53548.2021.9725756>.
- [62] Vinutha, K. et al.: A Machine Learning based Facial Expression and Emotion Recognition for Human Computer Interaction through Fuzzy Logic System. In: 2023 International Conference on Inventive Computation Technologies (ICICT). pp. 166–173 IEEE, Lalitpur, Nepal (2023). <https://doi.org/10.1109/ICICT57646.2023.10134493>.
- [63] Yang, J. et al.: Facial Expression Recognition Based on Facial Action Unit. In: 2019 Tenth International Green and Sustainable Computing Conference (IGSC). pp. 1–6 IEEE, Alexandria, VA, USA (2019). <https://doi.org/10.1109/IGSC48788.2019.8957163>.
- [64] Yang, B. et al.: Facial Expression Recognition Using Weighted Mixture Deep Neural Network Based on DoubleChannel Facial Images. *IEEE Access.* 6, 4630–4640 (2018). <https://doi.org/10.1109/ACCESS.2017.2784096>.

YOLO-WP: A Lightweight and Efficient Algorithm for Small-Target Detection in Weld Seams of Small-Diameter Stainless Steel Pipes

Huaishu Hou, Yukun Sun*, Chaofei Jiao

School of Mechanical Engineering, Shanghai Institute of Technology, Shanghai, 201418, China

Abstract—To address the low detection efficiency and high computational resource demands of current welded pipe defect detection algorithms for small target defects, this paper proposes the YOLO-WP algorithm based on YOLOv5s. The improvements of YOLO-WP are mainly reflected in the following aspects: First, an innovative GhostFusion architecture is introduced in the backbone network. By replacing the C3 modules with C2f modules and integrating the Ghost CBS module inspired by Ghost convolution, cross-stage feature fusion is achieved, significantly enhancing computational efficiency and feature representation for small target defects. Second, the Slim-Neck lightweight design based on GSConv is employed in the neck to further optimize the network structure and reduce the number of parameters. Additionally, the SimAM lightweight attention mechanism is incorporated to improve the network's ability to extract defect features, and the Focal-EIoU loss is utilized to optimize CIoU loss, thereby enhancing small object detection and accelerating loss convergence. The experimental results show that the AP(D1) and mAP@0.5 of the YOLO-WP model are improved by 5.3% and 3%, respectively, over the original model. In addition, the number of model parameters and FLOPs are reduced by 40% and 45%, respectively, achieving a good balance between performance and efficiency. We evaluated the performance of YOLO-WP using other datasets and showed that YOLO-WP exhibits excellent applicability. Compared to existing mainstream detection algorithms, YOLO-WP is more advanced. The YOLO-WP model significantly enhances production quality in industrial defect detection, laying the foundation for building compact, high-performance embedded weld pipe surface defect detection systems.

Keywords—Welded pipe; lightweight model; defect detection; deep learning; feature extraction; attention mechanism

I. INTRODUCTION

Small-diameter stainless steel welded pipes are widely used across various fields, including oil and gas transportation, chemical production, medical equipment, and automotive components [1]. The widespread use is due to their excellent welding performance, lower manufacturing cost compared to seamless pipes, small diameter, lightweight design, high strength characteristics [2], and superior corrosion resistance [3-4]. The welding of small-diameter stainless steel pipes primarily employs the Gas Tungsten Arc Welding (GTAW) technique. In practical production, two common defects are observed on the weld surface. The first is weld voids, which can result from misalignment or omission during the welding process [5]. The second is welding porosity or pits, typically caused by tungsten electrode contamination or grinding wheel damage [6]. These

defects can significantly impact the performance of the pipes. Therefore, it is essential to perform real-time inspection of the weld seams of small-diameter stainless steel welded pipes on the production line. This approach enables the timely detection and correction of defects, preventing defective products from proceeding to subsequent offline inspection stages and avoiding unnecessary quality control costs.

Current methods for online inspection of weld surface defects in stainless steel pipes mainly include manual inspection, X-ray inspection, eddy current inspection, and ultrasonic inspection [7]. Manual inspection is considered inefficient [8]. X-ray and ultrasonic inspections require high operational skills from personnel [9]. Additionally, eddy current inspection signals are vulnerable to interference from external factors [10]. Therefore, there is an urgent need for a detection method that is efficient, easy to operate, and highly resistant to interference. Machine vision, a non-destructive testing method, can fulfill this requirement. Within recent years, with the progress of machine vision and deep learning technologies, numerous defect detection methodologies leveraging these approaches have been extensively applied in various inspection contexts, including food and agriculture, electronics fabrication, metal materials, the semiconductor sector and healthcare [11-15]. Nevertheless, machine vision-based inspection methods for detecting weld surface defects on small-diameter stainless steel welded pipes are still not widely used.

Employing machine vision for inspection not only eliminates human subjectivity but also enables quantitative defect descriptions, thereby minimizing variability in the results. This innovation enhances detection efficiency and accuracy, fostering the advancement of industrial automation. Currently, mainstream detection algorithms can be categorized into two types [16]: single-stage algorithms and two-stage algorithms. Advanced single-stage object detection algorithms include the YOLO series, DETR, SSD and CenterNet. Advanced two-stage object detection algorithms include Faster R-CNN, Mask R-CNN, Libra R-CNN and HTC. Yang [17] applied the YOLOv5 algorithm to the welded pipe defect detection and achieved excellent results. Compared to the representative two-stage object detection algorithm Faster R-CNN, YOLOv5 demonstrates superior precision and detection speed. Therefore, different algorithms should be selected and modified for different application scenarios to target specific tasks. There are many existing strategies for algorithm improvement that focus on optimizing model components to achieve desired outcomes. Zhou et al. [18]

introduced a novel model in the YOLOv5 algorithm that combines the advantages of the CSPlayer module with a global attention enhancement mechanism, improving accuracy for metal material detection. However, this approach increases model complexity and demands higher computational resources. Zhao et al. [19] enhanced the Faster R-CNN algorithm by replacing some traditional convolutional networks with deformable convolutional networks to improve the detection capability for small-size defects on steel strips. However, this approach leads to a significant increase in model parameters, making deployment more challenging and less suitable for direct application in industrial settings. Shao et al. [20] proposed the TD-Net network for detecting tiny defects in industrial products, addressing the limitations of current image-based defect detection methods in identifying small and irregularly shaped defects. However, the detection speed has decreased. Ji et al. [21] introduced the Yolo-tla algorithm, which integrates the C3CrossConv module into the YOLOv5 backbone, effectively reducing computational demand and parameter count, thus making the model more lightweight. However, the detection accuracy has slightly decreased. Han et al. [22] proposed the DFW-YOLO algorithm based on YOLOv5, which automatically calls defect indications, resolves redundant defect feature maps, and incorporates the FasterNet backbone to enhance the model's feature extraction capability. However, the detection accuracy for small target defects has decreased. Zhou et al. [23] proposed the SKS-YOLO algorithm, using EfficientNetv2 as the backbone, which significantly reduces computation and accelerates training speed while maintaining accuracy, and employs the Simplified Intersection over Union (SIoU) loss function to improve the model's capability in locating and detecting surface defects on steel plates. However, the ability to extract small target or high-frequency features has decreased in certain scenarios. Yuan et al. [24] presented the YOLO-HMC algorithm, which uses the HorNet network (MCBAM) as its backbone and incorporates an improved multi-convolution block attention module to enhance feature extraction capabilities. However, the model requires more computational resources. Despite these studies achieving breakthroughs in specific scenarios, existing algorithms still face challenges in balancing the detection accuracy of small targets, model lightweighting, and real-time performance.

However, the existing methods for detecting surface defects in small-diameter stainless steel welded pipes still face these challenges. Firstly, the irregular sizes of weld hole defects and the extremely small shapes of welding porosities further complicate the detection process. Secondly, the high computational resources required by deep learning models pose limitations on their application in online detection within actual production environments. To address these challenges, the online detection of surface defects in small-diameter stainless steel welded pipes demands models with high speed, accuracy, and ease of deployment. To this end, we developed an enhanced model called YOLO-WP. This study focuses on the following key areas:

1) *Network structure optimization*: By optimizing the deep learning network structure, we enhance the detection accuracy for small target defects and improve overall detection performance. This is achieved by removing redundant structures and modules within the network to design a lightweight model,

thereby increasing efficiency in resource-constrained environments while maintaining or even improving model performance.

2) *Incorporation of attention mechanism*: By introducing attention mechanisms, we improve detection accuracy and enhance the ability to detect small targets. Attention mechanisms allow the model to focus on critical features and regions, thereby improving the detection of small and complex defects.

3) *Loss function improvement*: By optimizing the loss function and tailoring it to the characteristics of the dataset, we further enhance the model's localization accuracy and classification performance for small target defects. This leads to improved overall detection performance.

II. ALGORITHM DESCRIPTION

A. Baseline YOLOv5s

The YOLO series of algorithms has evolved through several versions, with YOLOv5 being widely used in the field of industrial real-time detection caused by its excellent detection accuracy and outstanding detection speed. Innovations such as the CSPDarknet53 backbone, Feature Pyramid Network (FPN), adaptive anchor box computation, and advanced data augmentation techniques have enhanced the model's performance and flexibility. YOLOv5 offers multiple versions of the model, including YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x, to accommodate different computational needs and scenarios [25]. Among these, the YOLOv5s model features the smallest network depth and width, and the fastest detection speed, making it relatively well-suited for the requirements of industrial online detection. Accordingly, this paper selects YOLOv5s as the base model.

B. The Overview of YOLO-WP

Although YOLOv5s demonstrates advantages in speed, accuracy, and terminal applications, it still faces certain limitations in practical use, such as suboptimal localization accuracy, and high computational resource demands [26]. In particular, it often suffers from missed detections and inaccurate target localization when detecting small-scale defects. Hence, this paper proposes an online detection model named YOLO-WP, specifically designed for detecting weld seam surface defects in small-diameter stainless steel pipes. The design aims of the YOLO-WP model have two aspects: first, to improve operational efficiency in resource-constrained environments, and second, improve its capability in handling small-sized defects. Although the optimization and validation of this model primarily target the detection of weld seam surface defects in small-diameter stainless steel pipes, its architectural design and improvement strategies are equally applicable to other fields requiring efficient object detection. The structure of YOLO-WP model is illustrated in Fig. 1. The targeted improvements include:

1) *In the backbone network*, the paper proposes the GhostFusion architecture, which enhances feature expression through multiscale cross-stage fusion while maintaining efficient computation. In the neck network section, a lightweight Slim-Neck network, based on GSConv [27], is referenced to reduce network parameters and enhance computational resource efficiency.

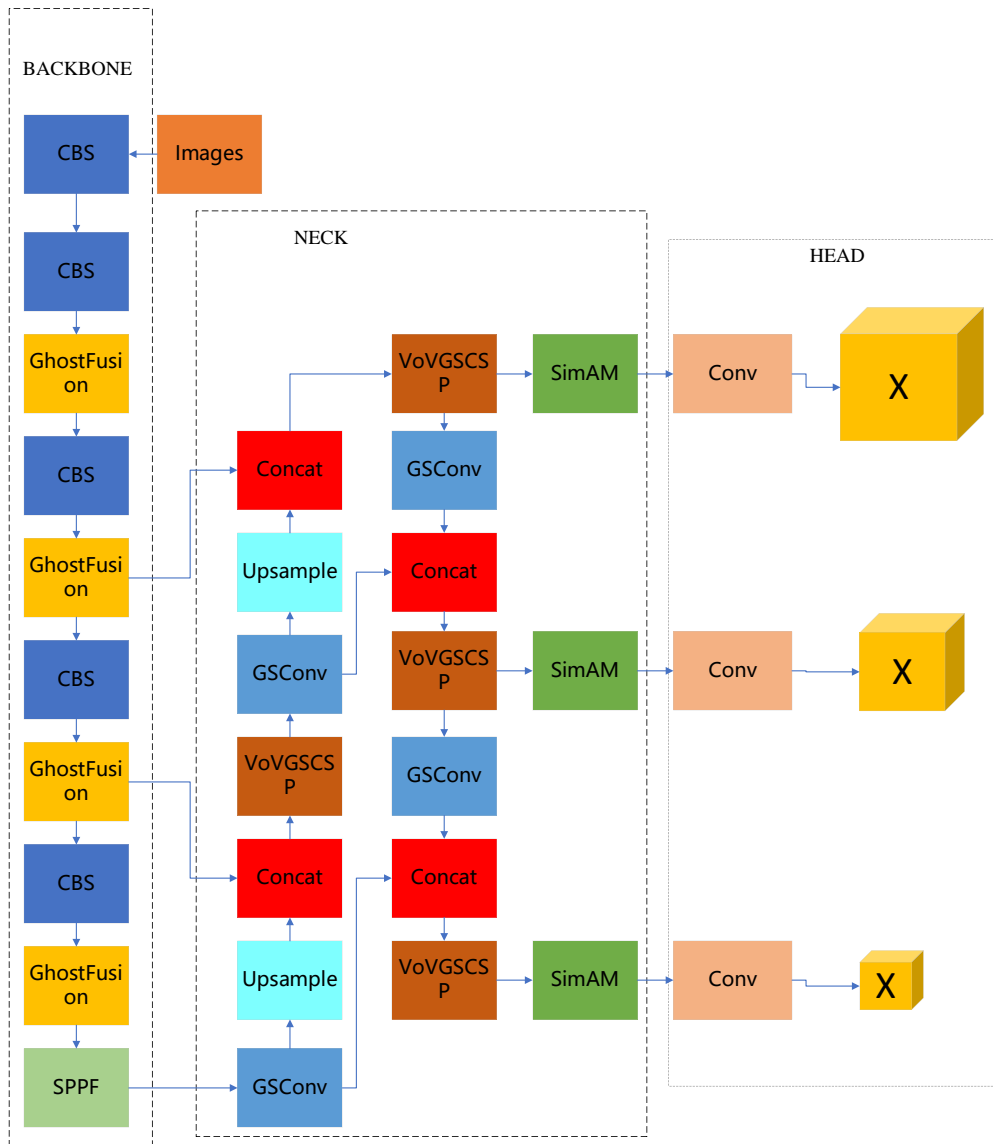


Fig. 1. Schematic model of YOLO-WP.

2) Adding the lightweight SimAM attention mechanism [28] to the neck network helps the model focus on key areas of the image, improves the fusion of feature maps across different scales, and enhances detection accuracy.

3) The Focal-EIOU loss [29] is utilized to substitute the original CIOU loss, aiming to enhance the detection of small-scale defects, address sample imbalance, and improve robustness on small and noisy datasets.

III. STRUCTURE OF KEY IMPROVEMENT COMPONENTS

A. GhostFusion Architecture

In the design of the backbone network, the innovative GhostFusion architecture is proposed. The construction process involves replacing the C3 modules with C2f modules and optimizing the CBS module in the C2f module by borrowing the core idea of Ghost convolution, and innovatively proposing the

Ghost CBS module, as depicted in Fig. 2. This design significantly enhances computational efficiency and feature representation capability through cross-stage fusion, achieving resource savings and improving detection performance for small target defects.

The core idea of Ghost Conv is to decompose conventional convolution operations into two stages: the main convolution stage and the ghost convolution stage. In the main convolution stage, 1×1 convolution kernels are used to extract condensed features, while in the ghost convolution stage, cheap 5×5 convolution kernels generate the remaining feature maps. The complete feature layer is formed by concatenating these two parts [30]. The detailed operation process is shown in Fig. 3. Additionally, the C2f module splits the input data into two parts through a Split operation. One part is directly retained, while the other part is processed through multiple BottleNeck structures to achieve multi-scale feature fusion [31].

The design greatly improves the computational efficiency and feature representation capability through cross-stage fusion, realizes resource saving, and improves the detection performance of small target defects. The core of the GhostFusion architecture lies in cross-stage feature fusion. By combining the Ghost CBS module with the C2f module, not only the computational efficiency is greatly improved, but also the semantic information of the features is significantly enhanced through the dynamic fusion of multi-scale features. This fusion mechanism enables the model to handle different sizes of receptive fields simultaneously, thus extracting more comprehensive feature information in complex scenes. The GhostFusion architecture performs well in the small target defect detection task. Through the synergy of the optimized Ghost CBS module and the C2f module, the model is able to efficiently extract features of small targets and further enhance the expressiveness of these features through the cross-stage fusion mechanism. This design not only improves the detection accuracy, but also ensures that the model can maintain efficient operation in resource-constrained scenarios. The GhostFusion architecture is designed with resource conservation and performance balance in mind. By introducing the efficient features of Ghost convolution, the

number of parameters and computation amount of the model can be significantly reduced, and at the same time, through the cross-stage fusion mechanism, the feature expression ability of the model is further enhanced. This design not only improves the operational efficiency of the model, but also ensures its high performance in complex tasks.

B. Slim-Neck Structure Based on GSConv Module

To elevate the model's detection speed and computational efficiency, this paper adopts a Slim-Neck structure based on the lightweight convolutional GSConv module [32]. Compared to traditional convolution operations, GSConv reduces the number of model parameters and computational complexity load by dividing the input features into multiple groups and independently performing convolution operations and depthwise separable convolutions on each group. GSConv introduces a mechanism that focuses on important feature channels, thus improving the model's feature extraction capability. The core principle of GSConv is to combine the characteristics of depthwise separable convolutions (DSC) and standard convolutions (SC) to achieve efficient feature map fusion and information flow. The structure of GSConv is shown in Fig. 4.

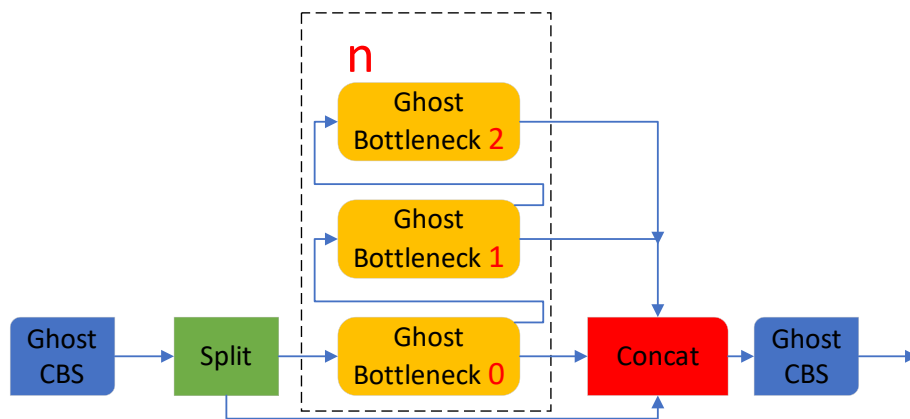


Fig. 2. Schematic diagram of the GhostFusion architecture.

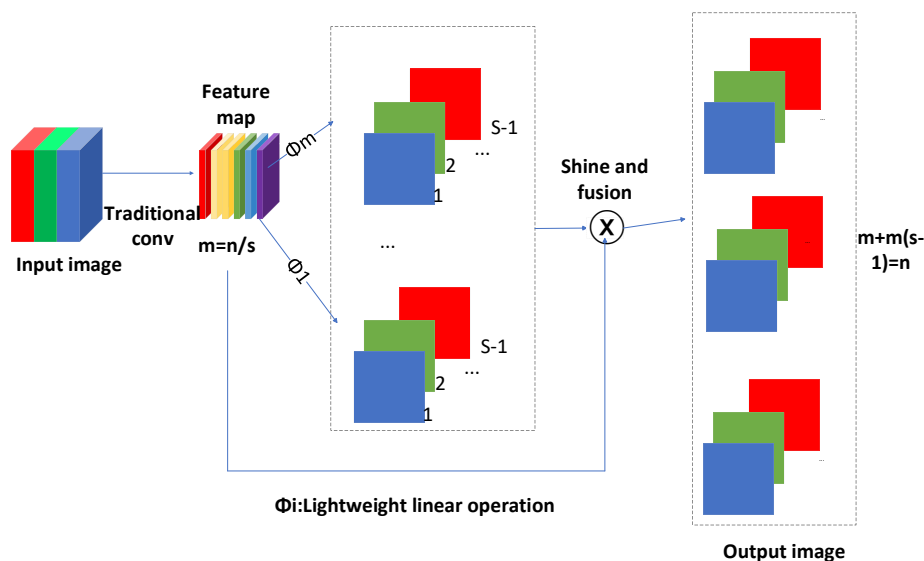


Fig. 3. GhostConv module.

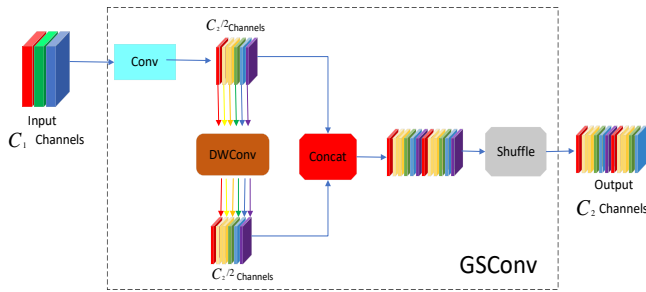


Fig. 4. The structure of GSConv.

First, GSConv inputs a downsampled standard convolution, followed by a depthwise convolution (DWConv), which concatenates the depthwise standard convolution (SC) and the depthwise separable convolution (DSC). Subsequently, a shuffle operation is applied to align the DSC output with the SC output, preserving channel and semantic information in the feature map.

The Slim-Neck lightweight neck network structure can effectively fuse and enhance features while maintaining the model's detection performance, even as it reduces computational load and the number of parameters.

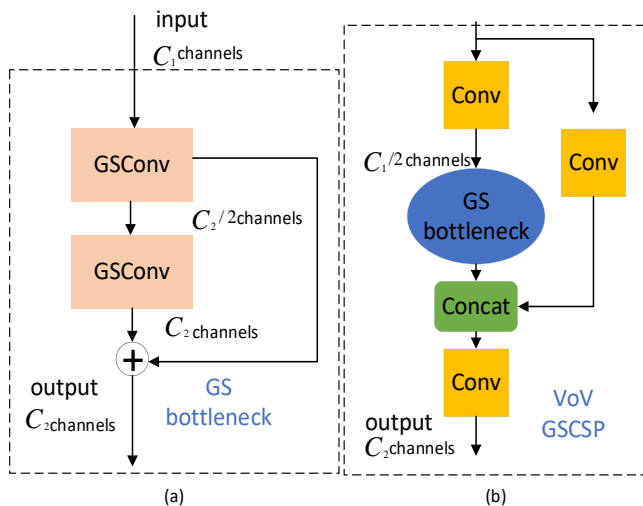


Fig. 5. Slim-Neck structure: (a) GSbottleneck module; (b) VoV-GSCSP module.

The basic building block of Slim-Neck is called VoV-GSCSP, which can replace the CSP layers comprised of standard convolutions. Among its components is the GSbottleneck, which uses GSConv as its building block, as shown in Fig. 5(a). The VoV-GSCSP module uses a one-shot aggregation method, shown in Fig. 5(b), to improve the model's target detection across different sizes by merging multi-scale feature maps, while reducing computational load and complexity.

C. SimAM Attention Mechanism

To enhance the model's representation of key features and improve its detection performance, this paper introduces an attention mechanism into the neck network of the model, thereby improving the Neck section and boosting the model's robustness. The attention mechanism typically includes channel atten-

tion and spatial attention. The parameter-free attention mechanism SimAM adopted in this paper combines both, as shown in Fig. 6. Adjacent pixels in an image usually have strong similarities, while distant pixels have weaker similarities. SimAM generates attention weights by calculating the similarity between each pixel and its neighbors in the feature map, thus inferring three-dimensional attention weights for the feature map. This effectively integrates channel and spatial attention, significantly improving the model's detection performance [33].

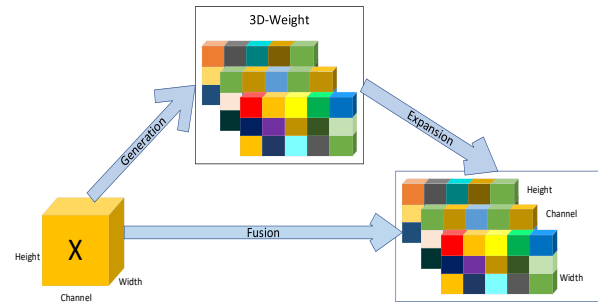


Fig. 6. The architecture of the SimAM attention mechanism module.

SimAM is grounded in the theory of visual neuroscience, where neurons with more information tend to exhibit more prominent activity compared to their adjacent neurons. In the task of surface defect detection for small-diameter stainless steel welded tubes, these neurons typically extract key features and should be assigned higher weights. This paper introduces the SimAM attention mechanism into the Neck network of the YOLOv5s model to optimize feature fusion and enhancement, while balancing network width, depth, and detection speed, thereby improving the accuracy of surface defect detection without increasing the network parameters. As shown in Eq. (1) ~ (4), SimAM evaluates neurons using an energy function for linear separability, where t represents the target neuron, x represents the adjacent neurons, and λ is a hyperparameter. The lower the energy e_t^* , the higher the distinguishability and importance of the neuron. As Eq. (4) shows, neurons are weighted based on their importance using $\frac{1}{e_t^*}$. SimAM assesses the importance of features using the energy function, providing higher interpretability and without introducing additional learnable parameters.

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (1)$$

$$\hat{\mu} = \frac{1}{M-1} \sum_{i=1}^{M-1} x_i \quad (2)$$

$$\hat{\sigma}^2 = \sum_{i=1}^{M-1} (x_i - \hat{\mu})^2 \quad (3)$$

$$\bar{X} = \text{sigmoid} \left(\frac{1}{E} \right) \odot X \quad (4)$$

D. Improvement of the Loss Function

The loss function determines the degree of agreement between the true values and the predicted values, and its performance largely reflects the model's effectiveness. In the YOLO algorithm, there are three types of loss functions: classification loss, confidence loss, and localization loss. Among these, the localization loss represents the error between the predicted bounding box and the ground truth bounding box. This paper

conducts research and improvement on the localization loss function. The original localization loss function of YOLOv5s is CIOU loss, which is centered around the concept of calculating the positional alignment error between the ground truth bounding box and the predicted bounding box based on the size of the IoU (Intersection over Union). The calculation process is expressed as:

$$L_{CIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (5)$$

Where b and b^{gt} represent the centroids of the prediction frame and the true frame respectively, ρ represents the computation of the Euclidean distance between the two centroids, so $\rho^2(b, b^{gt})$ is the distance between the centroid of the prediction frame and the centroid of the defective true bounding box, and c represents the diagonal distance of the smallest closed region that can contain both the prediction frame and the true frame. αv denotes the aspect ratio between the prediction frame and the true bounding box.

From Eq. (5), it is evident that despite CIOU loss function considering the overlap area, distance between center points, and aspect ratio of the regression bounding boxes, there are still some issues. Specifically, it adjusts based solely on the aspect ratio without considering the specific values of width and height. Additionally, the gradients of width and height have opposite signs, preventing simultaneous increase or decrease, leading to potentially amplifying the width or height during optimization when both the width and height of the anchor box are greater than the defect to be detected. Therefore, this paper selects the EIOU loss function, which modifies the aspect ratio adjustment in the CIOU loss function to specific width and height regression, enabling the model to converge faster and achieve higher accuracy, hence improving the detection efficiency for small target defects. The formula is as follows:

$$L_{EIOU} = L_{IOU} + L_{dis} + L_{asp} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \frac{\rho^2(w, w^{gt})}{C_w^2} + \frac{\rho^2(h, h^{gt})}{C_h^2} \quad (6)$$

Where w, h represents the width and height of the predicted box, w^{gt}, h^{gt} represent the width and height of the real box, C_w and C_h are the width and height of the smallest outer box that covers both boxes.

In this paper, variations in viewing angles and lighting may impact dataset quality during defect identification. In addition, the irregular appearance of defects complicates precise manual labelling, resulting in imperfect alignment between the aiming frame and defects. These factors cause dramatic fluctuations in loss value when training on low-quality samples, which can severely affect model performance. The goal of the proposed Focal II is to address the imbalance between high- and low-quality samples. Balance problem, and combined with EIOU loss to form Focal-EIOU loss.

$$L_{Focal-EIOU} = IOU^\gamma L_{EIOU} \quad (7)$$

Where γ is a constant, with a verified γ value of 0.5 giving the best results [34].

IV. EXPERIMENTAL PROCEDURE DESIGN

A. Experimental Platform and Parameter Design

During the model experiments, to maintain consistency with the comparison models and ensure the comparability of the experimental results, the SGD optimizer was used with an initial learning rate of 0.01, a momentum of 0.935, and a weight decay coefficient of 0.0005. The experiment was conducted with a batch size of 16 across 100 epochs. The experimental environment configuration is shown in Table I, and all experiments in this paper were conducted using this configuration.

TABLE I. EXPERIMENTAL ENVIRONMENT CONFIGURATION

Category	Configuration
CPU	Intel® Core™ i5-12490F Processor
GPU	NVIDIA GeForce RTX 3070ti 8G
RAM	32G
Operation System	Windows 10
Framework	PyTorch 2.0.0
Programming environment	Python 3.9
CUDA	11.8

B. Experimental Datasets

In this paper, we selected stainless steel welded pipes with defects, produced in actual manufacturing, with diameters ranging from 7mm to 9mm as samples. The defects studied in this paper are shown in Fig. 7. Due to the high visual similarity between welding sand holes and welding pores on these small-diameter stainless steel welded pipes, we categorized them as a single class.

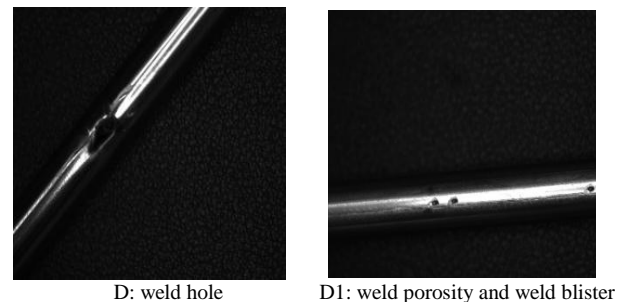


Fig. 7. The two types of defects studied in this paper.

In the absence of publicly available datasets dedicated to this field, this study established an experimental platform for image acquisition. An industrial matrix camera (model MV-CS004-10GM) was employed, precisely positioned above the welded pipe and aligned with the center of the annular aperture to ensure consistency and high quality in image acquisition. The smooth surface of the pipe reflects light centrally, resulting in brighter tones in the image, whereas weld void defects scatter the light due to their surface characteristics, creating darker areas that clearly outline the defects. Additionally, this study utilized an adjustable-brightness annular LED aperture as the light source and integrated a slider motor to accurately adjust the distance between the welded pipe and the camera, enabling rapid and precise focusing for pipes of different specifications and enhancing the flexibility and adaptability of the system. The specific setup is shown in Fig. 8.

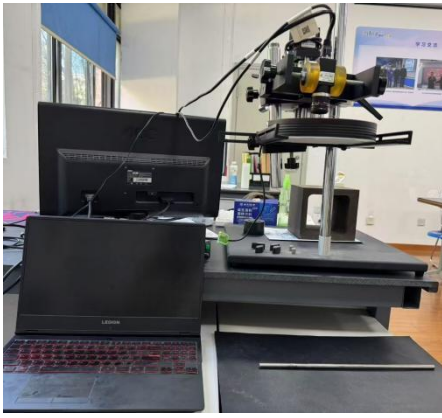


Fig. 8. Image acquisition platform.

To demonstrate the improved model's generalization capability, this study captured images of two defects types under varying lighting conditions, angles, distances, and focal lengths. For each type of defect, 2,000 images were selected to form the training set and 600 images for the validation set, with all images having a resolution of 640×640 pixels. Subsequently, the images were annotated, and the number of defect instances in both the training and validation sets was statistically analyzed. Detailed data are presented in Table II.

TABLE II. DEFECT DATA LABELING STATISTICS

Defect Type	Dataset Labels	
	Training Dataset	Testing Dataset
D: weld hole	2866	784
D1: weld porosity and weld blister	2995	813

V. EXPERIMENTAL RESULTS AND ANALYSES

A. Algorithm Evaluation Metrics

Accuracy P, mean accuracy mAP, recall R, parameters, model complexity (FLOPs) and FPS were used to evaluate the performance of the model.

$$P = \frac{TP}{TP+FP} \quad (8)$$

$$R = \frac{TP}{TP+FN} \quad (9)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{C} \quad (10)$$

Where TP is the number of correctly detected defects; FP is the number of incorrectly detected defects; FN is the number of undetected defects; AP denotes the accuracy of the detection; the value of mAP is obtained by averaging all the category APs; and C is the total number of detected categories.

When larger mAP and P values indicate higher detection accuracy, smaller parameters and FLOPs reflect a more lightweight model, and higher FPS reflects the faster algorithm detection speed.

B. YOLO-WP Ablation Experiments

To validate the effectiveness of the proposed improvements, this study conducted ablation experiments using YOLOv5s as the baseline model, incrementally adding improvement modules across six sets of experiments, as shown in the Table III. Compared to the first experimental group, the second group, after incorporating the GhostFusion efficient feature fusion network structure, saw an improvement of 3.2% in AP(D1) and 1.6% in mAP@0.5. Additionally, the number of parameters and FLOPs decreased by 22.2% and 33.5%, respectively, which not only enhanced computational efficiency but also improved the detection capability for small target defects. Compared to the second group, the third group introduced the Slim-Neck architecture based on the lightweight convolutional GSCConv module into the neck network. This resulted in a further reduction of 27.1% in parameters and 17.4% in FLOPs while maintaining the model's detection performance. Compared to the third set, the fourth set introduced the lightweight SimAM attention mechanism in the neck network. Although FPS decreased by 2.2, there was no increase in parameters and FLOPs. Meanwhile, AP(D), AP(D1), and mAP@0.5 improved by 0.5%, 1.2%, and 0.8%, respectively. In the fifth set, the Focal-EIOU loss was used to optimize CIOU loss, further improving the model's localization accuracy and convergence speed compared to the fifth set. AP(D), AP(D1), and mAP@0.5 increased by 0.5%, 0.8%, and 0.7%, respectively, with mAP@0.5 reaching 96.6%. The ablation experiments from the second to the fifth set demonstrate that each improvement method effectively optimizes the model. Compared to the baseline model YOLOv5s in the first set, the improved YOLO-WP model (sixth set) achieved a 5.3% increase in AP(D1) and a 3% increase in mAP@0.5, while reducing parameters and FLOPs by 40% and 45%, respectively. The YOLO-WP model efficiently detects small target defects with higher performance and lower computational cost, thereby enhancing overall detection accuracy. Although FPS decreased by 2.6, it still meets the requirements for online detection of surface defects in small-diameter stainless steel welded pipe seams.

TABLE III. ABLATION EXPERIMENTS

Group Model		AP(%)		P(%)	mAP@0.5 (%)	Parameters(106)	FLOPs (G)	FPS (F/S)
		D	DI					
1	YOLOv5s(Baseline)	97.1	90.3	93.3	93.7	7.2	15.5	105.3
2	YOLOv5s+GhostFusion	97.3	93.2	94.1	95.2	5.9	10.3	104.9
3	YOLOv5s+GhostFusion+Slim-Neck	97.1	93.1	93.9	95.1	4.3	8.5	103.7
4	YOLOv5s+GhostFusion+Slim-Neck+SimAM	97.6	94.3	94.5	95.9	4.3	8.5	101.5
5	YOLOv5s+GhostFu- sion+SlimNeck+SimAM+Focal-Elou	98.1	95.1	94.7	96.6	4.3	8.5	102.7

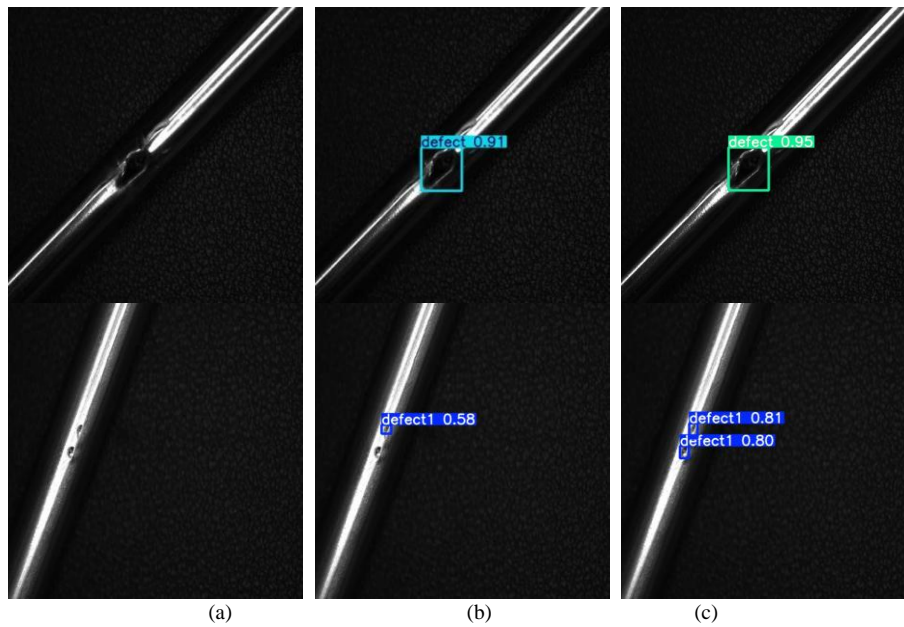


Fig. 9. Validation results for YOLO-WP and YOLOv5s. (a) Original image; (b) YOLOv5s; (c) YOLO-WP.

To offer a clearer illustration of the detection performance of the enhanced model, two images with two types of defects were randomly selected from the test set to evaluate the YOLO-WP model. This paper provides a visual comparison of the detection outcomes of the YOLO-WP model and the YOLOv5s model on images of small-diameter stainless steel welded pipes. As shown in Fig. 9, the YOLOv5s algorithm model exhibited missed detections and low accuracy in detecting small target defects. In contrast, the improved YOLO-WP algorithm model accurately located the weld seam surface defects within the images, demonstrating superior overall detection accuracy compared to the YOLOv5s algorithm model. This effectively addresses the issues of low detection efficiency and missed detections of small target on small-diameter stainless steel pipes.

C. Comparative Experiments with Multiple Datasets

To validate the generalization capability and robustness of the YOLO-WP model and to ensure its effectiveness across a wide range of applications, this paper designs experiments to test YOLOv5s and YOLO-WP on the MT Defects Dataset and NEU-DET Dataset. The MT Defects Dataset, used for magnetic tile surface defect detection, consists of 1,344 images that encompass five types of defects: pores, cracks, wear, fractures, and uneven surfaces. The NEU-DET dataset, utilized for detecting surface defects on hot-rolled steel strips, comprises 1,800 images that encompass six defect types: rolled scale, cracks, patches, pitted surfaces, inclusions, and scratches. Both datasets are divided into training and validation sets at a ratio of 7:3. The experimental results are shown in Tables IV and V. Table IV shows that the YOLO-WP model's mAP@0.5 increased by 0.5% compared to YOLOv5s. Additionally, the model's parameters and FLOPs are reduced by 38% and 43% respectively, while the detection speed is nearly unaffected. These results indicate that the YOLO-WP model achieves higher accuracy in detecting magnetic tile surface defects, demonstrates greater computational efficiency, and is easier to

deploy. As shown in Table V, the YOLO-WP model's mAP@0.5 increased by 0.9% compared to YOLOv5s. Moreover, the model's parameters and FLOPs were reduced by 36% and 43%, respectively, while the detection speed remained almost unchanged. These findings further demonstrate that the YOLO-WP model outperforms YOLOv5s in detecting surface defects in the hot-rolled steel strip dataset. The experimental results from both the MT Defects Dataset and the NEU-DET Dataset confirm that YOLO-WP surpasses YOLOv5s in terms of accuracy, model complexity, and computational efficiency. Overall, these experiments suggest that YOLO-WP offers greater practical value and stronger robustness for industrial applications compared to YOLOv5s.

TABLE IV. EXPERIMENTAL RESULTS OF MT DEFECTS DATASET

Model	mAP@0.5(%)	Parameters(10 ⁶)	FLOPs(G)	FPS(F/S)
YOLOv5s	89.1	7.3	15.4	87
YOLO-WP	89.6	4.5	8.7	86

TABLE V. EXPERIMENTAL RESULTS OF NEU-DET ON THE DATASET

Model	mAP@0.5(%)	Parameters(10 ⁶)	FLOPs(G)	FPS(F/S)
YOLOv5s	85.3	7.3	15.4	84
YOLO-WP	86.1	4.6	8.7	82

D. Comparison of Frontier Models

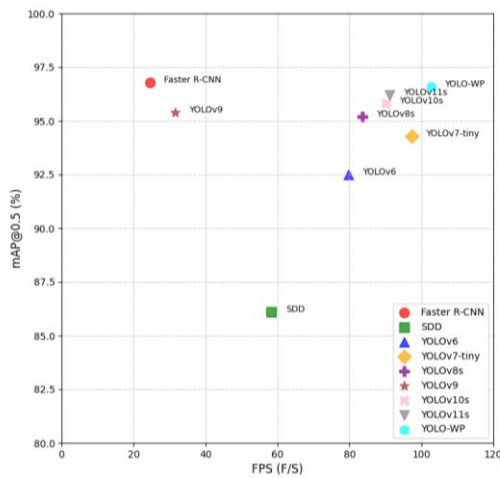
To more effectively assess the performance of the enhanced model introduced in this paper, we carried out comparative experiments using various object detection algorithms on a dataset for detecting surface defects in small-diameter stainless steel welded pipe seams. These algorithms include Faster R-CNN, SSD, YOLOv6, YOLOv7-tiny [35], YOLOv8s, YOLOv9, YOLOv10s [36] and YOLOv11s [37], along with our algorithm

(YOLO-WP). As shown in the Table VI, YOLO-WP achieved an mAP@0.5 of 96.6%, with 4.3 million parameters, 8.5 GFLOPs, and a speed of 102.7 FPS. The experimental results indicate that, compared to other algorithms, the YOLO-WP model has the smallest number of parameters and FLOPs, as well as the fastest detection speed. Although YOLO-WP's mAP@0.5 is 0.2% lower than that of Faster R-CNN, it significantly reduces the number of parameters and FLOPs.

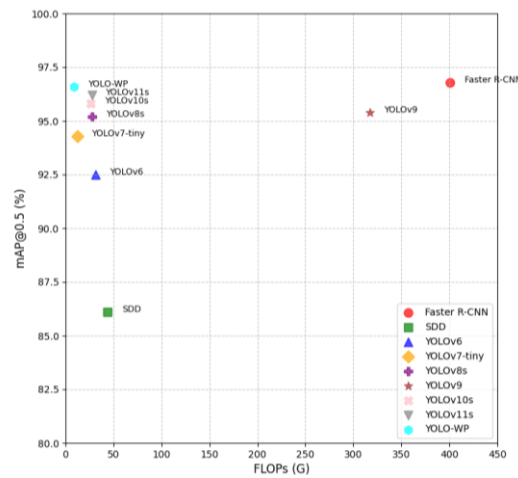
To better demonstrate the balanced advantages of the YOLO-WP model regarding detection accuracy, model complexity, and detection speed, this paper presents scatter plots illustrating the relationships among these factors. As shown in Fig. 10(a), the YOLO-WP model is positioned in the upper right corner of the two-dimensional coordinate system, reflecting its ability to balance speed and detection accuracy. In Fig. 10(b), the YOLO-WP model is positioned in the upper left corner of the two-dimensional coordinate system, indicating its capability to balance computational efficiency and detection accuracy.

TABLE VI. COMPARISON OF FRONTIER MODELS

Model	mAP@0.5(%)	Parameters(10 ⁶)	FLOPs(G)	FPS(F/S)
Faster R-CNN	96.8	135.4	400.8	24.5
SDD	86.1	63.5	43.7	58.3
YOLOv6	92.5	15.6	31.3	79.7
YOLOv7-tiny	94.3	6.3	12.7	97.4
YOLOv8s	95.2	10.3	27.5	83.6
YOLOv9	95.4	70.5	317.2	31.5
YOLOv10s	95.8	11.3	26.2	90.1
YOLOv11s	96.2	13.4	27.3	91.2
YOLO-WP(Ours)	96.6	4.3	8.5	102.7



(a) Scatter plot of the relationship between mAP@0.5 and FPS.



(b) Scatter plot of the relationship between mAP@0.5 and FLOPs.

Fig. 10. Relationship scatter plot.

Based on the comparative experimental data and analysis presented above, the improved algorithm proposed in this paper achieves a balance between accuracy and lightweight design, efficiently utilizing computational resources to reach an optimal balance between model accuracy and training weights. This further underscores the superiority of the YOLO-WP algorithm. By significantly reducing the number of parameters and FLOPs, the YOLO-WP algorithm lowers hardware requirements. Consequently, this improvement meets the demands for online detection of weld seam surface defects in small-diameter stainless steel pipes.

VI. CONCLUSIONS

To address the gap in detecting surface defects on small-diameter stainless steel pipe weld seams, and to overcome the limitations of YOLOv5s on terminal devices, which arise from insufficient computational power and poor detection capabilities for small object defects, this paper proposes the significantly improved YOLO-WP algorithm. By introducing the innovative GhostFusion architecture, Slim-Neck lightweight design, SimAM lightweight attention mechanism, and Focal-EIou

loss function optimization, the YOLO-WP model achieves a 5.3% and 3% increase in AP(D1) and mAP@0.5, respectively, compared to the original model. Additionally, the number of model parameters and FLOPs are reduced by 40% and 45%, respectively, significantly enhancing the efficiency of small target detection and the model's applicability in resource-constrained environments. Experimental results show that the YOLO-WP model achieves high detection accuracy, low complexity, minimal computational requirements, and rapid detection speeds. This model has demonstrated robustness across different datasets. Compared to other models, YOLO-WP exhibits strong competitiveness, improving production quality and reducing costs, thereby making it suitable for industrial online inspection.

However, this study still has certain limitations. Although we collected and trained common defect data during the online production process, which to some extent reduced the burden of subsequent offline detection, we have not yet achieved comprehensive coverage of surface defects in small-diameter stainless steel pipe weld seams. Moreover, YOLO-WP still faces challenges in dealing with more complex defect types, such as

occluded or extremely small targets. In future research, we plan to introduce more types of defects for study and verify the capability of our algorithm model in offline detection of surface defects in stainless steel pipe weld seams. We will also specifically improve a set of algorithm models suitable for offline detection to achieve collaborative work between offline and online detection, thereby fully applying visual inspection to all quality control processes.

This study provides a novel solution for the detection of surface defects in small-diameter stainless steel pipe weld seams. By optimizing the network architecture, incorporating lightweight design, and improving the loss function, the model's detection performance in resource-constrained environments has been significantly enhanced. These improvements not only offer an efficient and accurate detection tool for industrial online inspection but also provide valuable references for future research in related fields.

VII. AUTHORS' CONTRIBUTION

Conceptualization, Yukun Sun and Huaishu Hou; methodology, Yukun Sun and Chaofei Jiao; validation, Yukun Sun; investigation, Huaishu Hou; data curation, Huaishu Hou and Yukun Sun. writing— original draft preparation, Yukun Sun; writing—review and editing, Huaishu Hou. All authors have read and agreed to the published version of the manuscript.

FUNDING

None

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

- [1] Bettahar K, Bouabdallah M, Badji R, et al. Microstructure and mechanical behavior in dissimilar 13Cr/2205 stainless steel welded pipes. *Materials & Design*, 2015, 85: 221-229.
- [2] Kumar M V, Balasubramanian V. Microstructure and tensile properties of friction welded SUS 304HCu austenitic stainless steel tubes. *International Journal of Pressure Vessels and Piping*, 2014, 113: 25-31.
- [3] Ressa J, Monrrabal G, Díaz A, et al. Microbiologically influenced corrosion of welded AISI 304 stainless steel pipe in well water. *Engineering Failure Analysis*, 2020, 116: 104734.
- [4] Baddoo N R. Stainless steel in construction: A review of research, applications, challenges and opportunities. *Journal of constructional steel research*, 2008, 64(11): 1199-1206.
- [5] Wang Y, Guo Z, Bai X, et al. Effect of weld defects on the mechanical properties of stainless-steel weldments on large cruise ship. *Ocean Engineering*, 2021, 235: 109385.
- [6] Ressa J, Monrrabal G, Díaz A, et al. Microbiologically influenced corrosion of welded AISI 304 stainless steel pipe in well water. *Engineering Failure Analysis*, 2020, 116: 104734.
- [7] Lee J K, Bae D S, Lee S P, et al. Evaluation on defect in the weld of stainless steel materials using nondestructive technique. *Fusion Engineering and Design*, 2014, 89(7-8): 1739-1745.
- [8] Ge J, Zhu Z, He D, et al. A vision-based algorithm for seam detection in a PAW process for large-diameter stainless steel pipes. *The international journal of advanced manufacturing technology*, 2005, 26: 1006-1011.
- [9] Gök D A. Destructive and non-destructive testings of 304 austenitic stainless steel produced by investment casting method. *Nondestructive Testing and Evaluation*, 2024: 1-13.
- [10] Peng J, Xu Z, Chen H, et al. Detection of brazing defects in stainless steel core plate using the first peak value of pulsed eddy current testing signals. *Construction and Building Materials*, 2023, 408: 133636.
- [11] Huang C Y, Hong J H, Huang E. Developing a machine vision inspection system for electronics failure analysis. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 2019, 9(9): 1912-1925.
- [12] Ahmad H M, Rahimi A. Deep learning methods for object detection in smart manufacturing: A survey. *Journal of Manufacturing Systems*, 2022, 64: 181-196.
- [13] Badgajar C M, Poulouse A, Gan H. Agricultural object detection with You Only Look Once (YOLO) Algorithm: A bibliometric and systematic literature review. *Computers and Electronics in Agriculture*, 2024, 223: 109090.
- [14] Usamentiaga R, Lema D G, Pedrayes O D, et al. Automated surface defect detection in metals: a comparative review of object detection and semantic segmentation using deep learning. *IEEE Transactions on Industry Applications*, 2022, 58(3): 4203-4213.
- [15] Li Z, Dong M, Wen S, et al. CLU-CNNs: Object detection for medical images. *Neurocomputing*, 2019, 350: 53-59.
- [16] Demetriou D, Mavromatidis P, Robert P M, et al. Real-time construction demolition waste detection using state-of-the-art deep learning methods; single-stage vs two-stage detectors. *Waste Management*, 2023, 167: 194-203.
- [17] Yang D, Cui Y, Yu Z, et al. Deep learning based steel pipe weld defect detection. *Applied Artificial Intelligence*, 2021, 35(15): 1237-1249.
- [18] Zhou C, Lu Z, Lv Z, et al. Metal surface defect detection based on improved YOLOv5. *Scientific Reports*, 2023, 13(1): 20803.
- [19] Zhao W, Chen F, Huang H, et al. A new steel defect detection algorithm based on deep learning. *Computational Intelligence and Neuroscience*, 2021, 2021(1): 5592878.
- [20] Shao R, Zhou M, Li M, et al. TD-Net: tiny defect detection network for industrial products. *Complex & Intelligent Systems*, 2024: 1-12.
- [21] Ji C L, Yu T, Gao P, et al. Yolo-tla: An Efficient and Lightweight Small Object Detection Model based on YOLOv5. *Journal of Real-Time Image Processing*, 2024, 21(4): 141.
- [22] Han Z, Li S, Chen X, et al. DFW-YOLO: YOLOv5-based algorithm using phased array ultrasonic testing for weld defect recognition. *Nondestructive Testing and Evaluation*, 2024: 1-24.
- [23] Zhou S, Ao S, Yang Z, et al. Surface Defect Detection of Steel Plate Based on SKS-YOLO. *IEEE Access*, 2024.
- [24] Yuan M, Zhou Y, Ren X, et al. YOLO-HMC: An improved method for PCB surface defect detection. *IEEE Transactions on Instrumentation and Measurement*, 2024.
- [25] Jiang P, Ergu D, Liu F, et al. A Review of Yolo algorithm developments. *Procedia computer science*, 2022, 199: 1066-1073.
- [26] Liu H, Sun F, Gu J, et al. Sf-yolov5: A lightweight small object detection algorithm based on improved feature fusion mode[J]. *Sensors*, 2022, 22(15): 5817.
- [27] Wang L, Zhang Y, Lin Y, et al. Ship Detection Algorithm Based on YOLOv5 Network Improved with Lightweight Convolution and Attention Mechanism[J]. *Algorithms*, 2023, 16(12): 534.
- [28] Shang J, Wang J, Liu S, et al. Small target detection algorithm for UAV aerial photography based on improved YOLOv5s[J]. *Electronics*, 2023, 12(11): 2434.
- [29] Huang B, Liu J, Liu X, et al. Improved YOLOv5 Network for Steel Surface Defect Detection[J]. *Metals*, 2023, 13(8): 1439.
- [30] Han K, Wang Y, Xu C, et al. GhostNets on heterogeneous devices via cheap operations. *International Journal of Computer Vision*, 2022, 130(4): 1050-1069.

- [31] Wang Z, Zhou D, Guo C, et al. Yolo-global: a real-time target detector for mineral particles. *Journal of Real-Time Image Processing*, 2024, 21(3): 1-13.
- [32] Li H, Li J, Wei H, et al. Slim-neck by GSConv: a lightweight-design for real-time detector architectures. *Journal of Real-Time Image Processing*, 2024, 21(3): 62.
- [33] Yang L, Zhang R Y, Li L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks[C]//International conference on machine learning. PMLR, 2021: 11863-11874.
- [34] Zhang Y F, Ren W, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression. *Neurocomputing*, 2022, 506: 146-157.
- [35] Hu S, Zhao F, Lu H, et al. Improving YOLOv7-tiny for infrared and visible light image object detection on drones. *Remote Sensing*, 2023, 15(13): 3214.
- [36] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. arXiv preprint arXiv:2405.14458, 2024.
- [37] Khanam R, Hussain M. Yolov11: An overview of the key architectural enhancements[J]. arXiv preprint arXiv:2410.17725, 2024.

Determination of Pre Coding Elements and Activities for a Pre Coding Program Model for Kindergarten Children Using the Fuzzy Delphi Method (FDM)

Siti Naimah Rahman¹, Norly Jamil^{2*}, Intan Farahana Abdul Rani³, Hafizul Fahri Hanafi⁴

Faculty of Human Development, Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak, Malaysia^{1, 2, 3}
Faculty of Computer and Meta-Technology, Universiti Pendidikan Sultan Idris, Tanjong Malim, Perak, Malaysia⁴

Abstract—Computational Thinking (CT) skills are becoming increasingly crucial in education, particularly in early childhood education. Pre coding, which involves hands-on activities with real objects, has been shown to be quite effective in fostering kindergarten children's computational skills. Pre coding, on the other hand, is essential for boosting children's CT skills, but teachers frequently lack the information necessary to teach these skills successfully. Their successful adoption is hampered by the early childhood education community's lack of interest in CT skills and the sparse application of pre coding techniques. In order to help kindergarten instructors incorporate pre coding into their teaching and learning, this study focuses on defining the elements and activities described in a pre-coding program model. The study reviewed and compiled a list of prior literature's pre coding elements and activities. Subsequently, the Fuzzy Delphi Method (FDM) was utilised to refine and validate these elements and activities. Finally, the data collected from 11 selected experts relevant to this field of study were analysed using FDM to examine consensus. The results showed that the eight identified elements and 24 pre coding activities fulfilled the following required criteria: a threshold value (d) of lower than or equal to 0.2, an agreement percentage over 75%, and a fuzzy score value (A) higher than 0.5. These findings demonstrated the suitability of the identified pre coding elements and activities for integration into a pre coding program model for kindergarten children. In summary, this study provides valuable guidance for kindergarten teachers in implementing practical pre coding activities to enhance CT skills among children.

Keywords—Expert consensus; pre coding; element; activity; kindergarten children

I. INTRODUCTION

Computational thinking (hereafter called CT) refers to a set of cognitive skills for solving problems [1 - 3]. It is also considered a thinking process [4, 5] that involves an array of cognitive skills, including critical thinking, problem-solving, logical reasoning, and creative thinking [6, 7]. In view of this, CT has emerged as a fundamental skill that everyone needs to understand and master [8]. CT skill should be integrated into compulsory school education [9]. Moreover, CT has become one of the most effective approaches for teaching students, including early childhood, in line with the development of global modernisation [10, 11].

The CT teaching approach has also received increasing attention in education and research [12], with vast implementation across many countries, including the United

States, the United Kingdom, Estonia, Australia, and Singapore [6, 9]. Coding and pre coding are widely recognised as two standardise methods used for teaching CT [13, 14]. Coding emphasises the use of digital devices, such as computers, as the primary learning tool in computer science education [15 - 18]. Teaching CT through coding typically involves learning programming, which is deemed one of the most effective methods to nurture CT skills [19, 20]. Whereas, pre coding does not require the utilisation of digital devices, it offers an alternatif approach to fostering CT skills [21, 22].

Although coding and pre coding share the same goal, i.e., applying one's CT skills, their implementation differs. Specifically, coding involves digital devices and is more generally introduced at the primary, secondary, and higher education levels [22]. In contrast, pre coding is commonly introduced in the early stages of childhood as its implementation focuses on the active involvement of children through hands-on activities with concrete objects [21, 23], such as pencils and papers, puzzles, and wooden blocks [22, 24]. This learning method deeply resonates with children as it allows them to explore the real world and develop their CT skills [16]. In fact, pre coding is often associated with a simpler and fun implementation that corresponds with children's learning process and development stages [25, 26]. Furthermore, pre coding is particularly beneficial for students from B40 families with limited access to digital learning [27], making it a relevant, appropriate, and meaningful approach to children's education.

Nevertheless, the significance of pre coding in empowering CT skills among students has not been adequately conveyed to teachers [28, 29], including kindergarten teachers [30]. Even more critically, teachers are not equipped with sufficient knowledge to teach CT and pre coding skills [29, 31]. This matter has prevented teachers from successfully introducing CT skills through pre coding activities [32 - 35]. It was revealed that teachers in early childhood education were uninterested in CT skills, partly due to the lack of efforts to highlight their significance for children through pre coding [31, 36]. Besides, teachers are not always permitted to practice pre coding approaches and CT skills in their teaching sessions [30].

Considering the issues and problems encountered by teachers, this study aims to identify the key elements of pre coding and appropriate activities to developing structured model of a pre coding activity program. These pre coding

model program elements serve as a guideline for kindergarten teachers to practise pre coding to encourage CT skills in children from an early age. This study also systematically discussed the setting and verifying elements of pre coding activities for the program model based on expert consensus through the Fuzzy Delphi Method (FDM) for its implementation.

This paper is divided into several sections: Section II presents a concise literature review regarding elements and activities of pre coding. Section III specifies the methodology of this study. Section IV describes the data analysis process. Section V details the findings and discussion. Finally, Section VI concludes the study and recommends future works.

II. LITERATURE REVIEW

Pre coding is a type of unplugged activity, better known as unplugged coding [21, 37, 38]. This activity supports the development of CT skills without using electronic devices, such as computers, mobile phones, and tablets [16, 21, 23, 38-40]. As such, this approach emphasises hands-on activities and utilises easily accessible concrete materials, such as papers and pencils, cards, and puzzles [23, 24, 37, 39, 41]. This hands-on approach aligns with the constructivism theory that focuses on children's learning via active exploration and real-world interaction [42]. Hence, meaningful real-world experience can improve children's learning process.

Pre coding is viewed as a learning process for kindergarten and preschoolers that adopts physical movement activities and enjoyable games to develop elementary knowledge, nurture CT, and introduce core computer science concepts [16, 43-45]. Pre coding activities are typically conducted through indoor games using a wide range of materials, including pens and papers, cards, and game figurines [24, 46, 47]. Past studies concluded that pre coding incorporates physical activity with accessible materials to provide a fun and meaningful experience that develops CT skills in kindergarten children.

Pre coding is considered a suitable and meaningful learning approach for kindergarten children because it incorporates physical activities without utilising digital devices, such as computers, which may be perceived as complex tools for young learners [37, 44, 48]. It is also viewed as more relevant for children [49] as it emphasises learning in context rather than focusing solely on specific content related to pre coding subjects [9, 42]. As outlined in constructivism theory, pre coding concentrates on continuous learning through environmental experiences that support children's thinking process and active involvement [42]. Therefore, pre coding learning is typically incorporated with other subjects, such as language, mathematics, science, and visual arts [50].

In addition, the pre coding approach helps children develop their computational skills, which further promotes their problem-solving, logic, and creative thinking abilities [51]. This method also provides children with a deep-thinking experience when engaging in a task [23], enabling them to solve complex problems effectively and creatively [24, 52]. Children's mastery of CT skills also promotes their high-level thinking abilities, allowing them to think creatively, express their views in many ways, and analyse problems from different

viewpoints [53]. Thus, pre coding is essentially crucial in early childhood education.

In navigating today's digital world, this study assessed Malaysia's Industrial Revolution 4.0 (IR4.0) Policy, which underscores the need for the country to remain competitive within the digital ecosystem [54]. Among the essential skills required to address the challenges of IR4.0 are logical thinking, cognitive development, and creative thinking [17, 41, 55, 56]. As the Sustainable Development Goal (SDG) outlines, these skills are vital for achieving high-quality education. In order to meet the SDG targets and the IR4.0 goals, this study highlights the implementation of a pre coding program that nurtures CT skills in kindergarten children, ensuring these critical skills are developed from an early age.

There is also a growing demand to identify pre coding elements and activities that are suitable for implementation in early childhood education through a comprehensive literature review. Several researchers have conducted pre coding programs for children. Fig. 1 illustrates the definition of the respective pre coding elements, while Table I lists the pre coding activities based on previous studies.

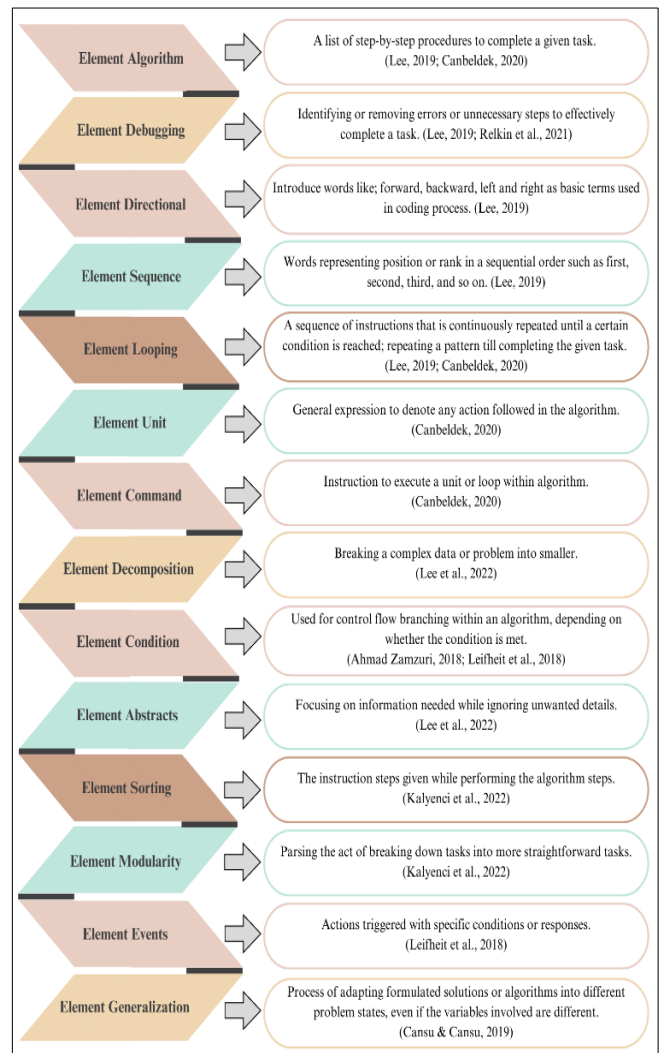


Fig. 1. Definition of pre-coding elements.

TABLE I. PRE CODING ACTIVITIES

No.	Pre coding activities	Previous studies
1.	Daily routine	[22]
2.	Storytelling	[16]
3.	Play	[16]
4.	Telling stories using books	[41]
5.	Coding sheet	[41]
6.	Treasure hunt	[41]
7.	Location search by map	[22]
8.	Following recipe	[22]
9.	Modelling how to perform a task	[22]
10.	Puzzle	[22]
11.	Activities using concrete materials	[16]
12.	Card use	[52, 57]
13.	LEGO pattern	[58]
14.	Sequencing stories	[58]
15.	Vocabulary building songs	[58]
16.	Direction game through cards	[58]
17.	Tic-tac-toe	[58]
18.	Hop scotch coding	[17]
19.	Neighbourhood walk activity map	[17]
20.	Robot Robi's Friend (activity map)	[17]
21.	Story card	[37]
22.	Coding through stories	[59]
23.	Storigami	[59]
24.	Robotic kits	[16]
25.	Tetris activity	[46]
26.	"Repetition Drawing" activity	[46]

Algorithmic elements are often prioritised in pre coding skills for children [4, 16, 23, 24, 37, 39, 40]. These pre coding elements, which are considered vital for children, teach them to follow a set of step-by-step instructions built to solve a task or problem [16, 24, 37]. These elements also encourage logical thinking involving data analysis processes and systematic problem-solving, rendering them suitable for teaching children [16, 38, 39].

Repetition control structure is another essential skill element in pre coding learning [23, 37]. This structural element refers to a set of continuous repeating instructions as long as specific conditions are fulfilled [23, 40]. The repetition control structure in a pre coding program aims to assist children in performing each task or activity based on the assigned conditions and counter value [23, 38, 46]. Hence, the string of this element becomes a key skill for children to master [38].

Furthermore, the sequence control structural element [37, 40, 58, 60] is a critical skill that needs to be mastered to understand the CT skill concept [40]. The sequence control structural element is typically introduced and implemented through daily routine activities [24]. Lee et al. [24] noted that

nurturing this element helps children recognise sequence patterns in their daily routines. In other words, daily routine activities can foster children's understanding of the sequence control structures. The study also revealed that children who successfully master this skill could indirectly anticipate future events and identify patterns or past patterns according to their understanding of daily routine activities. Hence, this skill is vital for children's development, as it boosts their cognitive abilities.

Three field experts assessed the initial pre coding element list to obtain confirmation and initial evaluation. The experts accepted only seven pre coding elements that were deemed appropriate for the early childhood pre coding program model, as presented in Table II.

TABLE II. PRE CODING ELEMENTS AFTER EXPERT INTERVIEWS

No.	Pre coding elements
1.	Algorithm
2.	Debugging
3.	Directional
4.	Sequence
5.	Looping
6.	Command
7.	Decomposition

The directional element is also a key pre coding skill [16, 17, 21, 24, 41]. Children need to master this element when learning to code, as it is frequently utilised in the coding process [41]. The directional element describes the interaction between one object and another, such as the spatial relationship, 'in front,' 'beside,' and 'at the edge' [17]. In addition, specific words, such as 'forward,' 'backward,' 'to the right,' and 'to the left,' are often employed to describe the concept of direction [41]. Using arrows and hand signals during pre coding learning is instrumental in helping children recognise the intended direction correctly and accurately [21]. In short, applying this element indirectly provides an easier, faster, and more meaningful understanding of the concept of direction.

Besides, error detection, or debugging, is a significant element of pre coding skills [21, 48, 61]. Debugging refers to the identification of unnecessary steps or errors to complete a task more effectively [21]. Additionally, debugging encourages children to explore, observe, reflect, and communicate when seeking solutions for their tasks [62] since activities or tasks involving this element are relatively open-ended in nature [48].

Lastly, the resolution element is a vital component of pre coding skills [24, 63, 64]. This element breaks down complex problems or systems into smaller parts to facilitate understanding of the solution process [24, 64]. Besides, simplifying the problems nurtures the thinking process to recognise specific patterns and dismiss irrelevant elements when solving problems [24]. Mastering the resolution element allows children to present various solutions when evaluating their strengths and limitations before selecting the optimal strategy for solving the problem [65]. Hence, the resolution element must be emphasised as these problem-solving skills help children to think faster when solving a problem.

III. METHODOLOGY

The Human Research Ethics Committee of Universitas Pendidikan Sultan Idris granted the approval for this study from January 31, 2023, to January 31, 2024. This study employed the Fuzzy Delphi Method (FDM) to obtain expert consensus on the elements and activities suitable to be incorporated in a pre coding model for preschoolers. The FDM adapts the classic Delphi method, which integrates fuzzy number sets while maintaining the Delphi method itself [66]. This method was selected because it shortens the cycle process and avoids data loss, thus enhancing economic efficacy in terms of time and cost [26]. Besides, FDM is an effective technique due to its

theory set fuzzy that resolves uncertainties against experts' consensus.

FDM is also a structured and systematic analytical procedure [67]. It has been broadly employed to validate the components for training contents due to its ability to obtain fuzzy score values in the form of ranks, which can serve as a determinant and priorities for an element based on expert consensus [27–29]. The design of this study consists of three stages: literary review, expert assessment, and FDM analysis. The study method is elucidated in Fig. 2.

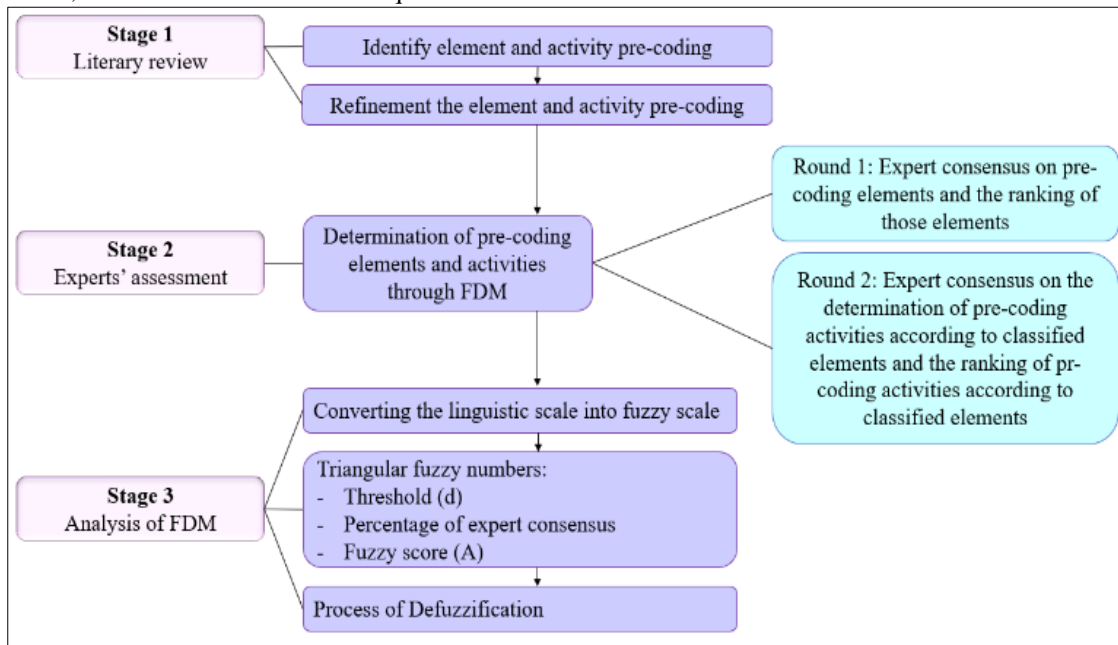


Fig. 2. Study method.

A. Literary Review Stage

The first stage aimed to identify the appropriate elements and activities for developing the pre coding program model for kindergarten children. A comprehensive literary review was carried out via several research databases, including Scopus, Elsevier, Springer Link, Research Gate, and Google Scholar. The data from this literary review comprised a preliminary list of pre coding elements and activities. The data were then evaluated by three experts specialising in computer science and CT skills through Google Meet interviews to assess the suitability and acceptability elements and pre coding activities for kindergarten.

B. Expert Assessment Stage

The FDM method was applied by constructing a questionnaire and analysing the data based on expert consensus. The questionnaire was developed based on the elements and activities identified from the literary review stage. The questionnaire consists of a 7-point linguistic scale a balanced range of response options, capturing a broader spectrum of attitudes, opinions, and behavior [68], as shown in Table III. In this process, three experts (Table IV) reviewed the questionnaire to ensure content validity, clarity of wording, and structural integrity.

TABLE III. THE 7-POINT LIKERT SCALE AND FUZZY SCALE

Linguistic variable	Likert scale	Fuzzy scale
Strongly disagree	1	(0.0,0.0,0.1)
Highly disagree	2	(0.0,0.1,0.3)
Disagree	3	(0.1,0.3,0.5)
Moderately agree	4	(0.3,0.5,0.7)
Agree	5	(0.5,0.7,0.9)
Highly agree	6	(0.7,0.9,1.0)
Strongly agree	7	(0.9,1.0,1.0)

TABLE IV. BACKGROUND OF THE THREE EXPERTS INVOLVED DURING THE DATA VALIDATION PROCESS

Expert no.	Expertise	Experience (years)	Organisation
1	Early childhood education	6	Public university
2	Early childhood education	6	Public university
3	Language and communication	6	Public university

TABLE V. BACKGROUND OF THE 11 EXPERTS INVOLVED IN THE FDM ANALYSIS

Expert no.	Field expertise	Experience (years)	Organisation
1	Early childhood education	6	Public University
2	Critical thinking skills	13	Public University
3	Early childhood education	20	Public University
4	Early childhood education	18	Other governmental agencies
5	Early childhood education	18	Other governmental agencies
6	Computer science (Coding)	21	Other government bodies
7	Computer science (Coding)	20	Other government bodies
8	Computer science (Coding)	20	Other government bodies
9	Computer science (Coding)	23	Other government bodies
10	Computer science (Coding)	15	Other government bodies
11	Information and Communication Technology (ICT)	7	Other government bodies

The questionnaire was distributed to 11 selected experts in fields related to the study [13], and the results were analysed using FDM. According to [13, 69-72], the appropriate number of experts for FDM is between 10 and 50. The experts were selected based on their expertise in the study context [73] and their work experience of over five years [74]. This study involved nine experts from government universities and two from other government bodies in Malaysia. This wide range of specialists guarantees a thorough comprehension of the topic by utilising both scholarly and real-world perspectives. Their diverse backgrounds give the study's conclusions a well-rounded viewpoint. Table V lists the demographic information of the selected experts.

The developed questionnaire facilitated the determination of pre coding elements and activities using the FDM. This FDM method was carried out in a face-to-face workshop attended by all 11 selected experts and consisted of two rounds. The first round aims at achieving expert consensus on the pre coding elements and their priority positions. The second round focuses on expert consensus in the context of the activities related to the classified elements and their respective priorities. The data were collected after each round and analysed using FDM.

1) *Round 1: Expert Consensus on Pre coding Elements and Priority Ranking of the Identified Elements:* All 11 experts participated in the first round to identify, evaluate, and confirm the pre coding elements. They discussed the pre coding element determination questionnaire to develop a suitable pre coding program model for kindergarten children. Based on the expert agreement, the elements were improved during the discussion by adding several new elements and removing those deemed irrelevant. All elements (added, rejected, or retained) aligned with the agreement and consensus reached by all experts during the FDM workshop.

The discussion proceeded with a voting process by the 11 experts to determine the priority position of the pre coding elements. Individual voting was performed by marking the agreement level for all items related to the pre coding program model, as agreed during the discussion. The voting results were analysed using FDM to determine the priority ranking of the elements.

2) *Round 2: Expert Consensus on the Determination and Priority of Pre coding Activities based on the Classified Elements.*

The second round involved expert consensus regarding the determination and priority of activities according to the classified elements. All 11 experts discussed to identify the pre coding activities based on the elements classified in the questionnaire. They shared their views and opinions to assess the appropriate level of the pre coding activities classified by elements for inclusion in the pre coding program model. The pre coding activities were also modified according to the classified elements, resulting in the addition of new pre coding activities and the removal of irrelevant ones. All pre coding activities (added, rejected, or retained) aligned with the agreement and consensus of the experts in the FDM workshop.

Subsequently, individual voting was performed to reach a consensus on the priority of pre coding activities based on the elements classified for inclusion in the pre coding program model. All 11 experts voted using a 7-point Likert scale to indicate their agreement level for each item. The findings were analysed using FDM to identify the priority of pre coding activities for each element.

IV. DATA ANALYSIS

A. Conversion of the Likert Scale to the Fuzzy Scale

Table VI shows that each Likert scale item has a corresponding fuzzy scale. In the FDM analysis process, the Likert scale value was converted to fuzzy numbers using Microsoft Excel's VLOOKUP function. Fuzzy set theory [75] was applied to convert expert agreement levels into suitable fuzzy number sets. Accordingly, the Likert scale findings from the experts were translated into fuzzy values consisting of three main values: the minimum value (m1), the most reasonable value (m2), and the maximum value (m3).

B. Data Analysis using the Fuzzy Delphi Method (FDM)

FDM data analysis comprised two key components: fuzzy triangular numbering (triangular fuzzy numbers) and fuzzy evaluation (defuzzification). Both parameters are vital when deciding to accept or reject an element based on expert consensus [39]. In particular, triangular fuzzy numbers influence the threshold value (d) and the percentage of expert agreement. Meanwhile, defuzzification impacts the fuzzy score value (A), which indicates the priority position of pre coding elements and their priority for each element [67].

The Likert scale data from the 11 experts were filled into a Microsoft Excel template for the FDM analysis. The data analysis involves assigning fuzzy triangular numbers (m1 to m3), followed by the analysis of four key aspects: (i) the average value of the fuzzy scale (m1, m2, and m3), (ii) the

threshold value (d), (iii) the percentage of expert consensus for each element and pre coding activity, and (iv) fuzzy score value (A) to determine the acceptance and priority of elements and activities available through defuzzification.

1) *Fuzzy scale average value (m1, m2, and m3)*: Fig. 3 presents a triangular graph of the mean against the triangular values (m1, m2, and m3). The m1, m2, and m3 values range from 0 to 1, which corresponds to the fuzzy numbers (0,1).

2) *Threshold Value (d)*: The threshold value (d) determines the expert consensus for each item in the questionnaire [76]. Based on the fuzzy number range (0,1), the threshold value (d) is calculated using two sets of fuzzy numbers, m (m1, m2, and m3) and n (n1, n2, and n3), as shown in Formula 1:

$$d(m, n) = \sqrt{1/3 [(m1 - n1)^2 + (m2 - n2)^2 + (m3 - n3)^2]} \quad (1)$$

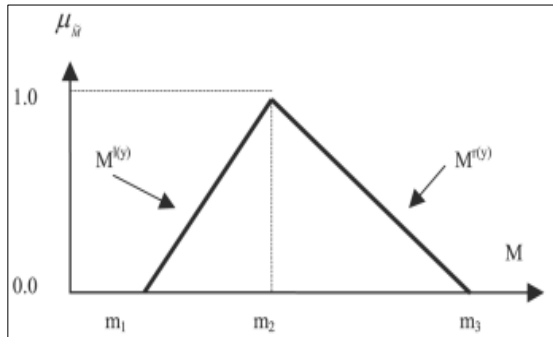


Fig. 3. Triangular graph representing the mean against the triangular values.

The data is considered to successfully reach expert agreement when the threshold value (d) is equal to or less than 0.2 [77]. Table VI describes the interpretation of the data based on the threshold value (d).

TABLE VI. DATA INTERPRETATION BASED ON THE THRESHOLD VALUE (D)

Threshold value (d)	Description	Interpretation
$d \leq 0.2$	The threshold value is equal to or less than 0.2	Accepted
$d > 0.2$	The threshold value is greater than 0.2	Rejected, or a second round may be conducted involving only experts who disagree

3) *Percentage of expert consensus*: This study also considered the percentage of expert agreement to determine the acceptance of each element and activity. An element or activity is accepted if the percentage of agreement is 75% or higher [14]. Otherwise, the element or activity is either eliminated or a second round should be conducted involving only the experts who disagreed.

4) *Fuzzy score value (A)*: The fuzzy score value (A) is obtained via defuzzification to determine the acceptance level of each item based on expert consensus. An item is accepted if its fuzzy score (A) achieves an a-cut value equal to or greater than 0.5 [78]. Formula 2 is used to calculate the fuzzy score value (A):

$$\text{Fuzzy score (A)} = (1/3) \times (m1 + m2 + m3) \quad (2)$$

In addition, the fuzzy score value (A) plays a role in determining the priority position of the pre coding elements and pre coding activities in the questionnaire. The setting of the priority position for these pre coding elements and activities is based on the results of expert discussion and agreement.

V. FINDINGS AND DISCUSSION

A face-to-face discussion in the FDM workshop involving 11 experts was conducted to evaluate and determine pre coding elements and activities for developing a suitable pre coding program model for kindergarten children. The experts successfully reached a consensus on eight elements and 24 pre coding activities; seven elements were retained, one was rejected, and two new elements were added. Next, the expert voting performed through the FDM analysis converted the Likert scale results into a fuzzy scale. The outcome showed that all eight elements and 24 pre coding activities met the conditions and reached expert consensus, where the threshold value (d) is between 0.092 and 0.204, which is < 0.2 . Table VII lists the FDM analysis element designation and pre coding activities.

TABLE VII. RESULTS OF THE FDM ANALYSIS ELEMENT DESIGNATION AND PRE CODING ACTIVITIES

Pre coding element	Number of items related to suitable pre coding activities
Algorithm element	5
Loop control structure element	3
Sequence control structure element	3
Direction indicator element	4
Error detection element	3
Decomposition element	2
Choice control structure element	2
Pattern recognition element	2

For the second condition, the study recorded over 75% of expert agreement for each element and pre coding activity, which ranged from 81.8% to 100%. Meanwhile, the third condition measures the fuzzy score value (A) to determine the acceptance level of each item, which needs to exceed 0.5. Based on the results, the fuzzy score value (A) for the pre coding elements and activities ranged from 0.788 to 0.924, proving that all pre coding elements and activities are acceptable and suitable for inclusion in the pre coding program model for kindergarten children.

The key point of this FDM analysis is its appropriateness and ability to confirm the identified pre coding elements and activities [79]. In addition, the FDM analysis assists in boosting the accuracy of pre coding elements and activities for the pre coding program model since the experts accepted all items that met the key FDM requirements based on the threshold value (d), percentage of expert agreement, and fuzzy score value (A). The results were further strengthened by the open discussions among the experts, which enabled them to present their views on the items found in the questionnaire [67]. These expert views were also considered to ensure that the results aligned with the study's context.

The success of this study stems from the proper selection of experts who shared their expertise in fields relevant to this study, including computer science with a speciality in coding skills, early childhood education, and thinking skills. The diverse pool of expertise facilitated the smooth FDM process and significantly assisted in determining the appropriate pre coding elements and activities for developing the pre coding model for kindergarten children.

The selection of FDM also influenced the quality of the study, as this method utilised the fuzzy theory to address the problem of ambiguity in data acquisition. The FDM also reduced boredom among experts and prevented data leakage during the data collection process [67] since its implementation is more organised, systematic, and shorter than traditional Delphi methods.

It should be noted that this study has a limited sample size of 11 experts. However, all selected experts have proven experience in their respective fields relevant to the study, including early childhood education, computing and meta-technology, and thinking skills. Despite the small sample size, the number of experts was sufficient, as the odd number of experts facilitated the process of reaching a consensus.

In short, this study reinforced the exceptional effectiveness of FDM [70, 80] for determining elements and activities for developing a pre coding program model suitable for kindergarten children. The strength of FDM, marked by its systematic procedure and enhanced accuracy of data analysis, particularly in reducing ambiguity, proved highly valuable for this study.

VI. CONCLUSION AND FUTURE RESEARCH

This study effectively identified key elements and activities for developing a pre coding program model suitable for kindergarten children. Based on the applied FDM approach with expert consensus, eight pre coding elements and 24 pre coding activities were deemed suitable for inclusion in the pre coding model for kindergarten children. This findings lies in several major contribution aspects.

The main contribution of this study is the development and validation of a pre coding program model specifically designed for preschool children in a systematic manner by establishing 8 pre coding elements and 24 suitable pre coding activities based on expert consensus. This study simultaneously fills an important gap in early childhood education related to CT. By emphasising pre coding (device-free activities), this study provides an accessible and non-digital-dependent approach to CT, making it highly relevant for underprivileged populations with limited access to digital. In the Malaysian context, the children from B40 families may be affected because they have limited access to digital devices, such as smartphones or laptops. So that, by integrating pre coding program model, it also helps overcome the challenges these children face in developing their CT skills.

Next, this study also contributes to encountering Malaysian kindergartens challenges during the implementation of pre coding in their teaching and learning. In other words, the pre coding program model becomes practical tools to empower

teachers in promoting the application of CT skills through pre coding activities. Directly, this can solve the problem of teachers in Malaysia who do not have knowledge about pre coding and some may not have even heard of the concept of pre coding [81]. However, it is not surprising if some teachers may still face issues in integrating pre coding into their instruction, even after being provided with a comprehensive model for guidance. Applying teachers' knowledge and enthusiasm for pre coding should be the primary priority in order to address this. After that, give teachers who are proficient in pre coding ongoing training.

This study also supports the national agenda of the country. The position pre coding as a new approach in early childhood education, align with the policies and objectives of IR4.0 and contribute to the achievement of Malaysia's SDG targets. Therefore, the application of inclusive quality education through pre coding activities is well-suited to the principles of Educational Sustainable Development (ESD), which seeks to meet current needs without compromising the ability of future generations to do the same. Apart from that, implementing the pre coding program for kindergarten children aligns with the Malaysian government's objective of fostering cognitive skills, such as logical thinking, problem-solving, and creative thinking in young learners, which are critical for navigating future technological landscapes and preparing the community for the demands of IR4.0.

Additionally the use of ranking-based elements is an important topic that requiring addressing. While this model contains elements that were outlined based on expert consensus, ranking may not be necessary when implementing activities related to these elements, as readiness and appropriateness are crucial in children's learning. Nevertheless, researchers recommend that algorithm elements be first introduced to children because they represent the most essential elements in coding. Algorithms offer a step-by-step implementation procedure, making it easier for children to understand and engage with coding concepts [16, 21, 24].

Finally, future researchers may verify the model experimentally. Test the efficacy of the suggested pre coding program model in enhancing kindergarten children CT abilities through experimental research in actual classroom environments. Then compare the results with those of other pre coding and coding techniques currently in use to evaluate the relative efficacy of the model. In the other hand, other research may create and execute pre coding pedagogy whereas focused on teacher training programs, making sure that instructors have the know-how to carry out the curriculum successfully. The most important part here is the researcher need to look into how these training sessions affected the teachers' self-assurance, comprehension, and pre coding classroom habits. Lastly integration pre coding with other learning domains such as mathematics, sciences, and language.

In conclusion, this research significantly contributes to transforming early childhood education by integrating computational thinking skills using pre coding activities and promotes an inclusive, useful, and creative approach to contemporary learning issues.

ACKNOWLEDGMENT

This study was funded by the Ministry of Higher Education under the Fundamental Research Grant Scheme (FRGS) with the number FRGS/1/2022/SSI07/UPSI/03/7.

REFERENCES

- [1] F. K. Cansu and S. K. Cansu, "An overview of computational thinking," *Int. J. Comput. Sci. Educ. Schools*, vol. 3, no. 1, pp. 17–30, 2019. <https://doi.org/10.21585/ijcses.v3i1.53>.
- [2] P. Curzon, J. Waite, K. Maton, and J. Donohue, "Using semantic waves to analyse the effectiveness of unplugged computing activities," in *WiPSCE '20: Workshop in Primary and Secondary Computing Education*, 2020. <https://doi.org/10.1145/3421590.3421606>.
- [3] B. Maraza-Quispe, A. Maurice, O. Melina, L. Marianela, L. Henry, W. Cornelio, and L. Ernesto, "Towards the development of computational thinking and mathematical logic through Scratch," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 2, 2021. <https://doi.org/10.14569/ijacsa.2021.0120242>.
- [4] X. Li, G. Sang, M. Valcke, and J. Van Braak, "Computational thinking integrated into the English language curriculum in primary education: A systematic review," *Educ. Inf. Technol.*, 2024. <https://doi.org/10.1007/s10639-024-12522-4>.
- [5] K. M. Yusoff, N. Sahari, T. Siti, and N. Mohd, "Validation of the components and elements of computational thinking for teaching and learning programming using the fuzzy Delphi method," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 1, 2021. <https://doi.org/10.14569/ijacsa.2021.0120111>.
- [6] M. U. Bers, "Coding, playgrounds and literacy in early childhood education: The development of KIBO robotics and ScratchJr," in *2018 IEEE Global Engineering Education Conference (EDUCON)*, 2018. <https://doi.org/10.1109/educon.2018.8363498>.
- [7] S. Grover and R. Pea, "Computational thinking: A competency whose time has come," in *Bloomsbury Academic eBooks*, 2018. <https://doi.org/10.5040/9781350057142.ch-003>.
- [8] J. M. Wing, "Computational thinking," *Commun. ACM*, vol. 49, no. 3, pp. 33–35, 2006.
- [9] A. Yadav and U. D. Berthelsen, "Computational thinking in education," in *Routledge eBooks*, 2021. <https://doi.org/10.4324/9781003102991>.
- [10] N. Lapawi and H. Husnin, "Investigating students' computational thinking skills on matter module," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 11, 2020. <https://doi.org/10.14569/ijacsa.2020.0111140>.
- [11] B. Zumaci and Z. Turan, "Educational robotics or unplugged coding activities in kindergartens? Comparison of the effects on pre-school children's computational thinking and executive function skills," *Think. Skills Creativity*, Article ID 101576, 2024. <https://doi.org/10.1016/j.tsc.2024.101576>.
- [12] S. K. Y. Leung, J. Wu, J. W. Li, Y. Lam, and O. Ng, "Examining young children's computational thinking through animation art," *Early Child. Educ. J.*, 2024. <https://doi.org/10.1007/s10643-024-01694-w>.
- [13] M. Adler and E. Ziglio, *Gazing into The Oracle: The Delphi Method and Its Application to Social Policy and Public Health*, London: Jessica Kingsley Publishers, 1996.
- [14] Y. Lin, H. Liao, S. Weng, and W. Dong, "Comparing the effects of plugged-in and unplugged activities on computational thinking development in young children," *Educ. Inf. Technol.*, 2023. <https://doi.org/10.1007/s10639-023-12181-x>.
- [15] D. P. McLennan, "Creating coding stories and games," *Teach. Young Child.*, vol. 10, no. 3, pp. 18–21, 2017. Retrieved from <https://www.naeyc.org/resources/pubs/tyc/feb2017/creating-coding-stories-and-games>.
- [16] S. Metin, "Activity-based unplugged coding during the pre-school period," *Int. J. Technol. Des. Educ.*, 2022. <https://doi.org/10.1007/s10798-020-09616-8>.
- [17] B. Somuncu and D. Aslan, "Effect of coding activities on pre-school children's mathematical reasoning skills," *Educ. Inf. Technol.*, 2022. <https://doi.org/10.1007/s10639-021-10618-9>.
- [18] S. Papavlasopoulou, M. N. Giannakos, and L. Jaccheri, "Exploring children's learning experience in constructionism-based coding activities through design-based research," *Comput. Human Behav.*, vol. 99, pp. 415–427, 2019. <https://doi.org/10.1016/j.chb.2019.01.008>.
- [19] C. Montuori, F. Gambarota, G. Altoé, and B. Arfé, "The cognitive effects of computational thinking: A systematic review and meta-analytic study," *Comput. Educ.*, vol. 210, Article ID 104961, 2024. <https://doi.org/10.1016/j.compedu.2023.104961>.
- [20] S. Y. Lye and J. H. L. Koh, "Review on teaching and learning of computational thinking through programming: What is next for K-12?" *Comput. Human Behav.*, vol. 41, pp. 51–61, 2014. <https://doi.org/10.1016/j.chb.2014.09.012>.
- [21] J. Lee and J. Junoh, "Implementing Unplugged Coding Activities in Early Childhood Classrooms," *Early Child. Educ. J.*, 2019. <https://doi.org/10.1007/s10643-019-00967-z>.
- [22] D. Wang, C. Zhang, and H. Wang, "T-Maze: A tangible programming tool for children," in *Proc. 10th Int. Conf. Interaction Design Children*, June 20–23, 2011, pp. 127–135.
- [23] A. Ç. Kirçali and N. Özdener, "A comparison of plugged and unplugged tools in teaching algorithms at the K-12 level for computational thinking skills," *Technol. Knowl. Learn.*, 2023. <https://doi.org/10.1007/s10758-021-09585-4>.
- [24] J. Lee, C. Joswick, and K. Pole, "Classroom play and activities to support computational thinking development in early childhood," *Early Child. Educ. J.*, 2022. <https://doi.org/10.1007/s10643-022-01319-0>.
- [25] M. Fleer, "Collective imagining in play," in *Children's Play and Development: Cultural Historical Perspectives*, pp. 73–87, 2013.
- [26] G. Futschek and J. Moschitz, "Developing algorithmic thinking by inventing and playing algorithms," in *Proc. 2010 Constructionist Approaches to Creative Learning, Thinking and Education: Lessons for the 21st Century (Constructionism 2010)*, pp. 1–10.
- [27] A. Devisakti, M. Muftahu, and H. Xiaoling, "Digital divide among B40 students in Malaysian higher education institutions," *Educ. Inf. Technol.*, vol. 29, no. 2, pp. 1857–1883, 2024. <https://doi.org/10.1007/s10639-023-11847-w>.
- [28] S. L. Mason and P. J. Rich, "Preparing elementary school teachers to teach computing, coding, and computational thinking," *Contemp. Issues Technol. Teach. Educ.*, vol. 19, no. 4, pp. 790–824, 2019.
- [29] P. J. Rich, R. A. Larsen, and S. L. Mason, "Measuring teacher beliefs about coding and computational thinking," *J. Res. Technol. Educ.*, pp. 1–21, 2020. <https://doi.org/10.1080/15391523.2020.1771232>.
- [30] X. C. Wang, Y. Choi, K. Benson, C. Eggleston, and D. Weber, "Teacher's role in fostering preschoolers' computational thinking: an exploratory case study," *Early Educ. Dev.*, vol. 32, no. 1, pp. 26–48, 2021. <https://doi.org/10.1080/10409289.2020.1759012>.
- [31] A. Strawhacker, M. Lee, and M. U. Bers, "Teaching tools, teachers' rules: Exploring the impact of teaching styles on young children's programming knowledge in ScratchJr," *Int. J. Technol. Des. Educ.*, vol. 28, no. 2, pp. 347–376, 2017. <https://doi.org/10.1007/s10798-017-9400-9>.
- [32] M. Çetin and H. Ö. Demircan, "Empowering technology and engineering for STEM education through programming robots: a systematic literature review," *Early Child Dev. Care*, vol. 190, no. 9, pp. 1323–1335, 2018. <https://doi.org/10.1080/03004430.2018.1534844>.
- [33] S. E. Jung and E. Won, "Systematic review of research trends in robotics education for young children," *Sustainability*, vol. 10, no. 4, p. 905, 2018. <https://doi.org/10.3390/su10040905>.
- [34] X. C. Wang, "Fostering young children's computational thinking: A systematic review," presented at *Early Childhood Educational Research Workshop*, Wenzhou University, Wenzhou, China, 2019.
- [35] B. Zhong and L. Xia, "A systematic review on exploring the potential of educational robotics in mathematics education," *Int. J. Sci. Math. Educ.*, vol. 18, no. 1, pp. 79–101, 2018. <https://doi.org/10.1007/s10763-018-09939-y>.
- [36] K. Brennan and M. Resnick, "New frameworks for studying and assessing the development of computational thinking," in *Proc. 2012 Annual Meeting of the American Educational Research Association*, Vancouver, Canada, pp. 1–25, 2012.

- [37] D. Kalyenci, Ş. Metin, and M. Başaran, "Test for assessing coding skills in early childhood," *Educ. Inf. Technol.*, 2022. <https://doi.org/10.1007/s10639-021-10803-w>.
- [38] M. F. Küçükara and P. Aksüt, "An example of unplugged coding education in pre-school period: Activity-based algorithm for problem-solving skills," *J. Inquiry Based Activities*, vol. 11, no. 2, pp. 81–91, 2021.
- [39] E. Polat and R. M. Yilmaz, "Unplugged versus plugged-in: examining basic programming achievement and computational thinking of 6th-grade students," *Educ. Inf. Technol.*, 2022. <https://doi.org/10.1007/s10639-022-10992-y>.
- [40] J. Del Olmo-Muñoz, R. Cózar-Gutiérrez, and J. A. González-Calero, "Computational thinking through unplugged activities in early years of Primary Education," *Comput. Educ.*, vol. 150, Article ID 103832, 2020. <https://doi.org/10.1016/j.compedu.2020.103832>.
- [41] J. Lee, "Coding in early childhood," *Contemp. Issues Early Child.*, Article ID 146394911984654, 2019. <https://doi.org/10.1177/1463949119846541>.
- [42] A. Csizmadia, B. Standl, and J. Waite, "Integrating the constructionist learning theory with computational thinking classroom activities," *Informatics Educ.*, vol. 18, no. 1, pp. 41–67, 2019. <https://doi.org/10.15388/infedu.2019.03>.
- [43] L. Leifheit, J. Jabs, M. Ninaus, K. Moeller, and K. Ostermann, "Programming unplugged: An evaluation of game-based methods for teaching computational thinking in primary school," in *ECGBL 2018 12th European Conference on Game-Based Learning*, Academic Conferences and Publishing Limited, p. 344, 2018.
- [44] T. Bell and J. Vahrenhold, "CS unplugged—How is it used, and does it work?" in *Adventures Between Lower Bounds and Higher Altitudes: Essays Dedicated to Juraj Hromkovič on the Occasion of His 60th Birthday*, pp. 497–521, 2018.
- [45] A. Nurhopipah, J. Suhaman, and M. T. Humanita, "Pembelajaran ilmu komputer tanpa komputer (unplugged activity) untuk melatih keterampilan logika anak," *J. Masyarakat Mandiri*, vol. 5, no. 5, pp. 2603–2614, 2021. <https://doi.org/10.31764/jmm.v5i5.5825>.
- [46] C. P. Brackmann, M. Román-González, G. Robles, J. Moreno-León, A. Casali, and D. Barone, "Development of computational thinking skills through unplugged activities in primary school," in *Proc. 12th Workshop on Primary and Secondary Computing Education*, pp. 65–72, 2017.
- [47] E. N. Caeli and A. Yadav, "Unplugged approaches to computational thinking: A historical perspective," *TechTrends*, vol. 64, pp. 29–36, 2020. <https://doi.org/10.1007/s11528-019-00410-5>.
- [48] E. Relkin, L. de Ruiter, and M. U. Bers, "TechCheck: Development and validation of an unplugged assessment of computational thinking in early childhood education," *J. Sci. Educ. Technol.*, vol. 29, no. 4, pp. 482–498, 2020. <https://doi.org/10.1007/s10956-020-09831-x>.
- [49] M. Fleer, "Collective imagining in play," in *Children's Play and Development: Cultural Historical Perspectives*, pp. 73–87, 2013.
- [50] R. Mohd Kusnan, N. H. Tarmuji, and M. K. Omar, "Sorotan Literatur Bersistematik: Aktiviti Pemikiran Komputasional dalam Pendidikan di Malaysia," *Malays. J. Soc. Sci. Humanit.*, vol. 5, no. 12, pp. 112–122, 2020. <https://doi.org/10.47405/mjssh.v5i12.581>.
- [51] K. Murcia, E. Cross, S. Mennell, J. Seitz, and D. Sabatino, "How to code a sandcastle: Fostering children's computational thinking through an unplugged coding experience," *J. Innov. Adv. Methodol. STEM Educ.*, vol. 1, no. 1, pp. 1–12, 2024. https://so13.tci-thaijo.org/index.php/j_iamstem.
- [52] Ü. Demir, "The effect of computer-free coding education for special education students on problem-solving skills," *Int. J. Comput. Sci. Educ. Schools*, vol. 4, no. 3, pp. 3–30, 2021. <https://doi.org/10.21585/ijcses.v4i3.95>.
- [53] A. F. Monteiro, M. Miranda-Pinto, and A. J. Osório, "Coding as literacy in pre-school: A case study," *Educ. Sci.*, vol. 11, no. 5, Article ID 198, 2021. <https://doi.org/10.3390/educsci11050198>.
- [54] Economic Planning Unit, Prime Minister's Department, "National Fourth Industrial Revolution (4IR) Policy," 2019. <https://www.ekonomi.gov.my/sites/default/files/2021-07/National-4IR-Policy.pdf>.
- [55] S. Çiftci and A. Bildiren, "The effect of coding courses on the cognitive abilities and problem-solving skills of pre-school children," *Comput. Sci. Educ.*, vol. 30, no. 1, pp. 3–21, 2019. <https://doi.org/10.1080/08993408.2019.1696169>.
- [56] V. Y. A. Prastika, N. Riyadi, and N. Siswanto, "Analysis of mathematical creative thinking level based on logical mathematical intelligence," *J. Phys. Conf. Ser.*, vol. 1796, Article ID 012011, 2021. <https://doi.org/10.1088/1742-6596/1796/1/012011>.
- [57] Y. Gülbahar, S. B. Kert, and F. Kalelioğlu, "The Self-Efficacy Perception Scale for Computational Thinking Skill: Validity and Reliability Study," *Türk Bilgisayar Ve Matematik Eğitimi Dergisi*, 2018. <https://doi.org/10.16949/turkbilmata.385097>.
- [58] A. Saxena, C. K. Lo, K. F. Hew, and G. K. W. Wong, "Designing unplugged and plugged activities to cultivate computational thinking: An exploratory study in early childhood education," *Asia-Pac. Educ. Res.*, vol. 29, no. 1, pp. 55–66, 2020. <https://doi.org/10.1007/s40299-019-00478-w>.
- [59] B. Çabuk, G. Afacan Adanir, and Y. Gülbahar, "How to teach coding through stories in early childhood classrooms," in *CTE-STEM 2022 Conference*, 2022. <https://doi.org/10.34641/ctestem.2022.449>.
- [60] Y. Lin, H. Liao, S. S. Weng, and D. Wang, "Comparing the effects of plugged-in and unplugged activities on computational thinking development in young children," *Educ. Inf. Technol.*, 2023. <https://doi.org/10.1007/s10639-023-12181-x>.
- [61] E. Relkin, L. E. de Ruiter, and M. U. Bers, "Learning to code and the acquisition of computational thinking by young children," *Comput. Educ.*, vol. 169, Article ID 104222, 2021. <https://doi.org/10.1016/j.compedu.2021.104222>.
- [62] M. Heikkilä and L. Mannila, "Debugging in programming as a multimodal practice in early childhood education settings," *Multimodal Technol. Interact.*, vol. 2, no. 3, Article ID 42, 2018. <https://doi.org/10.3390/mti2030042>.
- [63] F. L. K. Samudin and T. Y. Meng, *Sains Komputer Tingkatan 1*, Kementerian Pendidikan Malaysia, 2016.
- [64] C. S. Geck, Y. K. Hooi, Zaliha, and Fatimah, *Sains Komputer Tingkatan 4*, Kementerian Pendidikan Malaysia, 2016.
- [65] G. Dietz, J. Landay, and H. Gweon, "Building blocks of computational thinking: Young children's developing capacities for problem decomposition," *Cogn. Sci.*, pp. 1647–1653, 2019.
- [66] T. J. Murray, L. L. Pipino, and J. P. Van Gigch, "A pilot study of fuzzy set modification of Delphi," *Hum. Syst. Manage.*, vol. 5, no. 1, pp. 76–80, 1985.
- [67] M. R. Mohd Jamil and N. Mat Noh, *Kepelbagaian Metodologi Dalam Penyelidikan Reka Bentuk Dan Pembangunan*, Selangor: Qaisar Prestige Resources, 2020.
- [68] Russo, G. M., Tomei, P. A., Serra, B., & Mello, S. (2021). Differences in the use of 5-or 7-point likert scale: an application in food safety culture. *Organizational Cultures*, 21(2), 1.
- [69] A. L. Delbecq, A. H. Van de Ven, and D. H. Gustafson, *Group Techniques for Program Planning: A Guide to Nominal Group and Delphi Processes*, Scott, Foresman, 1975. <http://eduk.info/xmlui/handle/11515/11368>.
- [70] N. Yusof, N. L. Hashim, and A. Hussain, "A review of fuzzy Delphi method application in human-computer interaction studies," *AIP Conf. Proc.*, 2022. <https://doi.org/10.1063/5.0094417>.
- [71] C. Harteis, "Delphi-technique as a method for research on professional learning," in *Methods for Researching Professional Learning and Development*, Springer, Cham, 2022, pp. 351–371. https://link.springer.com/chapter/10.1007/978-3-031-08518-5_16.
- [72] C. Okoli and S. D. Pawlowski, "The Delphi method as a research tool: An example, design considerations and applications," *Inf. Manage.*, vol. 42, no. 1, pp. 15–29, 2004. <https://doi.org/10.1016/j.im.2003.11.002>.
- [73] J. Nworie, "Using the Delphi Technique in Educational Technology Research," *TechTrends*, vol. 55, no. 5, pp. 24–30, 2011. <https://doi.org/10.1007/s11528-011-0524-6>.
- [74] D. C. Berliner, "Describing the behavior and documenting the accomplishments of expert teachers," *Bull. Sci. Technol. Soc.*, vol. 24, no. 3, pp. 200–212, 2004. <https://doi.org/10.1177/0270467604265535>.

- [75] L. A. Zadeh, "Fuzzy sets and systems," *Int. J. Gen. Syst.*, vol. 17, no. 2–3, pp. 129–138, 1990. <https://doi.org/10.1080/03081079008935104>.
- [76] J. Valenzuela, "How to develop computational thinkers," 2020. <https://www.iste.org/explore/how-develop-computational-thinkers>.
- [77] C.-H. Cheng and Y. Lin, "Evaluating the best main battle tank using fuzzy decision theory with linguistic criteria evaluation," *Eur. J. Oper. Res.*, vol. 142, no. 1, pp. 174–186, 2002. [https://doi.org/10.1016/s0377-2217\(01\)00280-6](https://doi.org/10.1016/s0377-2217(01)00280-6).
- [78] S. Bodjanova, "Median alpha-levels of a fuzzy number," *Fuzzy Sets Syst.*, vol. 157, no. 7, pp. 879–891, 2006.
- [79] S. Saedah, T. L. A. Muhammad Ridhuan, and M. R. Rozaini, *Pendekatan Penyelidikan Rekabentuk dan Pembangunan (PRP): Aplikasi kepada Penyelidikan Pendidikan*, Tanjung Malim, Perak: Universiti Pendidikan Sultan Idris (UPSI), 2020.
- [80] A. F. R. L. De Hierro, M. Sánchez, D. Puente-Fernández, R. Montoya-Juárez, and C. Roldán, "A Fuzzy Delphi consensus methodology based on a fuzzy ranking," *Mathematics*, vol. 9, no. 18, Article ID 2323, 2021. <https://doi.org/10.3390/math9182323>.
- [81] Rahman, S. N., Jamil, N., Rani, I. F. A., Basir, J. M., & Omar, R. (2024). Perspective of Private Kindergarten Teachers on Pre coding Program in Early Childhood Education/Pandangan Guru Tadika Swasta Terhadap Program Pre coding dalam Pendidikan Awal Kanak-Kanak. *Sains Humanika*, 16(3), 145-152. <https://doi.org/10.11113/sh.v16n3.2127>

A Novel Internet of Things and Cloud Computing-Driven Deep Learning Framework for Disease Prediction and Monitoring

Bo GUO*, Lei NIU

School of Computer and Information Engineering, Fuyang Normal University, Fuyang, 236037, China

Abstract—In smart cities, the e-healthcare systems aided by Internet of Things (IoT) technologies play a significant role in proficient health monitoring services. The sensitivity and number of users in health networks highlights the necessity of treating security attacks. In the era of rapid internet connectivity and cloud computing services, patient medical information is most sensitive, and its electronic representation poses privacy and security concerns. Moreover, it is challenging for the traditional classifier to process a massive amount of health data and classify patients' health statuses. To address this matter, this paper presents a novel healthcare model, IoT-CDLDPM, to estimate patients' disease levels using original data and fuzzy entropy extracted from patients' remote locations. IoT-CDLDPM incorporates a deep learning classifier to analyze extensive patient-related data and provides efficient and accurate health status predictions. Furthermore, the proposed model presents the secured storage structure of the individual's health data in cloud servers. To give the authenticity of the health data, two new cryptographic algorithms are presented that encrypt and decrypt the data securely transmitted through the network. A comparison with existing methods reveals that the proposed system significantly reduces computation time, with a recorded time of 0.5 seconds, outperforming DSVS, PP-ESAP, and DRDA by up to 80%. Furthermore, the proposed cryptographic model enhances security levels, achieving a range between 99.4% and 99.8% across multiple experimental setups, surpassing other widely used encryption algorithms such as AES, RSA, and ECC-DH.

Keywords—IoT-driven healthcare; deep learning; fuzzy entropy; secure data storage; cryptography

I. INTRODUCTION

The convergence of cutting-edge technologies has recently led to revolutionary changes in the healthcare sector. Among these, the Internet of Things (IoT) stands out as a pivotal paradigm, transforming health monitoring and management [1, 2]. IoT denotes a network of connected items and sensors communicating seamlessly over the Internet, facilitating real-time data gathering and dissemination [3]. In healthcare, IoT enables the creation of smart environments where medical devices, wearables, and sensors collaborate to gather patient-specific information [4, 5]. This interconnectedness empowers healthcare professionals with timely and comprehensive data, fostering more accurate diagnostics, personalized treatments, and efficient disease management [6].

Cloud computing has become a cornerstone in reshaping healthcare systems infrastructure. The cloud offers a flexible

and centralized system for keeping and managing vast healthcare data [7]. It provides the flexibility to access information from anywhere, at any time, facilitating seamless collaboration among healthcare providers and enabling the delivery of telemedicine services [8]. Moreover, the cloud's robust storage capabilities alleviate the burden of data management, ensuring the security and accessibility of patient records [9]. Complementing these advancements, deep learning, a branch of artificial intelligence, has proven instrumental in deciphering intricate patterns within voluminous datasets [10]. Deep learning algorithms recognize complex relationships in healthcare data, making them particularly adept at disease prediction and classification tasks [11, 12]. Leveraging deep learning within the healthcare domain enhances the accuracy of diagnostics and prognostics, leading to the advent of precision medicine [13]. This paper explores the synergistic integration of IoT, cloud computing, and deep learning in designing a novel healthcare monitoring system, addressing the challenges posed by disease prediction and data security in the era of digital health [14].

This paper makes several noteworthy contributions to healthcare monitoring and data security. First, it introduces a novel, robust, and secure storage algorithm designed to maintain the consistency and safety of data stored in cloud databases. Second, the study proposes an innovative deep learning framework to predict health statistics collected via IoT sensors. Third, an encryption scheme is presented to securely protect the stored data, complemented by a corresponding decryption algorithm for accurate data retrieval. Additionally, the paper introduces intelligent fuzzy rules, contributing to effective decision-making based on medical IoT data.

Moreover, this research presents a new formula for ranking patient data, enhancing the prioritization of critical health information. The study also implements spatial and temporal constraints on a Convolutional Neural Network (CNN) classifier, refining its ability to accurately predict patients' health conditions. Finally, the paper systematically conducts various experiments to evaluate the effectiveness of the developed health-tracking approach. Collectively, these contributions advance the current understanding and capabilities in healthcare data security, predictive analytics, and decision-making within the context of IoT-enabled health monitoring systems.

This article is divided into several sections. Section 2 delves into the backgrounds, offering an in-depth exploration of the

contextual foundations relevant to the study. Section 3 elucidates the proposed framework, outlining the intricacies of the developed healthcare monitoring system. Section 4 summarizes the experimental observations, providing a detailed assessment of the system's efficiency in various scenarios. In section 5, the paper concludes by highlighting results, implications, and future research topics.

II. BACKGROUND

Integrating IoT, cloud computing, and deep learning in healthcare requires a thorough understanding of the challenges and opportunities within the rapidly evolving digital health landscape. Table I compares publications highlighting different methods and techniques of enhancing healthcare systems through these technologies. This section offers insight into the existing state of healthcare systems, emphasizing the growing reliance on interconnected devices, the significance of secure data storage, and the pivotal role of advanced data analytics in disease prediction and monitoring.

TABLE I. AN OVERVIEW OF RELATED WORKS

Study	Objective	Key techniques	Performance metrics
[15]	Identify and trace cyber-attack events in IoT networks	Network data flow extraction, PSO for deep learning parameter optimization, PSO-based DNN	Superior performance in detecting and tracing cyber-attacks
[16]	Early detection of thyroid infections	Fog computing, AI, ensemble-based classifier, encryption and decryption	Accuracy, precision, specificity, sensitivity, F1 score
[17]	Remote patient monitoring to reduce hospital visits	IoT, AI, NN configuration optimization, IoT protocols for data transmission	Not specified
[18]	Enhance privacy-preserving healthcare systems	Fog-enabled model, CNN with Bi-LSTM, Medical Entity Recognition, delta sanitizer	Recall, precision, F1-score, utility preservation
[19]	Detection of cardiovascular diseases	IoT, deep learning, BiLSTM for feature extraction, AFO for hyperparameter optimization, FDNN classifier	Accuracy (maximum 93.4%)

In the era of ubiquitous IoT technologies, everyday devices seamlessly connect to the Internet, delivering intelligent functions and on-demand capabilities to users. Despite their lightweight structure and low power consumption, these devices often expose themselves to cyber risks, adversely impacting their functionality within network systems. A significant challenge in securing IoT networks revolves around identifying and tracking sources of cyber-attack events, particularly in the context of obfuscated and encrypted network traffic.

Addressing this challenge, Koroniotis, et al. [15] have developed a novel forensic network methodology known as the Particle Deep Framework (PDF). This framework delineates

the phases of digital investigation aimed at detecting and monitoring malicious activities within IoT systems. The PDF introduces three distinctive features: the extraction of data flow patterns and verification of data integrity, tailored explicitly for encrypted networks; the utilization of a Particle Swarm Optimization (PSO) algorithm for the adaptive tuning of deep learning variables; and the design of a Deep Neural Network (DNN) utilizing PSO, designed to identify and monitor anomalies within IoT networks associated with home automation. To assess the efficacy of the presented PDF, evaluations are conducted using UNSW_NB15 and Bot-IoT sources, and comparative analyses are performed using different deep learning algorithms. The test outcomes underscore the superior ability of the PDF in detecting and tracing cyber-attack events compared to alternative strategies.

Various physiological activities are regulated by the thyroid gland, a crucial organ of the endocrine system. These processes include building proteins, energy metabolism, and hormone response. Accurate characterization and reconstruction of the thyroid are essential for detecting thyroid conditions, as alterations in the gland's shape and size indicate potential health issues. Understanding the origins and progression of thyroid diseases is paramount, necessitating focused research in this domain. The intersection of IoT, artificial intelligence, and cloud computing offers immediate computation capabilities with diverse applications in the healthcare sector. Machine learning algorithms are increasingly used in critical decisions. Individuals with thyroid conditions require a reliable and time-sensitive Quality of Service (QoS) framework.

Singh, et al. [16] have innovatively integrated artificial intelligence and fog computing into intelligent healthcare, establishing a reliable mechanism for quickly diagnosing thyroid infections. A novel ensemble-based classifier is introduced for identifying thyroid patients, utilizing UCI datasets, and simulations are conducted using Python programming. In addition to detection accuracy, the proposed framework emphasizes security through authentication and encryption. The effectiveness of the proposed framework is comprehensively assessed for power, RAM, and bandwidth usage. Simultaneously, the potential classifier's effectiveness is evaluated based on F1 score, sensitivity, specificity, precision, and accuracy. The results demonstrate that the developed methodology and classifier significantly outperform traditional methods in addressing thyroid disease detection complexities.

Singh, et al. [17] have developed e-health tools and telemonitoring systems to reduce hospitalizations, particularly in epidemic situations. This initiative leverages artificial intelligence and IoT to tackle these challenges effectively. This research aims to determine the most suitable and efficient configuration of hidden layers and encoding functions for a Neural Network (NN). Subsequently, the information transmitted through IoT networks is elucidated. The NN, an integral project component, scrutinizes information received from sensor data to make informed decisions. The selected condition is subsequently conveyed to the attending medical professional. This innovative tool empowers patients to independently recognize and predict illnesses, aiding healthcare professionals in remote disease detection and analysis.

Significantly, this is achieved without physical hospital visits, enhancing healthcare accessibility and efficiency.

Traditional health systems often struggle to manage vast volumes of biomedical data, leading to cloud-based storage and sharing. However, this approach introduces security challenges, particularly regarding privacy and confidentiality breaches. To address these issues, Moqurrab, et al. [18] have introduced an innovative fog-based data privacy model named "delta sanitizer", leveraging deep learning to enhance healthcare systems. The algorithm developed is built upon a Convolutional Neural Network with Bidirectional Long Short-Term Memory (Bi-LSTM) and is proficient in recognizing health-related entities. Statistical findings indicate that the delta sanitizer model surpasses existing models, achieving a recall of 91.1%, a precision of 92.6%, and an F1-score of 92%. Notably, the sanitization model demonstrates a 28.7% improvement in utility preservation compared to contemporary approaches. This underscores the efficacy of the proposed model in balancing the imperatives of privacy preservation and data utility in biomedical contexts.

Technological advances in the IoT, sensing technologies, and wearables have led to significant enhancements in healthcare quality, shifting from traditional healthcare approaches to continuous monitoring. Sensors attached to biomedical devices capture bio-signals generated by human actions, with the biomedical electrocardiogram (ECG) signal being a standard and non-invasive method for examining and diagnosing cardiovascular diseases (CVDs) rapidly. Given the challenges posed by the increasing number of patients and the diverse ECG signal patterns, computer-assisted automated diagnostic tools play a crucial role in ECG signal classification. In response to this need, Khanna, et al. [19] have introduced an innovative healthcare disease diagnosis model that integrates IoT and deep learning algorithms to analyze biomedical ECG signals. The model's primary objective is CVD detection through deep learning models of ECG signals. Bidirectional Long Short-Term Memory (BiLSTM) enhances the model's ability to extract meaningful feature vectors from ECG signals. The performance of the BiLSTM is further improved by leveraging the Artificial Flora Optimization (AFO) algorithm as a hyperparameter optimizer. A Fuzzy Deep Neural Network (FDNN) classifier assigns appropriate class labels to ECG signals. The model's accuracy is rigorously evaluated using biomedical ECG signals, and the test results confirm its superiority, achieving a maximum accuracy of 93.4%. This underscores the potential of the proposed model in advancing healthcare diagnostics through the fusion of IoT and deep learning technologies.

III. PROPOSED FRAMEWORK

Fig. 1 presents a comprehensive overview of the proposed healthcare monitoring system, comprising nine core modules: IoT devices, cloud database, temporal manager, rule base, rule manager, prediction, secure storage, decision manager, and user interface. Patient health data is captured by IoT devices and transmitted to a data collection agent, which stores the information in a cloud-based database. The user interface facilitates data retrieval from this cloud repository, enabling seamless access. Extracted data is then channeled to the decision manager for subsequent processing and secure storage. The latter incorporates a robust security framework comprising encryption, decryption, and key generation components. A novel RSA-based encryption algorithm safeguards data, while the corresponding decryption algorithm ensures data integrity. The efficient key generation algorithm underpins the entire cryptographic process. Encrypted data is persistently stored in the cloud database and can be retrieved upon user request through the user interface, with the decision manager orchestrating the process.

Patient statistics are forwarded to the prediction component for disease level estimation. This module leverages a novel deep learning architecture, the Fuzzy-Temporal Convolutional Neural Network (FTCNN), to accurately determine the severity of diseases. The temporal manager ensures data timeliness, while the spatial manager verifies patient location. The rule manager constructs and finalizes fuzzy rules stored in the rule base for subsequent disease prediction. The decision manager guides rule generation and interprets prediction outcomes, conveying results to physicians and patients. The IoT-CDLDPM framework encompasses three primary components: IoT-based data acquisition, secure data storage, and advanced disease prediction, which are discussed in this section.

Initial patient data is collected from remote locations using IoT devices tailored to specific diseases such as cancer, cardiovascular conditions, and diabetes. These devices employ specialized sensors to capture relevant patient symptoms, including glucose levels, heart rate, and electrocardiogram data. Extracted features are organized into individual patient records, each uniquely identified. The collected data is securely transferred to the cloud database through a coordinated process involving the decision manager, user panel, and data gathering component, with a secure storage component playing a critical role. The data capture component aggregates data and forwards it to the user interface, identifying essential characteristics and passing them to the decision manager for security processing. Subsequently, the decision manager transmits preprocessed data to the storage component for encryption, decryption, and secure cloud storage.

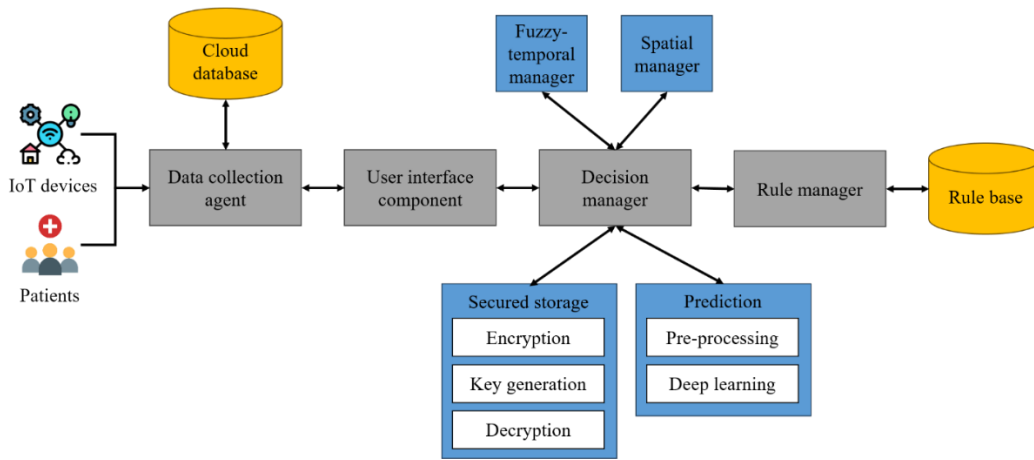


Fig. 1. An overview of the proposed healthcare monitoring system.

The proposed secure storage framework incorporates novel key generation, encryption, and decryption algorithms to safeguard medical and patient data. The initial phase involves key generation using an Elliptic Curve-based Key Generator (ECKG). This algorithm extracts a 4-bit cloud user code, partitioning it into two binary values (a and b). A prime number (p) is selected to define the Galois Field (GFp). Subsequently, the ECKG employs the Diffie-Hellman key exchange protocol to generate public keys P_A and P_B .

An additional layer of protection is introduced through the novel RSA-based Key Generator (RSAKG) to enhance security further. This algorithm derives key pairs (e and d) from specific points (q and r) on the elliptic curve. The RSAKGA process involves calculating n and U and generating public and private keys.

We employ the elliptic curve cryptography-based cyclic encryption procedure (ECC-CEP), which involves two sequential stages. The first stage utilizes elliptic curve-based encryption, while the second employs RSA-based encryption, enhancing the overall security of the process. This comprehensive approach ensures the confidentiality and integrity of the transmitted data.

To complete an entire cryptographic cycle, the suggested elliptic curve cryptography and RSA-enabled multi-decryption scheme are applied to decrypt the original text. This algorithm involves a two-stage process where the first stage utilizes RSA-based decryption, and the second involves ECC-based decryption. Integrating these cryptographic techniques establishes a robust and multilayered security framework for ensuring the privacy and integrity of healthcare data.

The predictive model comprises two principal modules to assess disease severity based on symptoms and patient feedback. The initial module utilizes a deep learning approach to analyze symptom-based severity, evaluate patient feedback textually, and estimate sentiment scores specific to individual diseases to determine their severity level. Subsequently, the model incorporates severity rating features alongside user feedback, allowing for evaluating severity-based ratings and indicating the disease status for specific data instances.

Severity classification relies on the extraction of salient features from patient data. This study proposes an MCST-CNN architecture to accurately determine disease severity and compute patient polarity scores. To refine severity categorization, Latent Dirichlet Allocation (LDA) is employed to cluster extracted severity levels. The MCST-CNN model is a specialized CNN architecture comprising four distinct channels for abnormal, medium, low, and average severity states. Dataset features are mapped to a linear matrix via a lookup function, resulting in a matrix $X \in R^{nk}$. The severity level embedding channel refines severity estimation by incorporating a 45-dimensional severity analyzer vector.

The convolutional layer is instrumental in extracting salient features from medical datasets, reports, and physician notes. By applying filters of varying sizes, this layer effectively identifies crucial attributes for feature and severity level embedding. Given a filter $wt_x \in R^{h \times k}$, where h refers to the height of matrix x embedded in a particular channel, feature extraction is performed according to Eq. (1) within defined temporal and spatial boundaries. ATT denotes an asymmetric map and b represents a bias term. The resulting attribute map, $CH_x \in R^{n-h+1}$, is calculated for a specific time interval ($t1$ to $t2$) as outlined in Eq. (2). For severity merging and attribute embedding within the embedding channel, distinct filters $wt_z \in R^{h \times 1}$ are employed to generate attribute maps as described in Eq. (3). This approach enables the generation of diverse feature representations and attributes.

$$CH_i\langle t1, t2, sp \rangle = ATT(wt_x \cdot x_{i+h} + b) \quad (1)$$

$$CH_x\langle t1, t2, sp \rangle = [CH_1^x, CH_2^x, \dots, CH_{n-h+1}^x] \quad (2)$$

$$CH_z\langle t1, t2, sp \rangle = [CH_1^z, CH_2^z, \dots, CH_{n-h+1}^z] \quad (3)$$

The pooling operation is pivotal in capturing maximum features from input values, typically expressed as shown in Eq. (4). Following this operation, the ultimate attributes are obtained by concatenating the semantically significant attributes using a filter. Typically, this process is denoted as

$$CH = CH \oplus CH \quad (5)$$

Eq. (5) illustrates the resulting final attributes, where the terms n and m denote distinct thresholds for useful and attribute-specific components, correspondingly.

$$CH_x = MAX(CH_x) \text{ and } CH_z = MAX(CH_z) \quad (4)$$

$$CH = \begin{matrix} 1 & n & 1 & m \\ CH \oplus & \dots \oplus & CH \oplus & CH \oplus \dots CH \\ x & x & z & z \end{matrix} \quad (5)$$

Typically, the softmax function is utilized to compute the final attributes. In this research, the extraction of severity levels is framed as a sequential labelling task. The resulting output is represented by Eq. (6), in which O denotes the masking function and $rs \in R^{n+m}$ signifies a sample based on the Bernoulli pattern.

$$O\langle t1, t2, sp \rangle wt. (c \ o \ rs) + b \quad (6)$$

The dataset comprising patient records encompasses a diverse range of attributes associated with severity levels, although variations in the specific attributes are relevant to each severity group. Moreover, the attributes representing severity encompass various types of severity. Hence, it becomes imperative to cluster the pertinent attributes and establish mappings between the extracted severity-related attributes and their respective counterparts. The standard Linear Discriminant Analysis (LDA) method is employed to identify the relevant characteristics from the standardized dataset. Leveraging the LDA method enables segregating specific severity levels into distinct groups. Notably, the LDA method incorporates considerations of spatial and temporal factors, representing an improvement over prior approaches.

Predicting disease severity entails clustering pertinent features and calculating polarity scores for each severity level. Severity level ratings within a rating matrix are determined by computing polarity scores corresponding to severity levels and considering the resulting polarity score. This methodology calculates the rating for each disease severity based on relevant attributes associated with the dataset. The severity level rating is computed using Equation (7), where W_k denotes the word set DS_{ij} linked to severity score a_k , and $SVL(w)$ represents the attribute polarity based on their semantic content.

$$r_{ijk} \langle t1, t2, sp \rangle = \frac{\sum_{w \in W_k(DS_{ij})} SVL(w)}{W_k(DS_{ij})} \quad (7)$$

A severity-based weight estimation process is employed to determine severity-associated attribute weights, utilizing a three-dimensional attribute-factor (AF) tensor, WT . This tensor encapsulates the intricate relationships between attributes, users, and disease severity levels. The tensor WT undergoes decomposition as outlined in Eq. (8), where R signifies the top-rank component count, and the symbol \circ represents the outer product. The column vectors within factor matrices X , Y , and Z are denoted by x_r , y_r , and z_r , respectively. The dimensions of X , Y , and Z are $I \times R$, $J \times R$, and $K \times R$, correspondingly. Eq. (9) presents an element-wise equivalent of Eq. (8).

$$wt \approx \sum_{r=1}^R x_r \circ y_r \circ z_r \quad (8)$$

$$wt_{ijk} = (x_r, y_r, z_r) = \sum_{r=1}^R x_{ir} \cdot y_{jr} \cdot z_{kr} \quad (9)$$

Each row within the matrices x_r , y_r , and z_r corresponds to weight factors associated with patients, attributes, and severity levels, respectively. Disease prediction ratings, denoted as r_{ij} , are computed using the proposed prediction model,

incorporating severity levels and weight vectors as defined in Eq. (10).

$$r_{ij} = \sum_{k=1}^K w_{ijk} \cdot r_{ijk} \quad (10)$$

IV. EXPERIMENTAL RESULTS

The proposed disease monitoring system was engineered using Java within the NetBeans Integrated Development Environment (IDE) and leveraged the CloudSim simulation toolkit for performance evaluation. The system incorporates a standardized dataset from the University of California, Irvine (UCI) Machine Learning Repository, encompassing various diseases such as cardiovascular conditions, diabetes, and cancer. Analyzing this dataset provides a user-friendly approach to assessing disease severity, thereby contributing to the prevention of life-threatening ailments.

This section provides a detailed overview of the medical datasets employed in this study, specifically focusing on heart disease, diabetes, and cancer. Furthermore, the performance metrics utilized to evaluate the proposed health monitoring system are outlined, followed by a comprehensive presentation of experimental results. The evaluation of the suggested approach is divided into two primary domains: disease prediction and safe storage. Each domain is assessed using specific evaluation parameters.

The assessment of the secured storage component within the disease prediction system encompasses factors including decryption time, encryption time, and key generation time. The formulas for computing these times are delineated in Eq. (11), (12), and (13), respectively.

$$K = DTT + ET \quad (11)$$

$$ET = EDT - STT \quad (12)$$

$$DT = EDT - STT \quad (13)$$

In Eq. (11), DTT represents the data transferring time, while ET denotes the time required to encrypt the data. The encryption time is determined by the duration of converting the original data into its encrypted form. Fig. 2 depicts the analysis of crucial generation time for the developed secured storage system. The figure illustrates the results of five experiments conducted with varying numbers of cloud users (200, 400, 600, 800, and 1000). As observed from the graph, as the number of cloud consumers increases, the key generation time also increases.

In Eq. (12), EDT represents the end time of the encryption process, while STT signifies the start time. DT denotes the user's time spent decrypting the encrypted data, measured in milliseconds and expressed as Eq. (13).

Fig. 3 presents an analysis of the proposed algorithm's encryption time. To assess the model's performance, the evaluation involved five experiments using data sets of varying sizes: 200 KB, 400 KB, 600 KB, 800 KB, and 1 MB. As expected, the encryption time exhibited a positive correlation with data size. This is likely caused by the inherent properties

of elliptic curve cryptography used in the algorithm and the two-stage nature of the encryption and decryption processes.

Similar to the encryption process, Fig. 4 analyzes the decryption time associated with the proposed secure storage algorithm. The evaluation employed the same five data set sizes to evaluate decryption efficiency. The results demonstrate that decryption time scales proportionally with the size of the data being handled. This characteristic can be attributed to the two-stage nature of the decryption scheme employed in the method.

Fig. 5 compares the computational time required by the suggested secure storage approach and several existing systems. The evaluation employed a fixed data size of 10 GB

across five scenarios. The results indicate that the proposed algorithm exhibits lower computational time than existing systems.

Fig. 6 compares the security performance of the developed secure storage algorithm (ECCRS-DDA& ECC-TSEA) with several existing algorithms. The evaluation involved five experiments designed to assess the relative security of each approach. The results demonstrate that the proposed algorithm offers a superior level of protection compared to existing solutions.

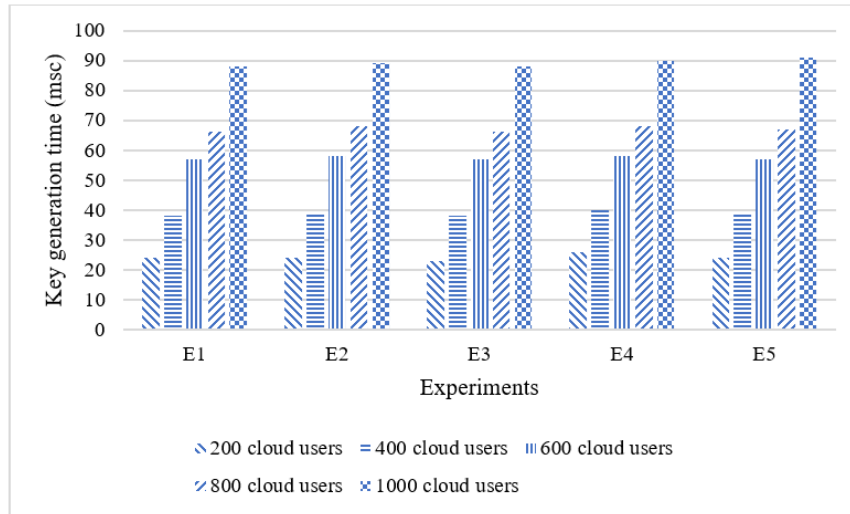


Fig. 2. Key generation time comparison.

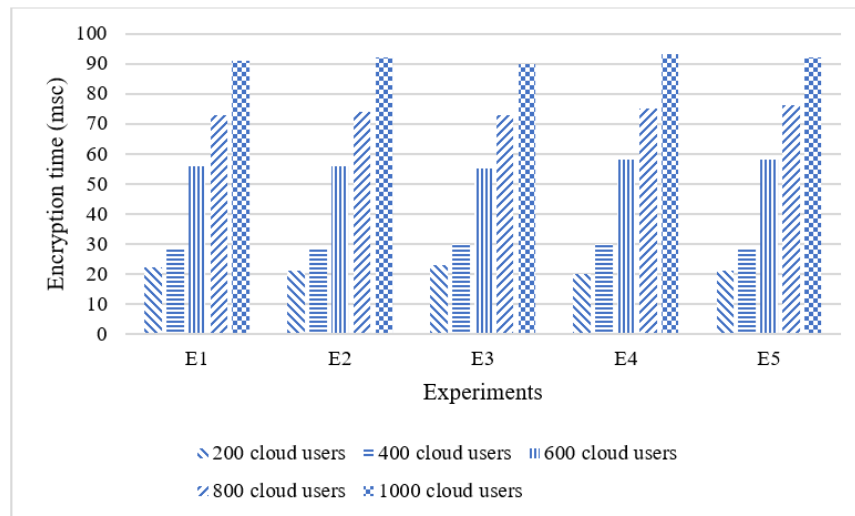


Fig. 3. Encryption time comparison.

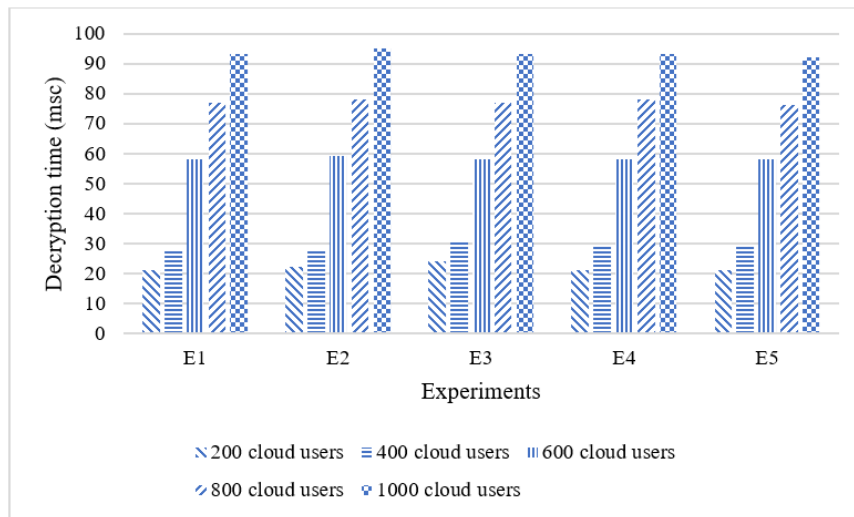


Fig. 4. Decryption time comparison.

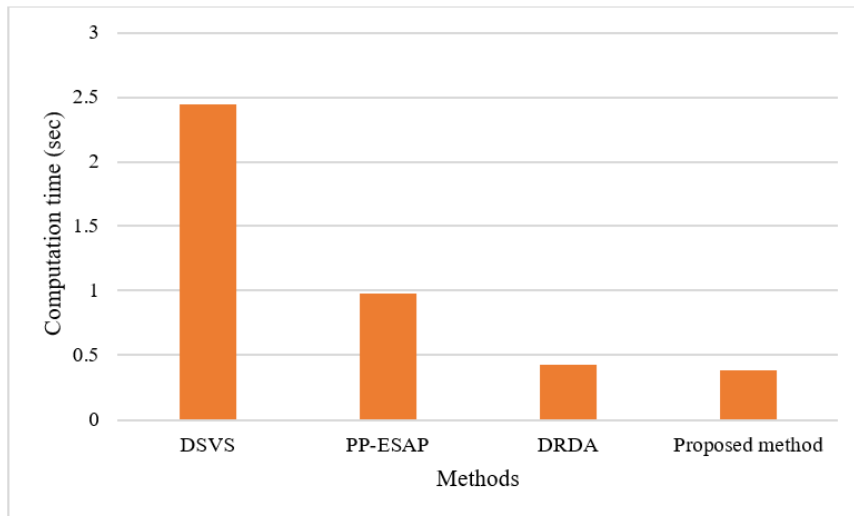


Fig. 5. Computation time comparison.

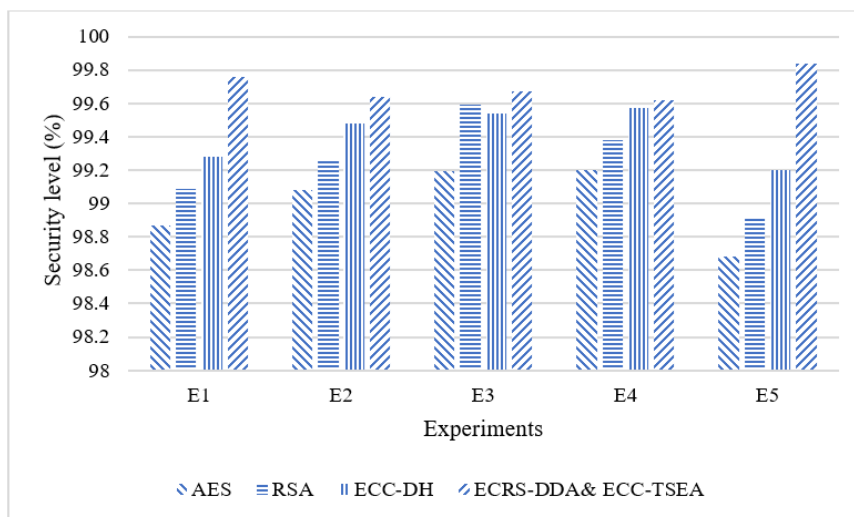


Fig. 6. Security level comparison.

V. CONCLUSION

In this research, an innovative healthcare surveillance system has been developed and deployed to assess the severity of critical illnesses, including diabetes and cardiovascular conditions. The system utilizes original data collected from patients residing in remote areas to predict disease levels. Additionally, a secure data storage model has been developed and integrated into the system to ensure the safe storage of patient data in cloud databases. Three novel algorithms have been formulated for key generation, encryption, and decryption procedures within the secure storage framework to bolster the system's security. These algorithms aim to protect sensitive patient information and prevent unauthorized access. A novel deep learning algorithm named IoT-CDLDPM has also been created and incorporated into the healthcare monitoring system. This algorithm enhances the efficiency of disease-level prediction by leveraging the power of deep learning techniques. The experimental outcomes derived from a series of trials in this study indicate the efficacy of the proposed healthcare system. The system achieved a prediction accuracy of 99.4%, demonstrating its high precision in assessing disease severity levels. Moreover, the system's security level is evaluated to be 99.7%, surpassing the performance of other healthcare systems.

REFERENCES

- [1] F. Kamalov, B. Pourghebleh, M. Gheisari, Y. Liu, and S. Moussa, "Internet of medical things privacy and security: Challenges, solutions, and future trends from a new perspective," *Sustainability*, vol. 15, no. 4, p. 3317, 2023, doi: <https://doi.org/10.3390/su15043317>.
- [2] M. Adil et al., "Healthcare internet of things: Security threats, challenges and future research directions," *IEEE Internet of Things Journal*, 2024.
- [3] B. Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy-efficient data fusion methods in the Internet of Things," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 15, p. e6959, 2022.
- [4] Z. Lu and X. Deng, "A cloud and IoT-enabled workload-aware Healthcare Framework using ant colony optimization algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 3, 2023.
- [5] V. Puri, A. Kataria, and V. Sharma, "Artificial intelligence-powered decentralized framework for Internet of Things in Healthcare 4.0," *Transactions on Emerging Telecommunications Technologies*, vol. 35, no. 4, p. e4245, 2024.
- [6] M. Riad, "IoT-based intelligent system For Alzheimer's Disease Detection & Monitoring," *International Journal of Advanced Science and Computer Applications*, vol. 3, no. 2, 2024.
- [7] V. Hayyolalam, B. Pourghebleh, A. A. P. Kazem, and A. Ghaffari, "Exploring the state-of-the-art service composition approaches in cloud manufacturing systems to enhance upcoming techniques," *The International Journal of Advanced Manufacturing Technology*, vol. 105, no. 1-4, pp. 471-498, 2019.
- [8] M. Hassan, A. Hussein, A. A. Nassr, R. Karoumi, U. M. Sayed, and M. AbdelRaheem, "Optimizing Structural Health Monitoring Systems Through Integrated Fog and Cloud Computing Within IoT Framework," *IEEE Access*, 2024.
- [9] V. Hayyolalam, B. Pourghebleh, M. R. Chehrehzad, and A. A. Pourhaji Kazem, "Single-objective service composition methods in cloud manufacturing systems: Recent techniques, classification, and future trends," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 5, p. e6698, 2022.
- [10] A. Azadi and M. Momayez, "Review on Constitutive Model for Simulation of Weak Rock Mass," *Geotechnics*, vol. 4, no. 3, pp. 872-892, 2024, doi: <https://doi.org/10.3390/geotechnics4030045>.
- [11] B. Omarov, A. Tursynova, and M. Uzak, "Deep Learning Enhanced Internet of Medical Things to Analyze Brain Computed Tomography Images of Stroke Patients," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, 2023, doi: <https://doi.org/10.14569/IJACSA.2023.0140874>.
- [12] M. D. Tezerjani, M. Khoshnazar, M. Tangestanizadeh, and Q. Yang, "A Survey on Reinforcement Learning Applications in SLAM," *arXiv preprint arXiv:2408.14518*, 2024, doi: <https://doi.org/10.48550/arXiv.2408.14518>.
- [13] S. Asif et al., "Advancements and Prospects of Machine Learning in Medical Diagnostics: Unveiling the Future of Diagnostic Precision," *Archives of Computational Methods in Engineering*, pp. 1-31, 2024.
- [14] S. Paul and C. Beulah Christalin Latha, "Machine Learning and IoT in Precision Healthcare," in *IoT and ML for Information Management: A Smart Healthcare Perspective*: Springer, 2024, pp. 201-234.
- [15] N. Koroniotis, N. Moustafa, and E. Sitnikova, "A new network forensic framework based on deep learning for Internet of Things networks: A particle deep framework," *Future Generation Computer Systems*, vol. 110, pp. 91-106, 2020.
- [16] P. D. Singh, G. Dhiman, and R. Sharma, "Internet of things for sustaining a smart and secure healthcare system," *Sustainable computing: informatics and systems*, vol. 33, p. 100622, 2022.
- [17] N. Singh, S. Sasirekha, A. Dhakne, B. S. Thrinath, D. Ramya, and R. Thiagarajan, "IOT enabled hybrid model with learning ability for E-health care systems," *Measurement: Sensors*, vol. 24, p. 100567, 2022.
- [18] S. A. Moqurrah et al., "A deep learning-based privacy-preserving model for smart healthcare in Internet of medical things using fog computing," *Wireless Personal Communications*, vol. 126, no. 3, pp. 2379-2401, 2022.
- [19] A. Khanna, P. Selvaraj, D. Gupta, T. H. Sheikh, P. K. Pareek, and V. Shankar, "Internet of things and deep learning enabled healthcare disease diagnosis using biomedical electrocardiogram signals," *Expert Systems*, vol. 40, no. 4, p. e12864, 2023.

Comparison of Artificial Neural Network and Long Short-Term Memory for Modelling Crude Palm Oil Production in Indonesia

Brodjol Sutijo Suprih Ulama*¹, Robi Ardana Putra², Fausania Hibatullah³, Mochammad Reza Habibi⁴,
Mochammad Abdilllah Nafis⁵
Department of Business Statistics, Faculty of Vocational Studies, Institut Teknologi Sepuluh Nopember,
Surabaya, Indonesia

Abstract—Indonesia is one of the largest producers and exporters of Crude Palm Oil (CPO), making CPO production crucial to the country's economic stability. Accurate forecasting of CPO production is essential for effective inventory management, export-import strategy, and economic planning. Traditional time series methods like ARIMA have limitations in modeling nonlinear data, leading to the adoption of machine learning approaches such as Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM). This study compares the performance of ANN, a general neural network, and LSTM, a neural network specifically designed for time series data, in predicting CPO production in Indonesia. Data from 2003 to 2022 were used to train and evaluate both models with various hyperparameter tuning configurations. The results indicate that while both models provide excellent forecasting accuracy, with MAPE values below 10%, the LSTM model achieved a lower out-of-sample MAPE of 5.78% compared to ANN's 6.87%, suggesting superior performance by LSTM in capturing seasonal patterns in CPO production. Consequently, LSTM is recommended as the preferred model for CPO production forecasting due to its enhanced ability to handle temporal dependencies and nonlinear patterns in the data.

Keywords—Artificial Neural Network (ANN); Crude Palm Oil (CPO); Long Short-Term Memory (LSTM)

I. INTRODUCTION

Indonesia is the largest agrarian country in Southeast Asia and one of the world's largest producers of Crude Palm Oil (CPO). According to data from the United States Foreign Agricultural Service, Indonesia's CPO production reached 47 million metric tons as of March 2024. This achievement places Indonesia as both the largest producer and exporter of crude palm oil (CPO) globally. Crude Palm Oil (CPO) is unrefined palm oil extracted from the mesocarp of the oil palm fruit, which remains in a raw state and requires further processing and refining to become pure palm oil [1]. CPO has many derivative products used in daily life, including cooking oil, margarine, biodiesel, soap, and detergent. [2].

As a primary producer in the international CPO market, crude palm oil is a major commodity that supports Indonesia's economy. According to data from the Central Statistics Agency (BPS), CPO contributes significantly and consistently to Indonesia's export value, accounting for around 12-13% during

the period from 2020 to 2022. Thus, the CPO industry makes a substantial contribution to national income through export activities. The export value plays a crucial role in the country's economic stability, as higher export values lead to an appreciation of the domestic currency. Currency appreciation can lower import prices, reducing inflationary pressures. Conversely, currency depreciation can increase import prices, triggering inflation [3]. Therefore, CPO production is a potential factor in the country's economic development.

Based on data from the Central Statistics Agency of Indonesia, CPO production fluctuated significantly from 2021 to 2022, with the lowest production occurring in February 2022, leading to a shortage of cooking oil that caused concern among the public. CPO is one of the largest contributors to the nation's export revenue. Therefore, fluctuations in CPO production can lead to instability in export income. A decline in export income could result in a trade balance deficit and depreciation of the domestic currency, which in turn may increase import prices and trigger inflation, negatively impacting the economy as a whole [3]. One way to manage and anticipate this risk is by forecasting CPO production. With accurate forecasting, the government and industry players can plan more effective export-import policies, manage inventory and prices more stably, and adjust investment strategies to reduce economic uncertainty, ensuring the optimal contribution of the palm oil sector to economic growth and national stability.

A commonly used forecasting method for modeling time series data is the Autoregressive Integrated Moving Average (ARIMA) model. ARIMA is a time series model derived from the Autoregressive Moving Average (ARMA) model, which combines non-stationary Autoregressive and Moving Average processes requiring differencing to achieve stationarity [4]. The ARIMA model is quite flexible in modeling time series data patterns because it can capture random data patterns, trends, seasonal, and even cyclical characteristics in the time series data. However, studies indicate that ARIMA is less suitable for modeling nonlinear time series data [5]. Recent advancements in technology have led to the use of machine learning methods, which are more suitable for addressing the limitations of traditional time series models. Machine learning algorithms excel in discovering and representing complex structural patterns in data, enabling better future forecasting based on these patterns [6]. Machine learning methods, especially deep

*Corresponding author

learning, have become powerful tools for handling time series data with nonlinearity or complexity factors [7]. The most commonly used deep learning method is the Artificial Neural Network (ANN). Artificial Neural Network (ANN) is a computational system inspired by the structure and operation of neural cells in the brain, modeling biological neural networks. ANN is a non-linear model with a flexible functional form and several parameters that cannot be interpreted. This characteristic enables ANN to solve unstructured and hard-to-define problems [8].

As technology advances, various deep learning methods have emerged. One deep learning method specifically designed for time series data is Long Short-Term Memory (LSTM) [9]. LSTM, a variant of Recurrent Neural Network (RNN), is widely used in time series data modeling. LSTM's architecture was developed to address the vanishing gradient problem often encountered in conventional RNNs. LSTM uses two key mechanisms, the forget gate and input gate, allowing the model to learn which information should be retained or forgotten at each time step. This advantage makes LSTM more flexible in managing its internal memory, which can be useful when working with complex or nonlinear data. Additionally, LSTM has a better ability to capture data patterns over longer periods, making it more suitable for applications requiring deeper and more accurate temporal dynamics modeling.

This study introduces a novel contribution to the field by focusing on the application and comparative analysis of two advanced neural network methods—Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM)—for forecasting crude palm oil (CPO) production in Indonesia. While ANN has been widely utilized for various predictive tasks, this research leverages the specialized capabilities of LSTM in handling sequential and time-series data, addressing the unique temporal patterns, seasonality, and variability inherent to Indonesia's CPO production. By tailoring these methods to the specific characteristics of Indonesia's agricultural sector, this study not only provides valuable insights into their effectiveness and limitations but also establishes a robust framework for improving forecasting accuracy in a critical industry that significantly impacts the country's economy.

This study introduces a novel contribution to the field by focusing on the application and comparative analysis of two advanced neural network methods—Artificial Neural Network (ANN) and Long Short-Term Memory (LSTM)—for forecasting crude palm oil (CPO) production in Indonesia. While ANN has been widely utilized for various predictive tasks, this research leverages the specialized capabilities of LSTM in handling sequential and time-series data, addressing the unique temporal patterns, seasonality, and variability inherent to Indonesia's CPO production. By tailoring these methods to the specific characteristics of Indonesia's agricultural sector, this study not only provides valuable insights into their effectiveness and limitations but also establishes a robust framework for improving forecasting accuracy in a critical industry that significantly impacts the country's economy.

II. RELATED WORK

Time series forecasting has been extensively explored using both traditional statistical models and modern machine learning

techniques. To better understand the strengths and limitations of these approaches, various studies have compared their performance across different domains. The Table I below summarizes key findings from relevant studies, highlighting the application of ARIMA, ANN, and LSTM models in diverse sectors and the insights gained from these comparisons

TABLE I. RESULTS OF CRUDE PALM OIL (CPO) PRODUCTION IN INDONESIA MODELLING WITH ANN

Domain	Key Findings	References
Various Sectors (Finance, Healthcare, Weather, Utilities)	Machine learning methods like ANN outperform ARIMA in scenarios with non-linear or complex patterns.	[10]
Energy Consumption (Commercial Buildings)	ANN is more reliable for non-linear datasets, capturing intricate dependencies and variability, while ARIMA is effective for linear and stationary data.	[11]
Global Crude Palm Oil (CPO) Prices	LSTM achieves superior accuracy (MAPE: 2.33%, RMSE: 34.708), better capturing complex temporal dependencies compared to ARIMA and hybrid models.	[12]
Economic and Financial Indicators (GDP Growth)	LSTM outperforms ARIMA in long-term forecasting for non-linear and non-stationary data, addressing ARIMA's limitations.	[13]
Stock Price Prediction (Real Estate)	LSTM provides better accuracy for non-stationary, complex, and cyclical data, while ARIMA performs better for linear and stationary datasets.	[14]

Conventional methods like ARIMA have been widely utilized for time series forecasting due to their ability to capture trends, seasonality, and cycles in data. However, ARIMA has notable limitations in addressing non-linear and complex data, such as the patterns of CPO production in Indonesia, which are influenced by various factors including seasonality, policy changes, and international price fluctuations. These studies collectively highlight the evolution of time series forecasting techniques, showcasing the growing prominence of machine learning methods like LSTM and ANN in overcoming the limitations of traditional models such as ARIMA. ANN excels in modeling non-linear and unstructured relationships, while LSTM, a variant of Recurrent Neural Networks (RNN), is specifically designed to handle temporal dynamics and long-term dependencies in time series data. However, a research gap persists in exploring the application of these models in integrated or hybrid approaches tailored for specific domains, such as agriculture and commodity price forecasting. This study aims to address this gap by applying and comparing ANN and LSTM for forecasting CPO production in Indonesia, to capture the intricate characteristics and variability of the data more effectively while providing a robust framework for planning and decision-making in the CPO industry.

III. METHODOLOGY

The methodology diagram (Fig. 1) illustrates the key stages in modeling and forecasting Crude Palm Oil (CPO) production in Indonesia using Artificial Neural Network (ANN) and Long

Short-Term Memory (LSTM) models. Initially, the data is divided into training and testing subsets, ensuring a robust evaluation of the models. To achieve this, several data splitting ratios, such as 70:30, 80:20, and 90:10, are explored to analyze the impact of data availability on the training process.

In the model design and training phase, ANN and LSTM models are configured with specific hyperparameters to capture the unique patterns in the data. These configurations include batch sizes of 8 and 16, neuron counts of 50 and 100, hidden layers of 1 and 2, epochs of 50 and 100, and learning rates of 0.001 and 0.01. The use of a six-month window size for LSTM is particularly emphasized to account for the seasonal characteristics of CPO production. Adam optimization is applied to update the model weights effectively, ensuring optimal learning from the data.

During the evaluation stage, the models are assessed using the Mean Absolute Percentage Error (MAPE) metric to validate their forecasting accuracy. The systematic exploration of data splits and hyperparameter combinations ensures a comprehensive understanding of the models' capabilities in forecasting CPO production. The final step occurs in the output gate component, where the sigmoid activation function (σ) is applied to produce the output value o_t and process the cell state (C_t) with tanh activation.

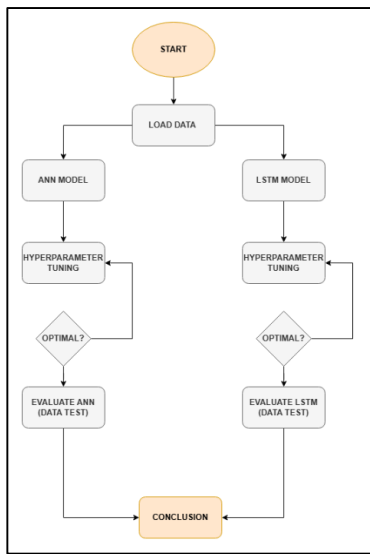


Fig. 1. Methodology diagram.

IV. ARTIFICIAL NEURAL NETWORK (ANN)

An Artificial Neural Network (ANN) is a computational system inspired by the structure and function of neural cells in the brain, essentially modeling biological neural networks. ANN is an example of a non-linear model with a flexible functional form and several parameters that cannot be interpreted, similar to parametric models. However, this allows ANN to tackle unstructured and challenging-to-define problems. The process within an ANN begins with input data received by neurons, which are grouped into layers. Information received from the input layer is passed sequentially through the layers in the ANN until it reaches the output layer. Layers positioned between the input and output are known as hidden layers. A Neural Network

is determined by three main components: the pattern of relationships between units (network architecture), the method for updating weights in connection links (training method or algorithm), and the activation function [8].

A. Components of ANN

This form of artificial intelligence is developed to mimic the workings of the human biological nervous system (neurons). Each component of an ANN system can be analogized to parts of the human neuron system, such as dendrites (parts that receive input/signals), the cell body (processes inputs), and the axon (transmits this input to other neurons/outputs) [15] (Fig. 2).

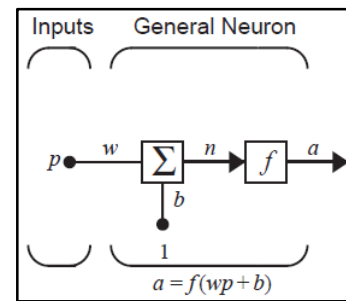


Fig. 2. Components of ANN.

Each piece of information received by the dendrites (input) is summed and sent through the axon to the dendrites (output) of another neuron. This information will only be received by the other neuron if it meets a certain threshold value. Neurons that receive information and transmit it to other neurons are considered to be activated. Neurons receive information from other neurons in the form of values known as weights, which also indicate the strength of the connection between neurons.

B. ANN Architecture

ANN architecture is the pattern of relationships between neurons. Neurons that share the same weight pattern and activation function are grouped in the same layer. Information usually flows from the input layer to the output layer, often passing through hidden layers. Some neural network architectures include:

1) Single-layer network

This network has only one layer containing weights connected to each other. There is no hidden layer in this type of network, and all input units are connected to every output unit (Fig. 3).

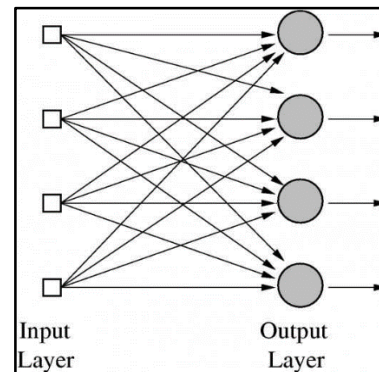


Fig. 3. Single-layer network.

2) Multi-layer network

This network has one or more layers (hidden layers and weights) between the input layer and the output layer. A multi-layer network can solve more complex problems than a single-layer network (Fig. 4).

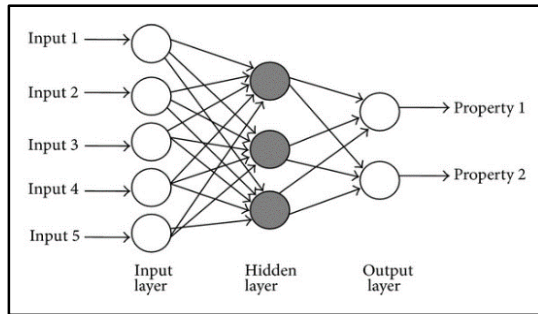


Fig. 4. Multi-layer network.

V. LONG SHORT TERM MEMORY (LSTM)

Long Short-Term Memory (LSTM) is a neural network architecture that shares similarities with Recurrent Neural Networks (RNNs) [16]. LSTM was first introduced by Hochreiter and Schmidhuber in 1997. LSTM falls within the category of Recurrent Neural Networks (RNNs), where it has repeating units that function similarly to neural network sequences [17]. LSTM was designed to overcome the limitations of conventional RNNs, specifically the vanishing gradient problem. The vanishing gradient problem occurs when gradient values decrease significantly towards the last layers, preventing weight updates and causing the model to struggle to improve or converge. Unlike conventional RNNs, LSTM uses additional signals passed from one time step to the next, known as the cell state. LSTM has strong generalization capabilities and effective learning abilities for both large and small datasets. It is particularly advantageous in processing non-linear data, which enhances forecasting accuracy [18].

LSTM's design makes it particularly suitable for solving complex problems that involve sequential or time-dependent data. For instance, in forecasting tasks where patterns such as seasonality, trends, or temporal dependencies play a critical role, LSTM excels by learning these non-linear relationships. Its strong generalization capabilities and ability to learn from both large and small datasets make it highly effective for addressing real-world challenges. By leveraging its strengths, LSTM has been successfully applied in various domains, such as financial time-series prediction, speech recognition, and production forecasting. For example, in modeling Crude Palm Oil (CPO) production, LSTM can identify seasonal trends and long-term dependencies, enabling more accurate and reliable forecasts. These features make LSTM a powerful solution for problems where conventional methods often struggle, ensuring better decision-making and planning in dynamic environments.

The LSTM architecture consists of memory cells, an input gate, a forget gate, and an output gate. The LSTM cell can store data for a specific duration. Intuitively, the input gate controls the extent to which new information can enter the cell, while the forget gate controls how much information remains within the cell, and the output gate manages which information exits the

cell to calculate the activation of the LSTM model [16]. Below are descriptions of the gates in an LSTM (Fig. 5).

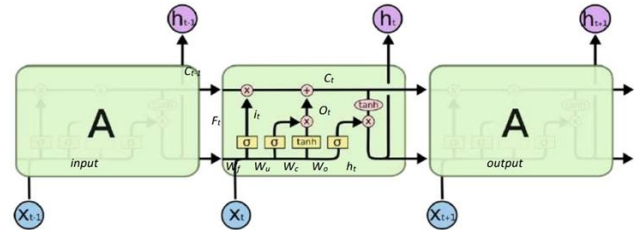


Fig. 5. Architecture of LSTM.

VI. RESULT

This section begins by discussing the characteristics of the Crude Palm Oil production in Indonesia, modeling with ANN, modeling with LSTM, and a comparison of the results from the ANN and LSTM models.

A. Crude Palm Oil in Indonesia

The production data of Crude Palm Oil (CPO) in Indonesia is very important, especially for the government. The government requires information on Crude Palm Oil (CPO) production in Indonesia to assist in the effective planning and management of CPO supplies, reduce economic risks due to fluctuations in CPO production, and devise better export-import strategies for CPO to enhance the stability of the national economy. The characteristics of the CPO production data in Indonesia are illustrated in Fig. 6.

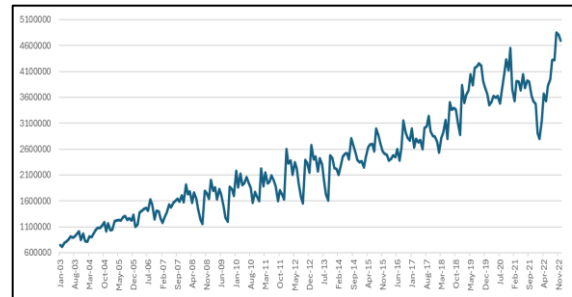


Fig. 6. Time series plot Crude Palm Oil (CPO) production in Indonesia.

Fig. 6 shows that the production of Crude Palm Oil (CPO) in Indonesia from January 2003 to December 2022 exhibits a seasonal pattern. CPO production in Indonesia from 2003 to 2013 showed a tendency to decline sharply at the end of the year, particularly in September, and had a tendency to rise sharply at the beginning of the year, especially in January. In other words, there tends to be a decrease at the end of the year, followed by a rise at the beginning of the year. This is supported by the characteristics of oil palm, which grows well during the end of the rainy season, where the rainy season in Indonesia usually starts in November and ends between January and March. However, from 2014 to 2022, the production of Crude Palm Oil (CPO) in Indonesia has shown a new seasonal pattern, characterized by a tendency to decrease at the beginning of the year and an increase at the end of the year. This is believed to be due to increasingly unpredictable climate conditions year after year. Overall, the production of Crude Palm Oil (CPO) in Indonesia has shown an increasing trend during this period, indicating a consistent rise in production year after year. The

lowest production of Crude Palm Oil (CPO) in Indonesia was recorded at the beginning of the observation period, specifically in February 2003, while the highest production occurred at the end of the observation period, specifically in October 2022.

B. Forecasting CPO Production Using Artificial Neural Network

The process in an Artificial Neural Network (ANN) begins with the input received by neurons, which are grouped in layers. The information received from the input layer is sequentially passed to the subsequent layers in the ANN until it reaches the output layer. The determination of inputs in modeling the production of Crude Palm Oil (CPO) in Indonesia using ANN is based on Fig. 7.

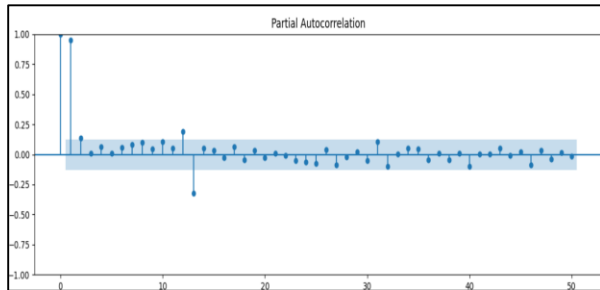


Fig. 7. PACF Crude Palm Oil (CPO) production in Indonesia.

Fig. 7 shows that the production of Crude Palm Oil (CPO) in Indonesia during period x_t has a strong relationship with production in periods x_{t-1} , x_{t-2} , x_{t-12} , and x_{t-13} . Based on this relationship, the Artificial Neural Network model uses the CPO production data from Indonesia during periods x_{t-1} , x_{t-2} , x_{t-12} , and x_{t-13} as inputs to predict CPO production in Indonesia for x_t period. This study will utilize several combinations of hyperparameters, namely a batch size of 8 and 16, neuron counts of 50 and 100, hidden layers of 1 and 2, epochs of 50 and 100, and learning rates of 0.001 and 0.01. Adam optimization will be used to update the weights of the ANN network, and MAPE out-of-sample will be employed to determine the best ANN model. In addition to these hyperparameters, this study will also use data splitting ratios to divide the dataset into training and testing subsets. Data splitting ratios refer to the proportion of data allocated for each subset, for example 80% for training, where the model learns from the data, and 20% for testing, where the model's performance is evaluated on unseen data. This approach helps to prevent overfitting and ensures the model's effectiveness on new data. The results of the modeling using the

Artificial Neural Network with several data partitioning scenarios and hyperparameter tuning are presented in Table II.

TABLE II. RESULTS OF CRUDE PALM OIL (CPO) PRODUCTION IN INDONESIA MODELLING WITH ANN

Split Ratio	Batch Size	Neurons	Hidden Layers	Epochs	Learning Rate	MAPE
70 : 30	8	150	2	150	0,01	6,87%
80 : 20	8	120	2	150	0,01	7,92%
90 : 10	8	150	2	150	0,01	6,94%

Table II shows that the best Artificial Neural Network model for predicting Crude Palm Oil (CPO) production in Indonesia is the model with a data partitioning ratio of 70:30, a batch size of 8, 150 neurons, 2 hidden layers, 150 epochs, and a learning rate of 0.01. This model achieves an out-of-sample MAPE value of 6.87%, which falls within the criteria for very good forecasting, as it is less than 10%. The best models in Table II indicate a trend that increasing the number of hidden layers and epochs reduces the model's error rate. Conversely, a smaller batch size tends to produce models with a lower error rate.

C. Forecasting CPO Production Using Long Short Term Memory

LSTM is a form of neural network. One important component in the formation of LSTM networks is the determination of the inputs used. This study will use the window size feature, or time steps, as input. The window sizes used in this research are variations of three, four, and six, representing quarterly, triannual, and semiannual periods. In addition to the input, another crucial component is the hyperparameters. This study will utilize several combinations of hyperparameters, namely a batch size of 8 and 16, neuron counts of 50 and 100, hidden layers of 1 and 2, epochs of 50 and 100, and learning rates of 0.001 and 0.01. Adam optimization will be used to update the weights of the LSTM network, and MAPE out-of-sample will be employed to determine the best LSTM model. In addition to these hyperparameters, this study will also use data splitting ratios to divide the dataset into training and testing subsets. Data splitting ratios refer to the proportion of data allocated for each subset, for example 80% for training, where the model learns from the data, and 20% for testing, where the model's performance is evaluated on unseen data. This approach helps to prevent overfitting and ensures the model's effectiveness on new data. The results of the modeling using Long Short-Term Memory with several data partitioning scenarios and hyperparameter tuning are shown in Table III.

TABLE III. RESULTS OF CRUDE PALM OIL (CPO) PRODUCTION IN INDONESIA MODELLING WITH LSTM

Split Ratio	Window Size	Batch Size	Neurons	Hidden Layers	Epochs	Learning Rate	MAPE
70 : 30	3	8	120	2	100	0,01	6.141%
	4	16	150	1	150	0,01	6.28%
	6	8	150	1	150	0,001	6.33%
80 : 20	3	8	150	1	150	0,001	5.81%
	4	16	120	1	100	0,01	5.85%
	6	16	120	2	150	0,001	5.78%
90 : 10	3	8	120	2	150	0,001	6.15%
	4	8	150	2	150	0,01	6.17%
	6	8	150	2	150	0,001	6.01%

Table III shows that the best Long Short-Term Memory (LSTM) model for a 70:30 data partitioning ratio is with a window size of 3 and a hyperparameter combination of a batch size of 16, 120 neurons, 2 hidden layers, 100 epochs, and a learning rate of 0.01. This LSTM model with the specified window size and hyperparameter combination achieves an out-of-sample MAPE of 6.14%. The best LSTM model for an 80:20 data partitioning ratio is with a window size of 6 and a hyperparameter combination of a batch size of 16, 120 neurons, 2 hidden layers, 150 epochs, and a learning rate of 0.001. This LSTM model achieves an out-of-sample MAPE of 5.78%. The best LSTM model for a 90:10 data partitioning ratio is with a window size of 6 and a hyperparameter combination of a batch size of 8, 120 neurons, 1 hidden layer, 150 epochs, and a learning rate of 0.001. This LSTM model achieves an out-of-sample MAPE of 6.01%. Therefore, the best model to be used in the dashboard to be developed is with an 80:20 data partitioning ratio, a window size of 6, and a hyperparameter combination of a batch size of 16, 120 neurons, 2 hidden layers, 150 epochs, and a learning rate of 0.001.

Table III also indicates that there is no definitive pattern of hyperparameters that consistently yields the best model. The best model obtained is the one with a window size of 6, which aligns with the seasonal pattern of Crude Palm Oil (CPO) production in Indonesia. For example, from 2003 to 2013, there was a tendency for production to decrease in September and increase in February, which corresponds to a time gap of six months from September to February.

D. Comparison of Artificial Neural Network and Long Short Term Memory

The results of modeling Crude Palm Oil (CPO) production in Indonesia using general neural network methods, namely Artificial Neural Network (ANN), and neural networks specifically designed for time series data, namely Long Short-Term Memory (LSTM) are presented in Table IV.

TABLE IV. COMPARISON ANN AND LSTM MODEL

Metode	MAPE
Artificial Neural Network (ANN)	6,87%
Long Short Term Memory (LSTM)	5,78%

Table IV shows that the modeling of Crude Palm Oil (CPO) production in Indonesia using the general neural network method, namely Artificial Neural Network (ANN), and the neural network method specifically designed for time series data, namely Long Short-Term Memory (LSTM), yields out-of-sample MAPE values of 6.87% and 5.78%, respectively. Based on the out-of-sample MAPE values, both methods can be classified as capable of modeling CPO production in Indonesia very well, as the MAPE values are both below 10%. When comparing the two methods, the Long Short-Term Memory (LSTM) method provides a lower out-of-sample MAPE compared to the Artificial Neural Network (ANN), indicating

that the LSTM method has a better model performance than the ANN.

VII. DISCUSSION

The findings of this study demonstrate the effectiveness of ANN and LSTM in predicting CPO production in Indonesia. Both models performed exceptionally well, with MAPE values below 10%, indicating their suitability for time series forecasting. However, LSTM showed superior performance with a MAPE of 5.78%, outperforming ANN, which achieved a MAPE of 6.87%. This suggests that LSTM's ability to process sequential data and capture long-term dependencies allows it to model the complex and dynamic seasonal variations in CPO production more effectively. While ANN is proficient at handling non-linear relationships, its focus on shorter-term dependencies may have limited its adaptability to the shifts in seasonal patterns observed during the study period.

The seasonal trends in CPO production, driven by climatic factors and the oil palm growth cycle, emphasize the importance of models that can capture long-term temporal dynamics. From 2003 to 2013, CPO production typically decreased toward the end of the year and increased at the start of the following year, but after 2014, these patterns shifted due to changing and unpredictable climate conditions. The LSTM model's use of a six-month window size proved effective in addressing these shifts, reflecting its ability to align with seasonal cycles. Furthermore, the study underscores the importance of optimizing hyperparameters to improve model performance. For ANN, better results were achieved with smaller batch sizes, a larger number of neurons, and deeper network layers. In contrast, for LSTM, the optimal configuration included a six-month window size, a batch size of 16, and specific combinations of hidden layers, neurons, and learning rates.

The practical applications of these findings are significant for the CPO industry and policymakers. Accurate forecasts can support decision-making in inventory management, price stabilization, and export-import planning, thereby enhancing economic stability. The superior performance of LSTM makes it particularly useful for addressing the challenges posed by evolving seasonal patterns and climate variability in CPO production. Nonetheless, the study has certain limitations, such as relying solely on historical production data without incorporating external factors like market demand, policy changes, or global price fluctuations. Future research could explore hybrid approaches that combine ANN, LSTM, and other machine learning techniques to further improve accuracy. Moreover, integrating exogenous variables, such as weather conditions or economic indicators, could lead to more comprehensive forecasting models. Overall, this research demonstrates the potential of advanced neural network models in tackling complex forecasting challenges in the agricultural sector and provides a strong foundation for future studies.

VIII. CONCLUSION

The LSTM model demonstrates superior performance compared to the ANN model in modeling and forecasting Crude Palm Oil (CPO) production in Indonesia. This conclusion is supported by a comparative analysis of Mean Absolute Percentage Error (MAPE) values, where the LSTM model

achieved a notably lower MAPE (5.78%) in out-of-sample predictions compared to the ANN model (6.87%). The lower MAPE value highlights the LSTM model's enhanced capability to minimize the deviation between predicted values and actual observations, ensuring higher predictive accuracy. The superior performance of LSTM is attributed to its architectural design, which is specifically optimized to capture long-term dependencies and temporal patterns in sequential data. This makes LSTM particularly effective in handling the seasonal fluctuations and nonlinear trends inherent in agricultural production processes like CPO. For example, the LSTM model successfully identified seasonal production trends and temporal dependencies, such as increased production during harvest periods and decreases during off-seasons, which the ANN model struggled to replicate. In contrast, while the ANN model performs well in general-purpose predictive tasks, it lacks the specialized mechanisms to account for sequential and temporal dynamics. Consequently, its predictions are less robust when applied to CPO production data with complex seasonality and long-term patterns.

These findings emphasize the importance of selecting appropriate modeling techniques tailored to the characteristics of the dataset. By leveraging the strengths of LSTM, stakeholders in the CPO industry can achieve more reliable forecasts, allowing for improved planning, efficient resource allocation, and informed decision-making to strengthen Indonesia's position as a global leader in palm oil production. Future research could explore the use of more recent and extensive datasets to validate the robustness of the LSTM model over a longer period. Additionally, investigating other advanced machine learning models, such as Transformers or hybrid models, could provide further improvements in forecasting accuracy. Incorporating external factors like climate change, government policies, and global market prices into the models may also enhance their predictive capabilities. Furthermore, optimizing hyperparameters using advanced techniques like Bayesian Optimization and testing the models in real-world inventory management and export-import strategies could provide valuable insights for practical applications. Finally, conducting similar studies in other major CPO-producing countries could help generalize the findings and adapt the models to different conditions.

ACKNOWLEDGMENT

The authors extend their gratitude for the financial support provided for this research, funded by Institut Teknologi Sepuluh Nopember through a Research Grant under contract number 1636/PKS/ITS/2024.

DATA AVAILABILITY

The data supporting the findings of this study are openly accessible and can be retrieved using the following link: <https://doi.org/10.17632/8smrcwbqgx.1>. All relevant datasets used and analyzed during the research have been thoroughly documented to ensure transparency and reproducibility.

CONFLICT OF INTEREST STATEMENT

The authors state that they have no conflicts of interest related to the publication of this article. Each author has been involved in the research and manuscript development without any financial, professional, or personal ties that could compromise the integrity of the findings or their interpretation.

REFERENCES

- [1] Y. Basiron and C. K. Weng, "The oil palm and its sustainability," *J Oil Palm Res*, vol. 32, no. 3, pp. 455–473, 2020.
- [2] K. Laia, "Peramalan produksi Crude Palm Oil (CPO) di Provinsi Riau dengan pendekatan model ARIMA (Autoregresif Integrated Moving Average)," 2019.
- [3] S. Kurnia, "Analisis faktor-faktor yang memengaruhi nilai tukar rupiah di Indonesia tahun 1999-2020," 2022.
- [4] S. Zhang, Y. Yang, and Z. Xu, "Application of ARIMA and Machine Learning in Time Series Data Analysis," *Journal of Applied Research in Technology*, vol. 16, no. 4, pp. 543–555, 2019.
- [5] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "The M5 Competition: Results, findings, and conclusions," *Int J Forecast*, vol. 36, no. 1, pp. 54–74, 2020.
- [6] T. K. Shih and C. H. Lin, "Comparative Study of LSTM, ARIMA, and Prophet Models for Time Series Forecasting," *Data Sci J*, vol. 20, no. 1, pp. 24–37, 2021.
- [7] S. H. Y. Tyas, "Tinjauan Pustaka Sistematis: Perkembangan Metode Peramalan Harga," 2022.
- [8] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2020.
- [9] M. I. Anshory, Y. Priyandari, and Y. Yuniaristanto, "Peramalan Penjualan Sediaan Farmasi Menggunakan Long Short-term Memory: Studi Kasus pada Apotik Suganda," *Performa: Media Ilmiah Teknik Industri*, vol. 19, no. 2, 2020.
- [10] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A review of time series forecasting applications using ARIMA models and machine learning approaches: Financial, health, and utility applications," *Journal of Artificial Intelligence Research*, 2023.
- [11] B. Yildiz, J. I. Bilbao, and A. B. Sproul, "ANN vs ARIMA for energy consumption forecasting in commercial buildings," *Renew Energy*, vol. 156, pp. 82–93, 2020.
- [12] A. Uskono, B. Smith, and C. Johnson, "Comparative Analysis of ARIMA and LSTM Models for Crude Palm Oil Price Forecasting," *International Journal of Agricultural Research*, vol. 12, no. 4, pp. 567–579, 2023.
- [13] Y. Yan and J. Chen, "Comparing ARIMA and LSTM for GDP Growth Forecasting: Evidence from Emerging Economies," *Econ Model*, vol. 95, pp. 224–235, 2021.
- [14] International Journal of Advanced Computer Science and Applications, "Time Series Forecasting using LSTM and ARIMA," *International Journal of Advanced Computer Science and Applications*, 2023.
- [15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2019.
- [16] S. I. N. Suwandi, R. Tyasnurita, and H. Muhayat, "Peramalan Emisi Karbon Menggunakan Metode SARIMA dan LSTM," *Journal of Computer Science and Informatics Engineering (J-Cosine)*, vol. 6, no. 1, pp. 73–80, 2022.
- [17] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE Trans Neural Netw Learn Syst*, vol. 28, no. 10, pp. 2222–2232, 2019.
- [18] H. Purnomo, H. Suyono, and R. N. Hasanah, "Peramalan Beban Jangka Pendek Sistem Kelistrikan Kota Batu Menggunakan Deep Learning Long Short-Term Memory," *Transmisi: Jurnal Ilmiah Teknik Elektro*, vol. 23, no. 3, pp. 97–102, 2021.

Enhanced Jaya Algorithm for Quality-of-Service-Aware Service Composition in the Internet of Things

Yan SHI

Hebei Chemical & Pharmaceutical College, Shi Jiazhuang 050026, China

Abstract—The Internet of Things (IoT) has shifted how devices and services interact, resulting in diverse innovations ranging from health and smart cities to industrial automation. Nevertheless, at its core, IoT continues to face one of the major tough tasks of Quality of Service-aware Service Composition (QoS-SC), as these IoT settings are normally transient and unpredictable. This paper proposes an improved Jaya algorithm for QoS-SC and focuses on optimizing service selection with a balance between the main QoS attributes: execution time, cost, reliability, and scalability. The proposed approach was designed with adaptive mechanisms to avoid local optima stagnation and slow convergence and thus assure robust exploration and exploitation of the solution area. Incorporating these enhancements, the proposed algorithm outperforms prior metaheuristic approaches regarding QoS satisfaction and computational efficiency. Extensive experiments conducted over diverse IoT scenarios show the algorithm's scalability, demonstrating that it can achieve faster convergence with superior QoS optimization.

Keywords—Service composition; internet of things; quality of service; Jaya algorithm; optimization

I. INTRODUCTION

The Internet of Things (IoT) is a transformational paradigm connecting diverse devices through a harmonious and interoperable structure [1]. This would enable cooperation among many smart devices to deliver innovative services, including those within the domains of healthcare and smart cities, as well as industrial automation [2]. With the fast proliferation of these connected gadgets, IoT holds promise for an array of applications driven by the urge for sufficient communication and function [3]. However, device functionalities are highly diverse and limited by resource constraints such as battery life and processing capacity [4]. In this respect, integrating services from heterogeneous IoT devices into composite applications is essential for seamless service delivery while meeting user needs efficiently within set energy and resource constraints [5]. In addition, constitutive models for the simulation of weak rock masses can be applied to obtain insights into resource optimization and structural robustness in IoT-driven systems involving infrastructure and industrial automation [6].

In IoT environments, most individual atomic services are not competent at delivering complex user requirements independently [7]. Thus, combining atomic services with varying Quality of Service (QoS) attributes or characteristics like cost, reliability, and scalability leads to composite services [8]. The fulfillment of composite services depends on Service-Oriented Computing (SOC) principles, allowing the

composition of services into workflows that match a wide range of applications [9]. Indeed, this involves selecting an optimum from many service candidates considering constraints related to energy consumption, which are constantly changing with ever-changing user preferences and dynamic network conditions. With such enlargement and complications in IoT systems, guaranteeing service quality and dependability is challenging.

As a matter of fact, QoS-aware Service Composition (QoS-SC) involves selecting the best services from a vast pool of candidates while optimizing conflicting QoS criteria such as execution time, cost, and reliability [10]. The problem is compounded by its combinatorial nature, which makes it NP-hard [11]. Traditional metaheuristic methods often struggle with local optima stagnation and slow convergence, limiting their ability to address large-scale, dynamic IoT environments efficiently. To overcome these challenges, this study proposes an enhanced Jaya algorithm designed explicitly for QoS-SC in IoT. The algorithm balances exploration and exploitation by incorporating adaptive mechanisms and a stagnation-recovery strategy, improving convergence speed and solution quality. It also adapts to varying workflows, including sequential, parallel, and loop-based structures, to effectively model diverse IoT scenarios.

The contributions of this work are fourfold: (1) introducing an enhanced Jaya algorithm with adaptive mechanisms for QoS-SC, (2) developing a stagnation-recovery technique to overcome local optima, (3) evaluating the algorithm's performance against state-of-the-art methods across diverse IoT scenarios, and (4) demonstrating the scalability and computational efficiency of the proposed approach. This study presents a robust approach for optimizing service composition in dynamic IoT ecosystems.

The remainder of this paper is structured in the following way. Section II summarizes related research on QoS-aware service composition and optimization methods. The problem is formulated in Section III. Section IV describes the proposed algorithm in detail. Section V presents the experimental setup, outcomes, and comparisons with existing methodologies. Finally, Section VI summarizes the main conclusions and recommendations for further study.

II. RELATED WORK

The solutions to QoS-SC have been addressed in many research works by applying different optimization methods. For example, Sefati and Navimipour [12] presented a hybrid method using Hidden Markov Models (HMM) and Ant Colony Optimization (ACO) to address partial challenges in the composition of IoT services. HMM predicts QoS attributes by

learning the optimal emission and transition matrices via the Viterbi algorithm, while ACO estimates QoS to find the best service paths.

Vakili, et al. [13] proposed a service composition strategy based on the Grey Wolf Optimization (GWO) algorithm under the MapReduce methodology. This significantly improves cost, availability, and response time QoS attributes when discovering an optimal set of atomic services. In the end, the simulation results reduce cost and response time and improve the amount of energy saved regarding availability.

Asghari, et al. [14] propose a hybrid evolutionary algorithm (SFLA-GA) for privacy-preserving cloud service composition. A computational scheme selects the optimal QoS aggregation selection, while services are categorized according to their privacy level. Results indicated better fitness values and service selection compared to the existing algorithms.

Xiao [15] presented a service composition method leveraging cloud and fog computing and an improved Artificial Bee Colony (ABC) algorithm. The approach introduced a scheme for Dynamic Reduction to enhance convergence and balance exploration and diversification. Evaluations show reduced energy consumption compared to traditional algorithms and increased reliability and, thus, cost optimization.

Rajendran, et al. [16] proposed an enhanced eagle strategy algorithm for large-scale Dynamic Web Service Composition (DWSC) in cloud-based IoT environments, bio-inspired and much more computationally efficient with huge repository challenges. Therefore, the computation time would be faster and the QoS metrics much improved.

Tang, et al. [17] suggested an Improved Shuffled Frog Leaping Algorithm (ISFLA) using chaos and reverse learning theories to enhance population initialization and diversity. This technique used Gaussian mutation and a local update method to find the optimum IoT service composition. The simulation shows superior fitness values, quicker convergence, and better solution quality than SFLA and related techniques.

Ait Hacène Ouhadda, et al. [18] presented the Discrete Adaptive Lion Optimization Algorithm (DALOA), which is empowered by operators of exploration-exploitation strategies: roaming, mating, and migration. The approach divided the population into two groups: pride and nomads, to balance diversity with efficiency. These results indicated that DALOA provided near-optimal solutions within acceptable execution times and that this method outperformed the rest of the analyzed algorithms.

As highlighted in Table I, existing IoT service composition solutions still have a few highly valued shortcomings that can be improved in dynamic/large-scale environments. Most current solutions focus on optimizing single QoS attributes, such as response time or cost, in a non-holistic manner. Scalability remains a persistent problem, especially in methods like HMM-ACO and the Improved Eagle Strategy, when dealing with large-scale IoT repositories. Balancing exploration and exploitation is a core limitation in approaches such as

SFLA-GA and ISFLA; this often leads to convergence at premature stages or very suboptimal solutions. Most algorithms have underexplored privacy concerns, addressed in only a few methods, such as SFLA-GA. To address these lacunae, the current paper proposes an improved variant of the Jaya algorithm with an adaptive mechanism and stagnation-recovery strategy. This approach will maintain an equilibrium between exploration and exploitation while guaranteeing scalability, accelerated convergence, and holistic QoS optimization, considering dynamic repository updates and privacy issues.

III. PROBLEM DESCRIPTION

QoS-SC in IoT concerns integrating abstract services provided by different providers into workflows to fulfill users' needs. Workflow are series of expert-level services that are needed for task execution. Typical applications of such workflows in smart city contexts are journey-planning applications, whereby different sub-services, including booking transportation, route planning, and even some payment systems, are all composed into one integrated single service. In general, selecting a concrete option with many sub-services and various QoS attributes will be complex and dynamic. The process of QoS-SC is shown in Fig. 1.

TABLE I. PREVIOUS IoT SERVICE COMPOSITION METHODS

Study	Main contribution	Shortcomings addressed in our study
HMM-ACO [12]	Combined HMM for QoS prediction and ACO for optimal pathfinding, improving QoS metrics like availability and cost.	Lack of dynamic adaptation and scalability to large-scale IoT repositories, addressed by integrating adaptive mechanisms. Narrow focus on specific QoS attributes; our study proposes a holistic QoS optimization framework considering diverse attributes.
GWO with MapReduce [13]	Integrated GWO with MapReduce to optimize QoS attributes like energy, cost, and response time.	Insufficient balance between exploration and exploitation; our method enhances this balance for better convergence and solutions.
SFLA-GA [14]	Proposed a hybrid privacy-aware service composition using SFLA and GA, optimizing QoS while addressing privacy.	Limited adaptability to dynamic IoT environments; our study integrates real-time optimization mechanisms.
Enhanced ABC with fog and cloud [15]	Leveraged cloud and fog computing with ABC and dynamic reduction for improved convergence and energy efficiency.	Ineffective for handling real-time service updates; our algorithm ensures scalability and adaptability to dynamic conditions.
Improved eagle strategy [16]	Addressed large-scale DWSC with a bio-inspired algorithm, improving computation time and QoS metrics.	High computational complexity for large IoT networks; our approach improves efficiency while maintaining scalability.
ISFLA [17]	Enhanced SFLA with chaos theory and reverse learning for better population diversity and fitness.	Longer execution time for large-scale repositories; our study emphasizes faster convergence and scalability in diverse scenarios.
DALOA [18]	Introduced DALOA with strong exploration and exploitation balance using sub-population strategies.	

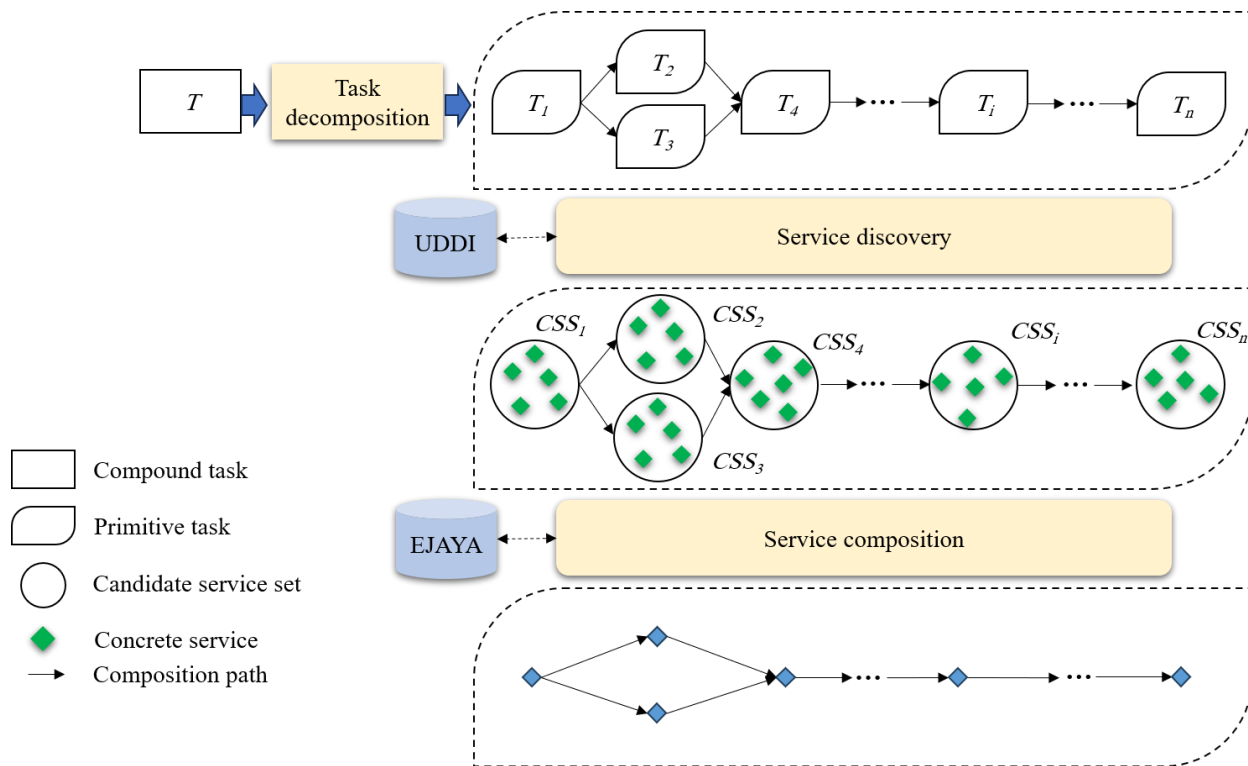


Fig. 1. An overview of QoS-SC process.

This inherent complexity naturally arises from the fact that functionally equivalent services feature distinct QoS metrics, namely response time, cost, and reliability. To handle this, IoT service composition is made up of five layers: a perception layer responsible for sensing; a network layer transferring services to the cloud; a cloud layer providing service databases; a composition layer that selects and composes services; and an application layer that enables users to interact. These layers have similarities to the structure of ISO network layers.

QoS evaluation is an indispensable process in service selection and composition in IoT environments, relying on seven key characteristics representative of various performance metrics and user requirements:

- Execution time: The time that elapses between a user request and the system's response. The shorter the execution time, the better the performance.
- Reliability: The ratio of completed service requests to the total number of requests, reflecting the dependability of the service.
- Execution cost: Represents the cost of utilizing a service. Lower costs are preferred.
- Availability: This gives the percentage of time a service continues to be operational and available over a given period.
- Scalability: The service's ability to adapt and function efficiently under changing demands or conditions.

- Reputation: A trust metric derived from user feedback; it can fall into the "very high," "high," "normal," "poor," or "very poor" categories.
- Response time: The time interval between a user's inquiry and the system's delivery of the requested service.

These attributes can be classified into two categories: cost indicators, where lower values are preferred, such as cost and execution time, and benefit indicators, where higher values are desired, including reliability and availability. Normalization ensures consistent evaluation. Raw QoS values are adjusted based on their minimum and maximum possible values. For cost-related QoS attributes (c_i), the normalization can be represented as by Eq. (1).

$$N(c_i) = \begin{cases} \frac{\max(C) - c_i}{\max(C) - \min(C)}, & \text{if } \max(C) \neq \min(C) \\ 1, & \text{if } \max(C) = \min(C) \end{cases} \quad (1)$$

Where $C(c_i)$ stands for the current cost value for the i^{th} QoS attribute, $\max(C)$ refers to the maximum cost value across all QoS attributes, and $\min(C)$ denotes the minimum cost value across all QoS attributes. For benefit-related QoS attributes (b_i), the normalization can be expressed using Eq. (2).

$$N(b_i) = \begin{cases} \frac{b_i - \min(B)}{\max(B) - \min(B)}, & \text{if } \max(B) \neq \min(B) \\ 1, & \text{if } \max(B) = \min(B) \end{cases} \quad (2)$$

Where (b_i) specifies the current benefit value for the i^{th} QoS attribute. $\max(B)$ and $\min(B)$ refer to maximum and minimum benefit value across all QoS attributes, respectively. Eq. (3) computes the fitness value for service composition by

weighting these normalized QoS values according to user preferences (w_i).

$$Fitness = \sum_{i=1}^r w_i \cdot N(q_i) \quad (3)$$

Where $N(q_i)$ refers to the normalized value of the i^{th} QoS attribute (either cost or benefit) and r represents the total number of QoS factors considered.

Service composition workflows describe how atomic services are arranged to form composite services. These workflows can significantly affect the aggregated QoS values. As shown in Fig. 2, the most common types are:

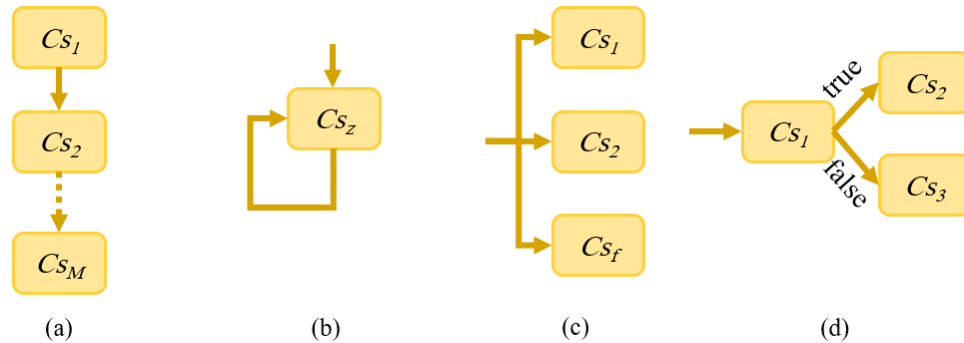


Fig. 2. Different workflow patterns for service composition: (a) Sequential, (b) Loop, (c) Parallel, and (d) Conditional (Switch).

The aggregation functions for different QoS attributes vary by workflow type, as summarized in Table II, where L refers to the number of iterations in a loop, and $t_{r,i,j}$, $c_{r,i,j}$, $a_{r,i,j}$, $r_{r,i,j}$ correspond to response time, execution cost, availability, and reliability, respectively, for the i^{th} task and j^{th} candidate service.

TABLE II. AGGREGATION FUNCTIONS FOR QOS ATTRIBUTES

Quality indicator	Loop	Parallel	Switch	Sequential
Reliability	$r_{r,i,j}^L$	$\max r_{r,i,j}$	$\prod r_{r,i,j}$	$\prod r_{r,i,j}$
Availability	$a_{r,i,j}^L$	$\max a_{r,i,j}$	$\prod a_{r,i,j}$	$\prod a_{r,i,j}$
Execution time	$L \cdot c_{r,i,j}$	$\min c_{r,i,j}$	$\sum c_{r,i,j}$	$\sum c_{r,i,j}$
Response time	$L \cdot t_{r,i,j}$	$\min t_{r,i,j}$	$\sum t_{r,i,j}$	$\sum t_{r,i,j}$

IV. ENHANCED JAYA ALGORITHM

The Enhanced Jaya Algorithm (EJAYA) was developed to address significant deficiencies in the traditional Jaya algorithm. Despite many applications to various optimization problems, Jaya's potential drawbacks include the possibility of convergence to a premature optimal solution due to its dependence on the information of the local optimum with reduced diversity while exploring the solution space for an appropriate solution [19]. These challenges could be overcome by EJAYA through several strategies directed toward local improvement of intensification and global improvement of exploration, ensuring an improvement by a factor greater than overall search efficiency and robustness. Such improvements seek to provide more enhanced balancing between diversification-segregated searching across extensive areas over the solution space and intensified structuring down into

- Sequential workflow: Services are executed one after the other in a sequence. QoS attributes like response time are typically aggregated using summation.
- Loop workflow: Services are repeated multiple times, with QoS attributes like response time multiplied by the number of iterations.
- Parallel workflow: Multiple services are executed simultaneously, with QoS attributes such as reliability aggregated using the maximum value.
- Switch workflow: Represents conditional execution paths where only one of the services is selected based on certain conditions.

up-coming regions. The traditional Jaya algorithm updates the position of a solution (x_i) within a population (N) using Eq. (4).

$$v_i = x_i + \lambda_1(x_{Best} - |x_i|) - \lambda_2(x_{Worst} - |x_i|), \quad i = 1, 2, \dots, N \quad (4)$$

Where x_{Best} and x_{Worst} are the best and worst solutions in the current population, λ_1 and λ_2 are random numbers in the range $[0,1]$, and v_i is the updated solution.

The decision to retain or discard the updated solution is based on its fitness value calculated by Eq. (5).

$$x_i = \begin{cases} v_i, & \text{if } f(v_i) \leq f(x_i) \\ x_i, & \text{otherwise} \end{cases} \quad (5)$$

This update process is straightforward but can lead to reduced population diversity, particularly in later iterations, when solutions begin to converge near the global best. The limitations of the basic Jaya algorithm include:

- Local optima stagnation: As the algorithm heavily relies on x_{Best} and x_{Worst} , the population may become trapped in local optima, reducing the probability of finding the global optimum.
- Reduced diversity: The absolute value symbol in the update equation contributes to a loss of diversity, making it challenging to explore new regions in the solution space effectively.
- Imbalance of exploration and exploitation: Basic Jaya lacks mechanisms to dynamically balance the search space exploration and the refinement of promising solutions.

To address these challenges, EJAYA introduces advanced strategies for local exploitation and global exploration, significantly improving its performance on complex optimization problems. Original JAYA locally updates the solutions by considering an upper attract point, P_u , and a lower attract point, P_l , so that the solution is attracted to more promising areas of the feasible solution space:

Upper attract point Eq. (6):

$$P_u = \lambda_3 \cdot x_{Best} + (1 - \lambda_3) \cdot M \quad (6)$$

Where M is the mean solution of the current population calculated using Eq. (7).

$$M = \frac{1}{N} \sum_{i=1}^N x_i \quad (7)$$

Lower attract point (Eq. 8):

$$P_l = \lambda_4 \cdot x_{Worst} + (1 - \lambda_4) \cdot M \quad (8)$$

These attract points provide additional flexibility, allowing solutions to gravitate toward the best and worst solutions while maintaining a strong connection to the mean of the population. This mechanism reduces premature convergence and improves diversity. The updated solution is calculated using Eq. (9):

$$v_i = x_i + \lambda_5(P_u - x_i) - \lambda_6(P_l - x_i), \quad i = 1, 2, \dots, N \quad (9)$$

Where λ_5 and λ_6 are random numbers in the range $[0, 1]$.

To enhance exploration, EJAYA incorporates a historical population (X_{old}) and a switch probability (P_{switch}), ensuring greater diversity and escaping local optima:

Historical population: The historical population is generated using Eq. 10.

$$X_{old} = \begin{cases} X, & \text{if } P_{switch} \leq 0.5 \\ \text{permute}(X), & \text{otherwise} \end{cases} \quad (10)$$

Where $\text{permute}(X)$ represents a random reordering of the population, introducing randomness and diversity.

Global exploration update: The solution is updated using Eq. (11).

$$v_i = x_i + \kappa(x_{old,i} - x_i), \quad i = 1, 2, \dots, N \quad (11)$$

Where κ is a random number sampled from a standard normal distribution. This process assures that the algorithm investigates unexplored areas in the solution space. Fig. 3 illustrates the pseudocode of EJAYA.

Input: Population size (N), Upper limits of variables (u), Lower limits of variables (l), Current number of function evaluations ($T_{current} = 0$), Maximum number of function evaluations (T_{max}).

Output: Optimal solution (x_{Best}).

Step 1: Initialization

Generate the initial population X and historical population X_{old} using Eq. 4.

Calculate the fitness values for all individuals in X and identify x_{Best} and x_{Worst} .

Update the function evaluations:

$$T_{current} = T_{current} + N$$

Step 2: Main loop

While $T_{current} < T_{max}$:

For each individual $i = 1$ to N :

Generate a random probability P_{select} in the range $[0, 1]$.

If $P_{select} > 0.5$:

Perform the **local exploitation strategy**:

Select x_{Best} and x_{Worst} .

Compute the mean solution M of the population using Eq. 7.

Update the individual using the local exploitation strategy described in Eq. 6, Eq. 8, and Eq. 9.

Else

Perform the **global exploration strategy**:

Use the historical population x_{old} and apply the global exploration strategy using Eq. 10 and Eq. 11.

End for

Update the function evaluations.

End While

Step 3: Output

Return the best solution (x_{Best}).

Fig. 3. The pseudocode of EJAYA.

V. EXPERIMENTAL RESULTS

The proposed EJAYA was evaluated for IoT service selection and composition using real datasets, containing 25 scenarios. Each dataset contained about 2500 real tasks, characterized by criteria such as cost, response time, availability, and dependability. In generating these scenarios, different numbers of abstract tasks n and concrete services m

for each abstract task were considered. The experiment analyzed the effectiveness of EJAYA in comparison with five algorithms: ABC, Particle Swarm Optimization (PSO), Discrete Dragonfly Algorithm (DDA), Genetic Algorithm (GA), and Ant Colony Optimization (ACO).

The computation environment used was on a Windows OS system with Intel Core i5 at 3.2 GHz, with 16 GB RAM. All

algorithms were implemented in MATLAB version 2020a. Each algorithm has been executed with a population size of 30 and for 30 runs, up to a maximum number of 1000 iterations for each execution. The performance of EJAYA was evaluated based on QoS fitness value that measures the QoS selection based on weighted QoS attribute; execution time, the time taken to converge to an optimal solution; and convergence rate, the ability of the algorithm to escape local optima and achieve better solutions over iterations.

EJAYA consistently outperformed all other algorithms across all scenarios. For example, as shown in Fig. 4, when the number of tasks varied from 10 to 100 and the number of concrete services ranged from 10 to 100, EJAYA achieved higher fitness values than other algorithms. Also, with a fixed $n=20$ and m ranging from 200 to 1000, EJAYA maintained superior performance, as illustrated in Fig. 5.

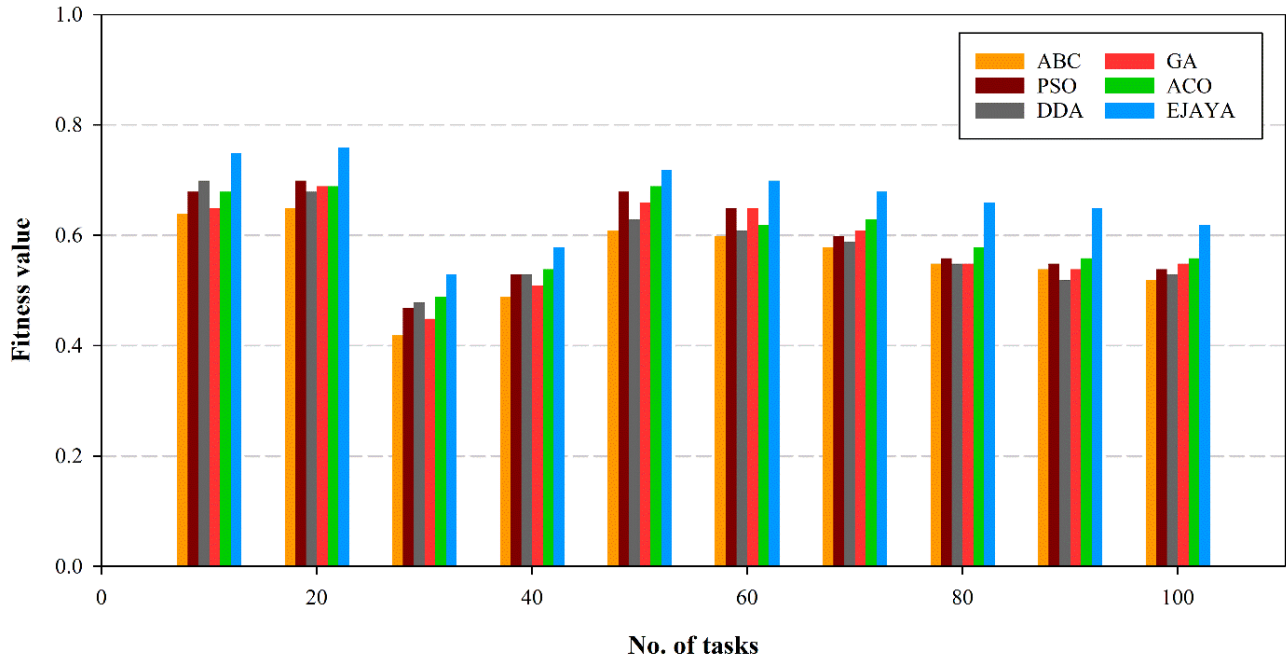


Fig. 4. Fitness values for algorithms (Scenario 1).

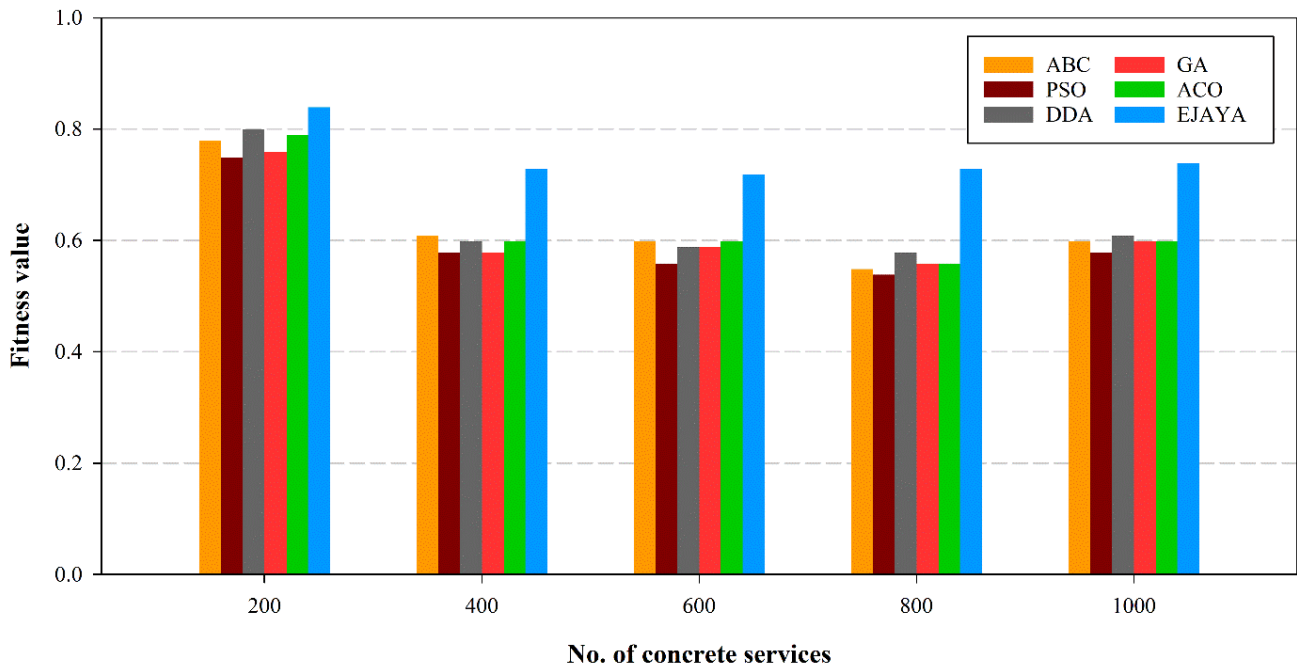


Fig. 5. Fitness values for algorithms (Scenario 2).

Fig. 6 shows the convergence curves, where EJAYA never got stuck and thus escaped from the local optima, where other algorithms were not capable of improving their solution after a number of iterations. EJAYA illustrated a gradually increasing trend over iterations in fitness values which describes that superior solutions are more quickly obtained.

Fig. 7 compares the execution time of EJAYA with those of other algorithms for an increasing number of concrete services when $n=20$. Note that EJAYA showed competitive computational efficiency, while its execution times were below those of most algorithms. For instance, for $m=1000$, its execution time in EJAYA was about 1.65 s, much faster compared to the other algorithms.

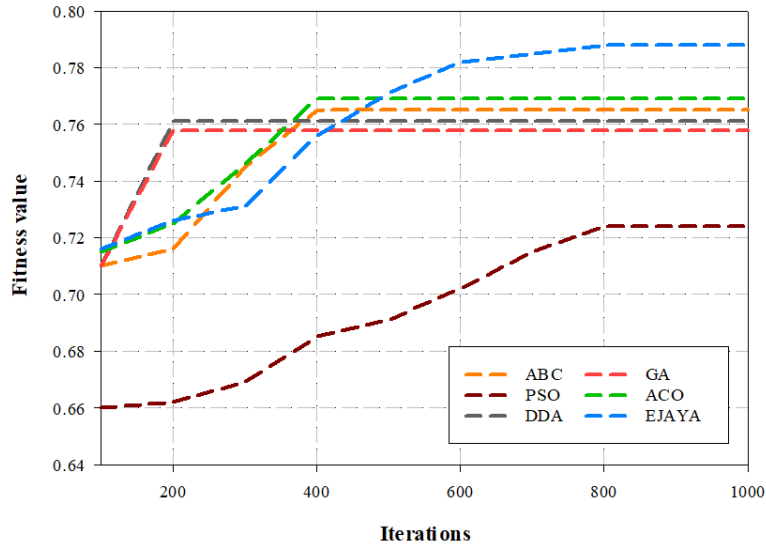


Fig. 6. Convergence curves.

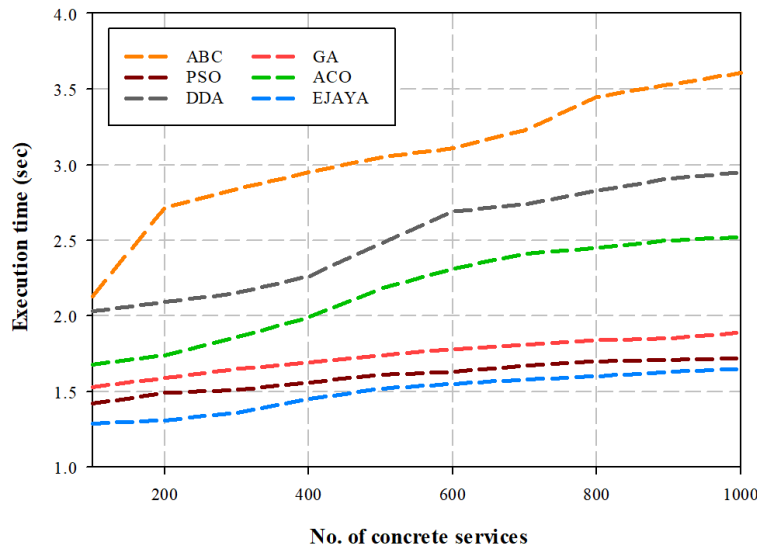


Fig. 7. Execution time comparison.

The experimental observations reveal that EJAYA outperforms most QoS-SC scenarios, indicating its strength in IoT environments. Compared to its competitors, EJAYA consistently delivers higher QoS fitness values, suggesting it can optimize service composition effectively. This is due to the adaptive mechanisms and stagnation recovery strategies of EJAYA, enabling it to escape local optimum and converge on better solutions. For example, in scenarios involving different numbers of tasks and concrete services, EJAYA reached higher fitness values, proving it is more scalable and flexible for different IoT settings. The above results confirm findings from

recent related work, which establishes the importance of adaptability and convergence in dynamic optimization problems, further asserting EJAYA's relevance as one of the methods for QoS-SC.

Most importantly, computational efficiency establishes the applicability of EJAYA in real-world applications. It provides faster convergence times, especially for large datasets, making it superior to other algorithms in terms of execution speed. This is consistent with the literature that suggests execution time is one of the most critical considerations in dynamic IoT

environments wherein the process of service composition should be executed fairly quickly. In addition, convergence curves illustrate that EJAYA improves performance steadily in each iteration and, therefore, avoids stagnation and achieves the best solutions. These results validate the design objectives of the algorithm and highlight its potential contributions to the area of QoS-aware service composition in IoT ecosystems.

VI. CONCLUSION

In this paper, we proposed the EJAYA, a robust optimization for QoS-SC in IoT environments. EJAYA has been developed to incorporate an advanced local exploitation strategy and a global exploration strategy to overcome some major drawbacks of the original Jaya algorithm, such as local optima susceptibility and reduced diversity. The algorithm employed upper and lower attract points, enhancing local search using historical populations for better global exploration. It was balanced between exploration and exploitation. The experimental outcomes proved that EJAYA outperformed the existing optimization algorithms, such as ACO, GA, DDA, PSO, and ABC. EJAYA achieved the highest QoS fitness values for all the tested datasets, escaped stagnation, and remained competitive in execution time. These results verify the efficiency of EJAYA in dealing with the complexities of large-scale service composition problems and obtaining optimal solutions with improved performance stability. EJAYA will be extended in the future for resource allocation problems in both edge and fog computing environments while incorporating dynamic scenarios to meet IoT real-time requirements. Besides, integrating machine learning techniques for further optimization and adaptability might result in a more solidification of the algorithm in the performance of highly dynamic IoT ecosystems.

ACKNOWLEDGMENT

This work was funded by Science Research Project of Hebei Education Department (No. ZC2022024).

REFERENCES

- [1] A. Shoomal, M. Jahanbakht, P. J. Componation, and D. Ozay, "Enhancing supply chain resilience and efficiency through internet of things integration: Challenges and opportunities," *Internet of Things*, p. 101324, 2024.
- [2] B. Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy-efficient data fusion methods in the Internet of Things," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 15, p. e6959, 2022.
- [3] B. Pourghebleh and V. Hayyolalam, "A comprehensive and systematic review of the load balancing mechanisms in the Internet of Things," *Cluster Computing*, vol. 23, no. 2, pp. 641-661, 2020.
- [4] S. Singh, P. K. Sharma, S. Y. Moon, and J. H. Park, "Advanced lightweight encryption algorithms for IoT devices: survey, challenges and solutions," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-18, 2024.
- [5] K. Halba, E. Griffor, A. Lbath, and A. Dahbura, "IoT capabilities composition and decomposition: A systematic review," *IEEE Access*, vol. 11, pp. 29959-30007, 2023.
- [6] A. Azadi and M. Momayez, "Review on Constitutive Model for Simulation of Weak Rock Mass," *Geotechnics*, vol. 4, no. 3, pp. 872-892, 2024, doi: <https://doi.org/10.3390/geotechnics4030045>.
- [7] D. Rastogi, P. Johri, S. Verma, V. Garg, and H. Kumar, "IoT Technology Enables Sophisticated Energy Management in Smart Factory," *Cyber Physical Energy Systems*, pp. 147-181, 2024.
- [8] B. Pourghebleh, V. Hayyolalam, and A. Aghaei Anvigh, "Service discovery in the Internet of Things: review of current trends and research challenges," *Wireless Networks*, vol. 26, no. 7, pp. 5371-5391, 2020.
- [9] S. K. Mishra and A. Sarkar, "An efficient clustering mechanism towards large scale service composition in IoT," *International Journal of Web and Grid Services*, vol. 19, no. 2, pp. 185-210, 2023.
- [10] V. Hayyolalam, B. Pourghebleh, M. R. Chehrehzad, and A. A. Pourhaji Kazem, "Single-objective service composition methods in cloud manufacturing systems: Recent techniques, classification, and future trends," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 5, p. e6698, 2022.
- [11] V. Hayyolalam, B. Pourghebleh, A. A. Pourhaji Kazem, and A. Ghaffari, "Exploring the state-of-the-art service composition approaches in cloud manufacturing systems to enhance upcoming techniques," *The International Journal of Advanced Manufacturing Technology*, vol. 105, pp. 471-498, 2019.
- [12] S. Sefati and N. J. Navimipour, "A qos-aware service composition mechanism in the internet of things using a hidden-markov-model-based optimization algorithm," *IEEE Internet of Things Journal*, vol. 8, no. 20, pp. 15620-15627, 2021.
- [13] A. Vakili, H. M. R. Al-Khafaji, M. Darbandi, A. Heidari, N. Jafari Navimipour, and M. Unal, "A new service composition method in the cloud-based internet of things environment using a grey wolf optimization algorithm and MapReduce framework," *Concurrency and Computation: Practice and Experience*, vol. 36, no. 16, p. e8091, 2024.
- [14] P. Asghari, A. M. Rahmani, and H. H. S. Javadi, "Privacy-aware cloud service composition based on QoS optimization in Internet of Things," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 11, pp. 5295-5320, 2022.
- [15] G. Xiao, "Toward Optimal Service Composition in the Internet of Things via Cloud-Fog Integration and Improved Artificial Bee Colony Algorithm," *International Journal of Advanced Computer Science & Applications*, vol. 15, no. 5, 2024.
- [16] V. Rajendran, R. K. Ramasamy, and W.-N. Mohd-Isa, "Improved eagle strategy algorithm for dynamic web service composition in the IoT: a conceptual approach," *Future Internet*, vol. 14, no. 2, p. 56, 2022.
- [17] Z. Tang, Y. Wu, J. Wang, and T. Ma, "IoT service composition based on improved Shuffled Frog Leaping Algorithm," *Heliyon*, vol. 10, no. 7, 2024.
- [18] S. Ait Hacène Ouhadda, S. Chibani Sadouki, A. Achroufene, and A. Tari, "A Discrete Adaptive Lion Optimization Algorithm for QoS-Driven IoT Service Composition with Global Constraints," *Journal of Network and Systems Management*, vol. 32, no. 2, p. 34, 2024.
- [19] E. H. Houssein, A. G. Gad, and Y. M. Wazery, "Jaya algorithm and applications: A comprehensive review," *Metaheuristics and Optimization in Computer and Electrical Engineering*, pp. 3-24, 2021.

Enhancing Facial Expressiveness in 3D Cartoon Animation Faces: Leveraging Advanced AI Models for Generative and Predictive Design

Langdi Liao¹, Lei Kang², Tingli Yue³, Aiting Zhou⁴, Ming Yang^{5*}

Design, Wuhan University of Communication, Wuhan, 430000, China^{1,3,5}

Design, Guangdong Polytechnic Institute, Guangzhou, 510000, China²

Nursing, affiliated Hospital Ofzunyi Medical University, Zunyi, 563000, China⁴

Abstract—An advanced system for facial landmark detection and 3D facial animation rigging is proposed, utilizing deep learning algorithms to accurately detect key facial points, such as the eyes, mouth, and eyebrows. These landmarks enable precise rigging of 3D models, facilitating realistic and controlled facial expressions. The system enhances animation efficiency and realism, providing robust solutions for applications in gaming, animation, and virtual reality. This approach integrates cutting-edge detection techniques with efficient rigging mechanisms. The AI-assisted rigging process reduces manual effort and ensures precise, dynamic animations. The study evaluates the system's accuracy in facial landmark detection, the efficiency of the rigging process, and its performance in generating consistent emotional expressions across animations. Additionally, the system's computational efficiency, scalability, and system performance are assessed, demonstrating its practicality for real-time applications. Pilot testing, emotion recognition consistency, and performance metrics reveal the system's robustness and effectiveness in producing realistic animations while reducing production time. This work contributes to the advancement of animation and virtual environments, offering a scalable solution for realistic facial expression generation and character animation. Future research will focus on refining the system and exploring its potential applications in interactive media and real-time animation.

Keywords—Facial landmark detection; 3D animation; deep learning; AI-assisted rigging; emotion recognition

I. INTRODUCTION

Facial expressions are a fundamental aspect of storytelling, communication, and emotional engagement in animated media. In 3D cartoon animation, creating expressive faces is a crucial element that bridges the gap between virtual characters and audience perception [1]. The ability to convey emotions such as joy, sadness, anger, fear, and surprise enables characters to resonate with viewers, immersing them in the narrative [2]. However, achieving this level of expressiveness is not without its challenges, especially in a 3D environment where facial rigging and animation require precision and creativity [3]. Traditional methods of designing facial expressions in 3D cartoon animation are both labor-intensive and time-consuming. Animators typically rely on manual keyframing [4], morph target blending, and complex rigging systems to create facial emotions. While these methods allow for detailed control, they pose significant limitations. Producing high-

quality facial animations demands extensive manual effort, expertise, and resources. Traditional processes lack automation, making them impractical for large-scale productions or real-time applications [3]. Achieving exaggerated and highly expressive facial animations requires significant trial and error, often restricting creative freedom. Maintaining consistency in facial expressions across different frames and characters can be difficult, particularly in projects with numerous assets [5]. These challenges highlight the need for advanced solutions that streamline the animation process while enhancing the expressiveness and realism of 3D cartoon characters.

The growing demand for high-quality animated content across entertainment, education, gaming, and virtual reality industries has pushed the boundaries of creativity and technology [6]. Audiences today expect not only visually appealing characters but also emotionally engaging performances that drive storytelling [7]. In this context, integrating AI-driven approaches into the facial animation pipeline offers promising opportunities. AI algorithms can automate key processes such as facial rigging, expression generation, and motion interpolation, significantly reducing production time. Generative models enable animators to explore a broader range of emotions and exaggerations, pushing creative possibilities beyond manual techniques [8]. Predictive AI models ensure consistency in facial expressions while preserving natural transitions between emotions [9]. AI-based tools lower the technical barriers for smaller animation studios and independent creators, democratizing access to advanced facial animation technologies [10].

The motivation for this study is to bridge the gap between traditional animation workflows and AI-powered tools, offering solutions that enhance expressiveness, streamline production, and foster innovation in 3D cartoon animation. This article leverages state-of-the-art AI models to generate and predict facial expressions for 3D cartoon characters. The methodology involves the following key steps i.e., existing datasets such as the Facial Expression Research Group Database (FERG) and synthetic datasets created using AI models (e.g., GANs) are utilized. These datasets include exaggerated facial expressions representing the seven basic emotions: anger, disgust, fear, happiness, sadness, surprise, and neutral. Deep generative models such as Generative Adversarial Networks (GANs) [11] and Variational

Autoencoders (VAEs) are used to synthesize new facial expressions based on input parameters. These models enable the generation of highly expressive and diverse facial animations. Machine learning techniques, including CNNs [12] and recurrent neural networks (RNNs), are applied to predict and interpolate facial expressions based on input features such as pose, texture, and landmarks. The generated expressions are evaluated for realism, emotional clarity, and consistency using qualitative and quantitative metrics. User studies are conducted to assess audience engagement and perception of the AI-generated animations [13].

The study uses a combination of publicly available and synthetic datasets to ensure diversity and coverage of facial expressions. FERF-DB is a well-known dataset comprising 55,000+ annotated images of cartoon characters with seven labeled expressions [14]. Synthetic AI-Generated Data to generate additional facial expressions that exhibit exaggerated emotions, enhancing the dataset's versatility. Custom Annotations for emotion intensity, landmark positions, and rigging points are added to improve the quality and usability of the dataset. By combining these datasets, the study ensures a robust foundation for training and testing AI models, enabling the generation of high-quality facial expressions for 3D cartoon characters.

The contribution of the article is well explained in the points below:

- This study introduces state-of-the-art AI models, including GANs and VAEs, to generate highly expressive and exaggerated facial expressions for 3D cartoon characters, pushing the boundaries of creative possibilities in animation.
- By combining the Facial Expression Research Group Database (FERG) with synthetic AI-generated data, the study ensures a comprehensive dataset that covers a wide range of facial expressions and emotional intensities, improving the versatility of animation generation.
- Utilizes CNNs and RNNs to predict and interpolate facial expressions based on parameters like pose, texture, and landmarks, ensuring consistency, realism, and natural transitions in the generated animations.

The study incorporates both qualitative and quantitative metrics to assess the realism, emotional clarity, and consistency of the AI-generated facial expressions, ensuring their effectiveness for 3D cartoon animation.

II. LITERATURE REVIEW

This study explores the use of GANs for generating facial expressions in animated characters [15]. The authors demonstrate how GANs can produce highly realistic and varied expressions, improving upon traditional animation methods. The study highlights the potential of GANs to handle different facial dynamics and offer a more flexible approach to character animation.

This research focuses on emotion recognition from 3D facial expressions, using convolutional neural networks

(CNNs) to identify emotions based on facial features [16]. The authors show that 3D models provide more accurate emotion recognition compared to 2D images, particularly in animated contexts. The study emphasizes the importance of texture and lighting in 3D emotion recognition systems. The paper investigates the use of VAEs to generate facial expressions, showcasing their ability to capture the underlying distribution of emotions [17]. The authors demonstrate that VAEs can create diverse facial expressions by learning the latent variables of facial movements. This approach enhances the expressiveness of animated characters, with smoother transitions between emotions.

This article discusses the application of machine learning techniques to achieve real-time facial animation for interactive applications [18]. The authors use deep neural networks to predict and animate facial expressions in real-time, significantly reducing the time and effort required in traditional animation pipelines. The study contributes to real-time facial animation for virtual characters in gaming and VR. This research focuses on automating the facial rigging process in 3D animation using machine learning algorithms [3]. The authors propose an AI-based approach to generate rigging parameters from minimal input data, reducing manual labor. The results show that the automated rigging system can match or exceed the quality of manually rigged models, improving efficiency in animation production. In this study, the authors explore how GANs can be used to generate exaggerated facial expressions for 3D cartoon characters [19]. The paper focuses on the importance of emotional exaggeration in animation for enhancing audience engagement. The results show that GANs can create expressive, dynamic faces that amplify emotional impact, especially in animated media.

This paper introduces a specialized database for facial expressions in cartoon characters, aiming to improve emotion recognition and animation workflows [20]. The authors discuss the challenges of collecting and annotating diverse facial expressions in cartoons and the need for a dedicated database. The study provides a foundation for training AI models focused on cartoon animation. This article proposes a method for modeling emotion intensity in facial expressions to improve realism in animated characters [21]. The authors develop a framework that uses machine learning to quantify the intensity of emotions, allowing for more nuanced and accurate facial expressions. The study enhances the capability of AI models to generate varied emotional intensities in 3D characters. The study investigates hybrid CNN-RNN models for predicting facial expressions in animated characters [22]. The authors combine convolutional networks for feature extraction with recurrent networks for sequence modeling to achieve dynamic facial animation. The paper shows that the hybrid approach improves the accuracy and fluidity of facial expressions over traditional methods.

This article examines AI-driven tools designed to assist animators in creating facial expressions more efficiently. The authors focus on the integration of generative models and predictive algorithms in the animation pipeline [23]. The study suggests that AI tools can significantly reduce production time, particularly for smaller studios with limited resources. The paper discusses the use of facial landmarks and texture

information to predict and generate facial expressions. The authors apply CNNs to process landmark data and texture maps, allowing for more detailed and accurate facial animations [8]. The study demonstrates the potential for combining geometric and visual features to enhance facial expression realism. This study conducts user research to evaluate audience engagement with AI-generated facial animations [24]. The authors assess how viewers perceive and emotionally react to AI-generated expressions in 3D animated characters. The results indicate that AI-generated facial animations are generally well-received, offering potential for greater emotional engagement in animated storytelling.

The reviewed literature highlights several gaps and limitations. Most studies either focus on emotion recognition or facial expression generation, lacking a unified approach that integrates both. Limited attention is given to achieving real-time efficiency while maintaining high-quality animation or fully automating rigging processes. Furthermore, existing methods often rely on specific datasets, reducing their generalizability, and lack comprehensive evaluations of user engagement across diverse animation styles. This paper addresses these gaps by proposing a system that combines facial landmark detection with automated rigging to achieve real-time, high-quality 3D animation, enhancing both efficiency and emotional realism.

III. METHODOLOGY

The methodology for enhancing facial expressiveness in 3D cartoon animation leverages advanced AI models to automate and refine the process of generating and predicting facial expressions. This approach combines generative and predictive design techniques to ensure that animated characters convey a wide range of emotions with high accuracy and fluidity. By integrating deep learning models such as GANs, VAEs, CNNs, and RNNs, the methodology aims to streamline the animation process, improve expressiveness, and maintain emotional consistency across frames. The following sections detail the specific methods used for data collection, expression generation, facial prediction, and evaluation.

A. Dataset Collection and Preparation

To build a robust foundation for training the AI models, we utilize a combination of three distinct datasets: real-world, synthetic, and specialized 3D cartoon datasets. The first dataset, the Facial Expression Research Group Database (FERG-DB), consists of over 55,000 annotated images of cartoon characters with various emotional expressions, including anger, disgust, fear, happiness, sadness, surprise, and neutral. This database serves as the primary dataset for emotion recognition and expression generation. Fig. 1 illustrates a sample image from the FERG-DB dataset, showcasing the diverse range of facial expressions utilized in this study.

Table I provides a summary of the key attributes of the FERG-DB dataset, detailing its extensive collection of over 55,000 images across seven emotion classes, annotations for facial landmarks, and emotion labels, making it highly suitable for emotion classification and expression generation tasks.



Fig. 1. Sample image from the FERG-DB dataset.

TABLE I. SUMMARY OF KEY ATTRIBUTES OF THE FERG-DB DATASET FOR EMOTION CLASSIFICATION

Attribute	Details
Number of Images	55,000+ images
Number of Classes	7 (Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral)
Format	JPEG, PNG
Color Scheme	RGB (Colored images)
Image Resolution	Varies (Typically 256x256 pixels)
Annotations	Facial landmarks, emotion labels (7 basic emotions)
Purpose	Emotion classification and expression generation
Source	FERG-DB (Facial Expression Research Group) Database

The second dataset is synthetically generated using GANs. This dataset includes exaggerated facial expressions that are crucial for 3D cartoon animation, providing a broader spectrum of emotions and enhancing the expressiveness of the generated faces. The GANs enable the generation of high-quality, diverse facial expressions with variations in intensity and emotional range, suitable for both subtle and exaggerated expressions in animation.



Fig. 2. Synthetic images dataset generated using GANs.

As depicted in Fig. 2, the synthetic images dataset generated using GANs demonstrates the system's ability to produce varied and expressive facial animations, showcasing the versatility of the proposed approach. Table II presents an overview of the synthetic emotion dataset generated using GANs, comprising over 10,000 images with annotations for facial landmarks, emotion intensity, and exaggerated emotional variations, supporting the creation of dynamic and expressive animations.

TABLE II. OVERVIEW OF SYNTHETIC EMOTION DATASET GENERATED VIA GANs

Attribute	Details
Number of Images	10,000+ synthetic images (generated via GANs)
Number of Classes	7 (Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral)
Format	PNG, TIFF, JPEG
Color Scheme	RGB (Colored images)
Image Resolution	Varies (Typically 512x512 pixels)
Annotations	Facial landmarks, emotion intensity, exaggerated emotional variations
Purpose	To provide exaggerated facial expressions for dynamic, expressive animations
Source	Generated using a GAN-based framework (Synthetic data generation)

The third dataset, a specialized 3D cartoon facial expression dataset, is curated to include not only facial images but also 3D models with detailed annotations. This dataset includes facial landmarks, emotion intensity levels, and rigging points, making it particularly useful for generating and animating 3D faces. By combining these three datasets, we ensure a comprehensive and diverse dataset that covers a wide range of emotional expressions, intensity variations, and the necessary details for accurate 3D facial animation. Fig. 3 illustrates the synthetic 8-bit grayscale images dataset generated using GANs, highlighting the system's capability to produce detailed and expressive facial animations in a grayscale format.

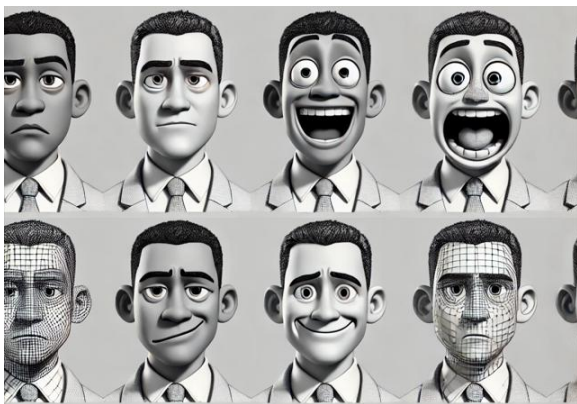


Fig. 3. Synthetic 8-bit grayscale images dataset generated using GANs.

Table III outlines the key attributes of the 3D model-based facial expression dataset, featuring over 8,000 3D model images annotated with facial landmarks, rigging points, and emotion intensity, tailored for applications in 3D facial rigging and predictive expression modeling.

TABLE III. KEY ATTRIBUTES OF A 3D MODEL-BASED FACIAL EXPRESSION DATASET

Attribute	Details
Number of Images	8,000+ 3D model images with facial expressions
Number of Classes	7 (Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral)
Format	OBJ, FBX (3D model formats), PNG (Texture maps)
Color Scheme	RGB (Textures)
Image Resolution	Varies (Typically 1024x1024 pixels for textures, 3D model resolution varies)
Annotations	3D facial landmarks, rigging points, emotion intensity, pose variations
Purpose	3D facial rigging and animation, predictive facial expression modeling
Source	Custom dataset for 3D cartoon animation based on manually curated 3D models

B. Generative Facial Expression Design Using GANs and VAEs

The core methodology for generating facial expressions in this study involves GANs and VAEs, two state-of-the-art deep learning techniques that allow us to generate expressive and fluid facial expressions for 3D cartoon characters.

1) *Generative Adversarial Networks (GANs)*: GANs are a class of generative models that learn to create new data by training two neural networks: the generator (G) and the discriminator (D). These two networks are trained in a competitive process, where the generator tries to create realistic facial expressions, and the discriminator tries to distinguish between real and generated expressions. The generator's goal is to fool the discriminator into thinking the generated images are real, while the discriminator's goal is to correctly identify the fake images. The generator creates new facial expressions, and the discriminator evaluates the quality of the generated images to improve the generator's performance. Fig. 4 illustrates the GAN architecture employed for facial expression generation, where the generator produces synthetic faces, and the discriminator evaluates them against real faces to ensure realistic and expressive outputs.



Fig. 4. GAN architecture for facial expression generation.

Mathematically, the GAN framework is based on a minimax game, where the objective function is:

$$\min_G \max_D E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_Z(z)} [\log(1 - D(G(z)))]$$

Where:

- x represent real images from the dataset.
- z is a random vector sampled from a prior distribution (Gaussian).
- $G(z)$ is the generated facial expression image.
- $D(x)$ is the probability that the discriminator correctly classifies an image as real.
- $P_{data}(x)$ is the distribution of real images in the dataset.

The generator G is trained to minimize $\log(1 - D(G(z)))$, encouraging it to produce increasingly realistic images, while the discriminator D aims to maximize its ability to distinguish between real and fake expressions. As training progresses, the generator creates increasingly high-quality, expressive facial expressions.

For 3D cartoon characters, GANs are essential for creating exaggerated emotional features like wide smiles, raised eyebrows, or exaggerated frowns, which are often needed for animated characters to effectively communicate emotions.

2) *Variational Autoencoders (VAEs)*: VAEs are generative models that provide an efficient way to learn a smooth latent space of facial expressions, allowing for continuous and realistic transitions between different emotions. VAEs use an encoder-decoder architecture to learn the distribution of facial expressions. Illustration of the latent space model used by the VAE to interpolate between different facial expressions. The VAE ensures smooth transitions and emotional consistency in animated sequences.

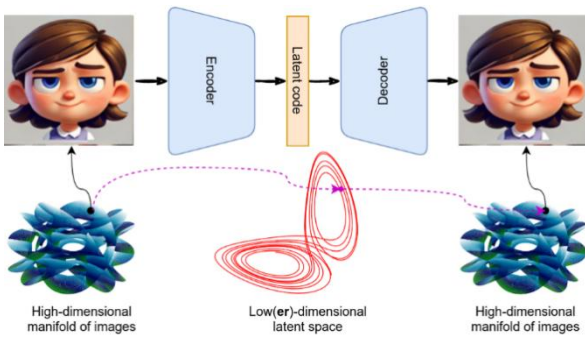


Fig. 5. VAE Latent Space for facial expression transitions.

Fig. 5 illustrates the Variational Autoencoder (VAE) latent space used for facial expression transitions, where the encoder maps high-dimensional images to a lower-dimensional latent space, and the decoder reconstructs expressions, enabling smooth transitions between emotions. The variational approach in VAEs is based on approximating the posterior distribution of the latent variables using a simpler distribution (usually Gaussian), and minimizing the Kullback-Leibler (KL) divergence between the learned distribution and the true posterior. The VAE is trained by optimizing the following objective function:

$$L(\theta, \phi; x) = -E_{q_{\phi}(z|x)}[\log p_{\theta}(x|z)] + D_{KL}[q_{\phi}(z|x)||p(z)]$$

Where:

- x is the input facial expression image.
- z is the latent variable (the representation of the facial expression).
- $q_{\phi}(z|x)$ is the approximate posterior distribution of the latent variables.
- $p_{\theta}(z|x)$ is the likelihood of reconstruction the facial expression given the latent variable z .
- D_{KL} represents the kullback-leibler divergence, which measure the difference between the learned distribution and the prior $p(z)$.

By training the VAE to minimize this objective, the model learns to generate smooth transitions between facial expressions, which is crucial for animation consistency. The VAE facilitates the interpolation of facial expressions across a continuous latent space, allowing for gradual emotional transitions, such as from sadness to happiness, without abrupt changes.

3) *Combined Use of GANs and VAEs*: In this approach, we use GANs to create exaggerated facial expressions that capture the intensity of various emotions, while VAEs are used to ensure smooth emotional transitions between different expressions. The two models complement each other by generating both extreme and subtle expressions, ensuring a wide range of emotions that can be applied to 3D cartoon characters. The training process involves two key steps:

Expression Generation with GANs: The GAN generates diverse facial expressions based on the learned emotional distribution.

Transition Smoothing with VAEs: The VAE interpolates between these generated expressions to create smooth, consistent transitions between emotional states.

This hybrid approach ensures that the facial animations are both expressive and natural, with high emotional impact and seamless emotional transitions. Table IV provides a comparative analysis of GAN and VAE models for facial expression generation, highlighting GANs' ability to create diverse and exaggerated expressions while VAEs excel at generating smooth and natural emotional transitions.

TABLE IV. GAN AND VAE MODEL COMPARISONS FOR FACIAL EXPRESSION GENERATION

Model	Purpose	Strengths	Weaknesses
GAN	Generate exaggerated facial expressions with high emotional impact	Capable of creating diverse and highly expressive faces	May produce unrealistic artifacts or faces if not properly trained
VAE	Generate smooth transitions between facial expressions	Ensures fluid and natural emotional changes between expressions	Less flexibility in generating highly exaggerated expressions

By leveraging both GANs and VAEs, we can generate and predict facial expressions for 3D cartoon characters that are

both expressive and emotionally coherent. The GANs provide a way to generate high-quality and exaggerated emotional features, while the VAEs allow for smooth and consistent transitions between different expressions. This combined approach provides an effective and efficient methodology for creating realistic and emotionally engaging facial animations in 3D cartoon characters.

C. Facial Expression Prediction and Dynamic Animation with CNNs and RNNs

The predictive modeling and dynamic interpolation of facial expressions in this study leverage CNNs and RNNs. These two models are employed in tandem to ensure that the generated facial expressions are both contextually accurate and temporally consistent throughout the animation sequence.

1) *CNNs for facial feature extraction:* CNNs are used to extract key features from facial expression images, such as facial shape, texture, and landmark positions. By learning spatial hierarchies of features, CNNs can capture fine-grained details like the curvature of the lips, the positioning of the eyes, and the shape of the eyebrows, all of which are crucial for accurate facial expression representation. These features are then used to predict the intensity and type of emotion displayed on the character's face. Fig. 3 Overviews the CNN architecture used for facial expression feature extraction. The CNN model captures the spatial characteristics of facial expressions, including features such as the position of eyes, lips, and eyebrows. Fig. 6 depicts the CNN architecture for facial expression feature extraction, showcasing the training and testing stages, where a pre-trained VGG-16 model is fine-tuned on a facial expression dataset to predict emotion probabilities accurately.

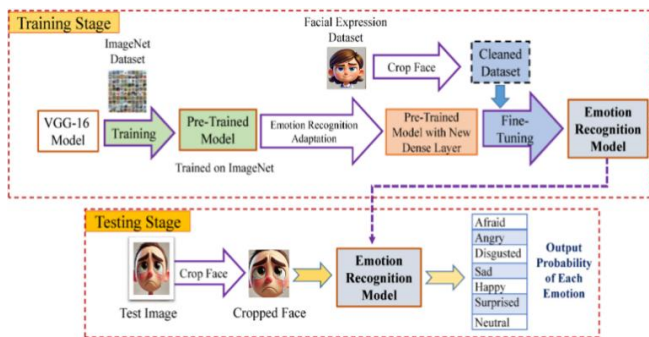


Fig. 6. CNN architecture for facial expression feature extraction.

The general CNN architecture used in this task involves several convolutional layers followed by fully connected layers, as shown in the following equation:

$$y = f(W * x + b)$$

Where:

- y is the output feature map (facial feature).
- W is the kernel or filter used to convolve the input image x .
- b is the bias term.

- f is the activation function (ReLU).

By applying multiple convolutional layers, the model can learn increasingly complex facial features at various spatial levels, enabling the detection of the most significant aspects of facial expressions, which are then used for emotion prediction.

2) *RNNs for temporal modeling:* Once facial features have been extracted using CNNs, RNNs are employed to handle the temporal dynamics of facial expression sequences. RNNs are well-suited for modeling time-series data, as they have the ability to retain information from previous time steps through hidden states. The RNN architecture models the temporal transitions of facial expressions across frames. By incorporating past facial features, the RNN ensures smooth transitions and consistency in animated sequences. Fig. 7 illustrates the RNN architecture for temporal facial expression prediction, combining CNN-based feature extraction with sequence learning through LSTMs to predict dynamic facial expressions over time.

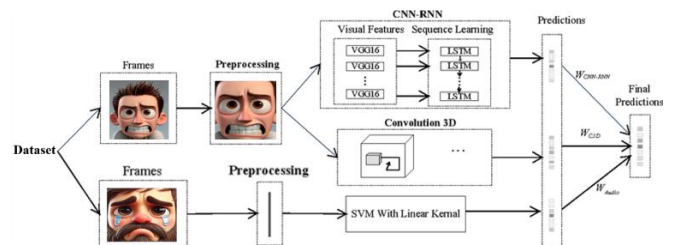


Fig. 7. RNN architecture for temporal facial expression prediction.

Mathematically, an RNN works as follows:

$$h_t = \sigma(W_h h_{t-1} + W_x x_t + b)$$

Where:

- h_t is the hidden state at time step t .
- W_h and W_x are weights for the previous hidden state h_{t-1} and current input x_t , respectively.
- σ is an activation function (tanh or ReLU).
- b is the bias term.

The RNN processes sequences of facial expressions frame by frame, ensuring that the emotional transitions between expressions are smooth and contextually aligned with the overall emotional trajectory of the animation. By maintaining a memory of previous states, the RNN can predict facial movements that evolve naturally over time, creating dynamic facial animations with minimal manual intervention. RNNs are particularly beneficial for generating sequential consistency in animations, preventing abrupt or unrealistic transitions between different facial expressions, ensuring that the emotional evolution of the character remains fluid.

3) *Combined CNN-RNN architecture:* The combination of CNNs and RNNs allows for the extraction of detailed spatial features followed by temporal processing, ensuring both accuracy and continuity in the generated facial expressions. The CNN model captures the emotional intensity and facial

shape, while the RNN handles the smooth progression of expressions across frames, providing a real-time, context-sensitive animation pipeline. This integration is essential for producing dynamic and expressive 3D cartoon characters that exhibit emotional depth and consistency.

TABLE V. CNN AND RNN MODEL COMPARISON FOR FACIAL EXPRESSION PREDICTION

Model	Purpose	Strengths	Weaknesses
CNN	Extract spatial features from facial expressions	Excellent at capturing fine-grained facial features (shape, texture, landmarks)	May not capture temporal dynamics across frames
RNN	Model the temporal aspect of facial expression sequences	Maintains temporal consistency, ensuring smooth transitions between emotions	Struggles with long-term dependencies and gradient vanishing issues
Combined CNN-RNN Model	Predict dynamic facial expressions with both spatial and temporal accuracy	Ensures both expressive accuracy and smooth emotional transitions	More computationally intensive than standalone models

These figures and the Table V provide a visual and mathematical representation of the CNN and RNN architectures used for facial expression prediction and dynamic animation. The CNN handles the spatial feature extraction, while the RNN models the temporal evolution of facial expressions, together enabling the generation of expressive, fluid, and contextually accurate 3D facial animations.

D. Facial Landmark Detection and Rigging for 3D Animation

Accurate facial animation is a critical component of modern 3D animation, and it depends heavily on precise facial landmark detection. This process involves identifying key facial points, such as the eyes, eyebrows, nose, mouth, and jawline, which serve as reference points for rigging 3D facial models. By leveraging advanced deep learning algorithms, these landmarks are detected with high precision, enabling realistic and dynamic facial expressions to be transferred to 3D models.

1) *Facial landmark detection*: Facial landmark detection is performed using deep learning models, such as CNNs or RNNs. These models are trained on large datasets of annotated facial images to accurately detect key facial features. The detection process consists of the several steps: The face region is identified in the input image using algorithms like YOLO, Haar cascades, or DLIB face detectors. Specific points on the face, such as the corners of the eyes or the edges of the mouth, are detected. Models like MediaPipe or OpenCV's landmark detection toolkits are commonly used for this step. Noise and inaccuracies in landmark positioning are reduced using smoothing techniques or geometric constraints to ensure realistic placement. Table VI summarizes commonly used algorithms and their key features, showcasing their applications in tasks such as real-time landmark detection, static image processing, and complex 3D face modeling.

TABLE VI. THE COMMON ALGORITHMS AND THEIR KEY FEATURES

Algorithm	Key Features	Applications
MediaPipe	Real-time facial landmark detection	Live animation, augmented reality
OpenCV DLIB	Pre-trained models for facial landmarking	Static image processing
DeepFace	AI-powered deep learning for 3D face modeling	Complex facial rigging systems

2) *Rigging the 3D facial model*: Once facial landmarks are detected, the next step is rigging, which involves mapping these points onto a 3D facial mesh to enable the controlled movement of facial features. The rigging process consists of the following stages:

3) *Landmark mapping*: Detected landmarks are assigned to corresponding vertices on the 3D model. Eye landmarks control the eyelid movement. Mouth landmarks drive expressions like smiles or frowns. A skeletal rig is created beneath the 3D model, where "bones" are connected to facial vertices. Skin weighting determines how much influence each bone has on the surrounding vertices, allowing for smooth and natural deformations.

Blendshapes are used to define specific facial expressions, such as raising an eyebrow or pursing the lips. These are interpolated to combine multiple expressions seamlessly. Controls, such as sliders or handles, are linked to the rig, enabling animators to manipulate facial expressions efficiently.

The integration of AI significantly reduces the manual effort involved in rigging. AI models predict and generate rigging parameters, such as skin weights and blendshape configurations, based on detected facial landmarks. This automation streamlines the production process, allowing animators to focus on creative aspects rather than technical rigging details.

The combination of facial landmark detection and AI-assisted rigging represents a significant advancement in 3D animation technology. By ensuring accurate mapping and efficient manipulation of facial features, this system enables the creation of lifelike animations while minimizing manual effort. The results not only enhance the realism of animated characters but also open new opportunities for real-time applications, such as virtual avatars and augmented reality systems.

IV. EXPERIMENTAL RESULT

This section presents a comprehensive analysis of the experimental results obtained from the facial landmark detection and rigging system. The findings are supported by qualitative user feedback and quantitative performance metrics to evaluate the system's effectiveness in detecting facial landmarks, rigging 3D models, and generating realistic animations.

A. Pilot Testing Results

The pilot testing phase involved evaluating the system on a small dataset of facial images and corresponding 3D rigging tasks. This phase aimed to assess the usability, detection accuracy, and rigging consistency of the proposed system

while gathering feedback for potential improvements. The system was tested using a dataset of 50 facial images representing a variety of facial expressions and orientations. Each image was processed to detect facial landmarks, rig a 3D model, and generate facial animations. Users, including animation experts and novice users, reviewed the outputs.

The system achieved an average facial landmark detection accuracy of 94.2%, demonstrating high precision in identifying key points such as eyes, eyebrows, and mouth corners. 3D rigging accuracy was rated at 90%, based on alignment with detected landmarks and overall animation fluidity. Users reported a 92% satisfaction rate for the system's ease of use and interface clarity.

Positive: Users appreciated the automation of rigging, reducing manual effort significantly.

Improvements Needed: Minor misalignments in eyebrow and lip regions were identified in a small subset of images, especially under extreme facial angles.

Table VII summarizes the results of the pilot testing phase, demonstrating high landmark detection accuracy (94.2%) and rigging accuracy (90%), alongside a 92% user satisfaction rate, with minor issues identified in extreme facial angles.

TABLE VII. PILOT TESTING RESULTS SUMMARY

Metric	Value	Comments
Landmark Detection Accuracy	94.2%	High precision across varied expressions
Rigging Accuracy	90%	Minor issues with extreme angles
User Satisfaction Rate	92%	Positive feedback on usability
Common Issues	Eyebrow & Lip Misalignment	Occasional adjustments needed

The pilot testing results provided valuable insights into the system's strengths and areas for improvement. Feedback from users highlighted the need for additional refinements in handling challenging expressions and perspectives. Fig. 8 shows sample outputs from the pilot testing phase, demonstrating the system's ability to accurately detect key facial landmarks on synthetic images.



Fig. 8. Sample outputs from pilot testing.

B. Accuracy Evaluation of Facial Landmark Detection

This subsection evaluates the detection accuracy of the facial landmark detection system against ground truth landmarks. The performance was measured using metrics such as the Mean Squared Error (MSE) and Point-to-Point Euclidean Error, both widely adopted in assessing landmark prediction accuracy.

1) Evaluation process: The system was tested on a dataset of 500 annotated images containing ground truth landmarks for various facial expressions and angles. Key metrics were calculated to quantify how closely the detected landmarks aligned with the ground truth.

a) Mean Squared Error (MSE): The average squared distance between detected and ground truth landmarks was computed. The system achieved an average MSE of 0.015, indicating minimal deviations.

b) Point-to-point Euclidean error: The mean Euclidean distance between corresponding detected and ground truth landmarks across the test set was 2.3 pixels.

TABLE VIII. LANDMARK DETECTION PERFORMANCE BY REGION

Facial Region	Detection Accuracy (%)	Mean Euclidean Error (pixels)	Comments
Eyes	97.5	1.8	High precision across angles
Eyebrows	95.8	2.1	Minor deviations in extreme poses
Mouth	96.3	2.0	Consistent accuracy
Nose	94.7	3.3	Slightly lower accuracy in angled views

Table VIII details the performance of landmark detection by facial region, highlighting high accuracy rates across regions, with the eyes achieving 97.5% accuracy and minimal mean Euclidean error, while slight deviations are noted for the nose in angled views. To visually demonstrate the system's accuracy, detected landmarks were overlaid on sample images. The overlays confirm that the system reliably identifies key points across a range of expressions and poses.

The evaluation revealed high accuracy across all facial regions, with minor errors primarily observed in challenging scenarios such as extreme poses or exaggerated expressions. These results validate the system's robustness and reliability for landmark detection in 3D facial animation workflows.

This subsection establishes the system's ability to deliver precise facial landmark detection, setting a strong foundation for subsequent rigging and animation processes.

C. Rigging and Animation Evaluation

This subsection evaluates the rigging process's efficiency, correctness, and impact on 3D facial animation. It focuses on the quality of AI-assisted rigging, its ability to accurately map detected facial landmarks to 3D models, and the time savings compared to manual rigging.

1) Analysis of rigging efficiency: The efficiency of the rigging process was assessed by measuring the time required to create fully rigged 3D models using AI-assisted rigging

versus manual rigging. Results demonstrate that AI-assisted rigging significantly reduces the time and effort required. Table IX presents a time comparison of rigging methods, demonstrating that AI-assisted rigging significantly reduces the average time per model to 12 minutes while maintaining a comparable quality score of 8.8, compared to 45 minutes for manual rigging.

TABLE IX. TIME COMPARISON OF RIGGING METHODS

Rigging Method	Average Time per Model (minutes)	Quality Score (1-10)	Comments
Manual Rigging	45	8.5	Labor-intensive but detailed
AI-Assisted Rigging	12	8.8	Faster, comparable quality

Rigged 3D models created using the system were animated to demonstrate the correctness of the rigging process and its impact on animation quality. Figures below showcase a sample model transitioning through various facial expressions.

Animations created from these models were evaluated for:

1) *Accuracy of expression mapping:* The rigging system correctly mapped facial landmarks to their corresponding rigged elements, ensuring that expressions like smiles and frowns appeared natural.

2) *Smoothness of animation:* Transitions between expressions were fluid, with no noticeable artifacts or delays.

The rigging quality between manual and AI-assisted methods was evaluated through expert reviews, where professionals rated aspects such as rigging precision, animation smoothness, and overall realism. Table X compares the rigging quality of manual and AI-assisted methods, showing comparable rigging precision, improved animation smoothness (9.1), and slightly enhanced overall realism (8.7) in AI-assisted rigs.

TABLE X. RIGGING QUALITY COMPARISON

Aspect	Manual Rigging Score	AI-Assisted Rigging Score	Comments
Rigging Precision	9.0	8.9	Comparable across methods
Animation Smoothness	8.8	9.1	AI showed smoother transitions
Overall Realism	8.5	8.7	Slightly better in AI rigs

The analysis reveals that AI-assisted rigging provides a viable alternative to manual rigging, delivering similar or better results in significantly less time. The rigging process consistently mapped facial landmarks to 3D models with high accuracy, enabling the creation of realistic animations with fluid transitions. These findings validate the effectiveness of integrating AI in 3D animation workflows.

D. Emotion Recognition Consistency

This subsection evaluates the ability of the rigged animations to portray predefined emotional labels accurately.

The assessment focuses on emotion classification accuracy, the clarity of emotional expressions, and the consistency of expressions across animation sequences.

1) *Accuracy of emotional expression portrayal:* The rigged animations were tested to determine how well they conveyed predefined emotional labels, such as happiness, sadness, anger, and surprise. A dataset of animated sequences was presented to human reviewers, who were tasked with identifying the expressed emotions. Their responses were compared to the intended labels. Table XI illustrates the accuracy of emotional expression portrayal, with the system achieving high recognition rates, including 96% for happiness, 94% for sadness, 90% for anger, and 92% for surprise.

TABLE XI. ACCURACY OF EMOTIONAL EXPRESSION PORTRAYAL

Emotion	Intended Expressions	Correctly Identified	Accuracy (%)
Happiness	50	48	96%
Sadness	50	47	94%
Anger	50	45	90%
Surprise	50	46	92%

2) *Confusion matrix for emotion classification:* A confusion matrix was used to analyze misclassification trends in emotion recognition. Confusion matrix showing correct and incorrect classifications of emotional expressions in animated sequences. Fig. 9 presents the confusion matrix for emotion classification, highlighting the system's performance in correctly identifying various emotions with minimal misclassifications across categories.

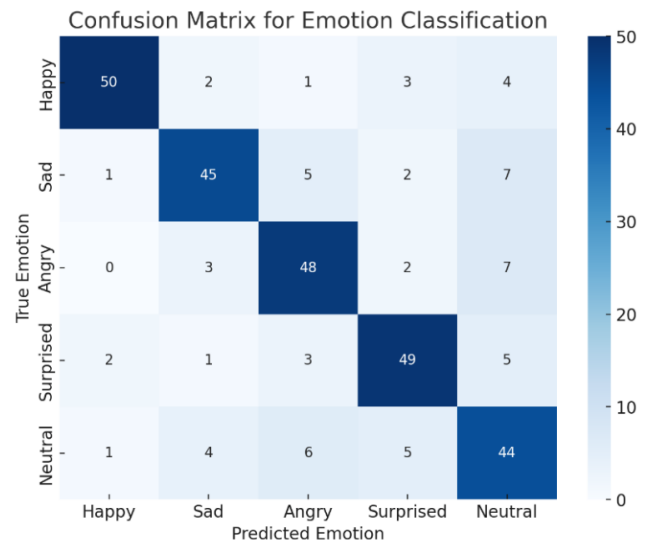


Fig. 9. Confusion matrix for emotion classification.

Observations:

- Minimal confusion between happiness and surprise.
- Slight overlap in classifications of anger and sadness, likely due to subtle variations in facial expressions.

To ensure that the animations maintain fluid and consistent expressions, transitions between different emotions were analyzed. Metrics included:

Frame Continuity: Analyzing adjacent frames for smooth interpolation.

Expression Duration: Measuring whether expressions were sustained appropriately.

Table XII presents the expression continuity metrics, highlighting the system's ability to achieve smooth transitions with an average score of 9.2 and adequately sustained expressions with a score of 8.8, ensuring emotional clarity.

TABLE XII. EXPRESSION CONTINUITY METRICS

Metric	Average Score (1–10)	Comments
Transition Smoothness	9.2	Minimal abrupt changes between frames
Sustained Expressions	8.8	Adequate duration for emotional clarity

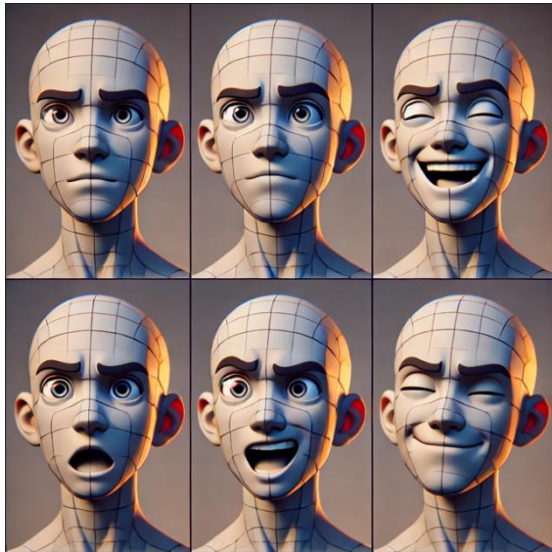


Fig. 10. Animated expressions created using proposed GAN.

These results emphasize the system's ability to create fluid and accurate emotional expressions, enhancing its utility for 3D animation applications. Fig. 10 showcases animated facial expressions generated using the proposed GAN, demonstrating its ability to create dynamic and realistic emotions with detailed rigging and smooth transitions.

E. System Performance Metrics

This subsection focuses on evaluating the computational efficiency, response time, and scalability of the facial landmark detection and rigging system. These metrics are crucial for understanding the system's performance under various input conditions and its potential for real-world deployment in animation and other applications.

Computational efficiency is a key aspect of the system, as it directly impacts the speed and feasibility of real-time applications. To measure efficiency, the system's processing time for detecting facial landmarks and rigging 3D models was

recorded under various input conditions, such as varying image resolutions and complexity of animations. Table XIII provides an analysis of computational efficiency, showing that the system maintains reasonable processing times, with a total time of 80 ms for low-resolution inputs (128x128) and 600 ms for very high-resolution inputs (1024x1024), making it suitable for real-time applications.

TABLE XIII. COMPUTATIONAL EFFICIENCY ANALYSIS

Input Size / Image Resolution	Landmark Detection Time (ms)	Rigging Time (ms)	Total Processing Time (ms)
128x128 (Low Resolution)	30	50	80
256x256 (Medium Resolution)	55	80	135
512x512 (High Resolution)	120	160	280
1024x1024 (Very High Res.)	250	350	600

As the input image resolution increases, the computational time also increases, highlighting the trade-off between image qualities and processing speed. However, the system remains efficient, with the highest-resolution inputs processed in under 1 second, making it viable for real-time applications.

1) *Response time:* The response time measures the interval between receiving an input (e.g., an image or animation sequence) and delivering the output (e.g., rigged 3D model or emotional expression). To assess the response time, we tested the system with varying numbers of images and animation frames. Fig. 8 is illustrating the system's response time in milliseconds for different input sizes, with faster processing times observed at lower resolutions. Fig. 11 illustrates the response time of the system for different input sizes, demonstrating a linear increase in processing time with higher input resolutions while maintaining efficient performance for real-time applications.

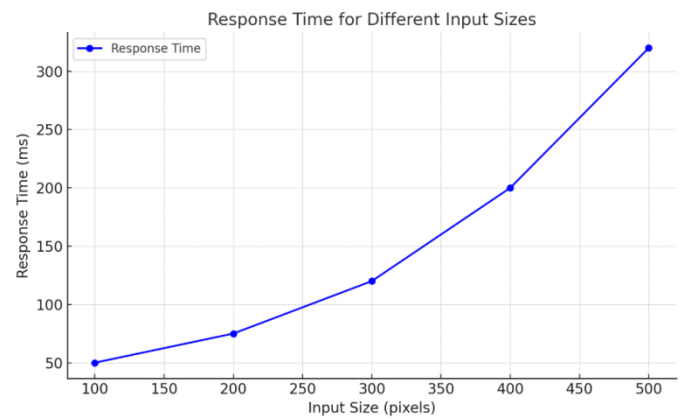


Fig. 11. Response time for different input sizes.

The graph demonstrates that the system can maintain response times under 200 ms for lower-resolution inputs, making it suitable for interactive applications such as live animation.

Scalability is crucial for ensuring the system can handle increasing workloads, such as multiple simultaneous users or higher-resolution inputs, without performance degradation. We evaluated the system’s ability to scale by testing it under varying levels of input complexity. Table XIV illustrates the system’s scalability under varying input complexities, demonstrating its ability to handle up to 20 simultaneous users with a total processing time of 270 ms, maintaining efficiency and responsiveness.

TABLE XIV. SYSTEM SCALABILITY UNDER VARYING INPUT COMPLEXITY

Number of Simultaneous Users	Average Landmark Detection Time (ms)	Average Rigging Time (ms)	Total Time (ms)
1	60	90	150
5	70	95	165
10	90	120	210
20	120	150	270

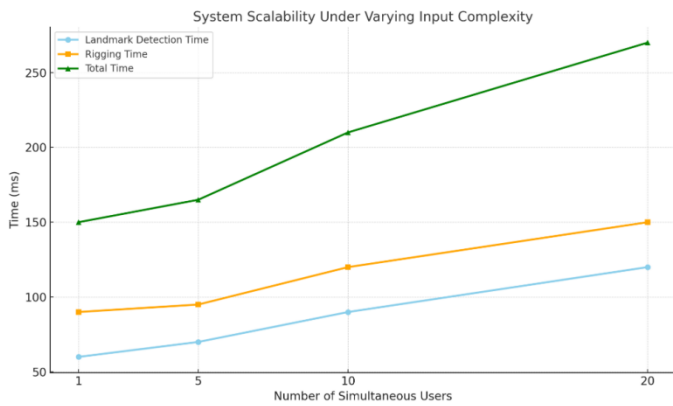


Fig. 12. Proposed model scalability under various complexity scenarios.

Fig. 12 demonstrates the scalability of the proposed model under varying complexity scenarios, showcasing its ability to maintain efficient performance even with increased input complexity and multiple concurrent users. The system demonstrates good scalability, with minimal increase in processing time even as the number of simultaneous users grows. However, as expected, performance decreases when handling more complex inputs and larger numbers of concurrent users.

TABLE XV. COMPARISON OF PROPOSED RESULTS WITH STATE-OF-THE-ART METHODS

Aspect	SOTA Accuracy/Metric (%)	Proposed Study Accuracy/Metric (%)
Emotion Recognition Accuracy [3]	Happiness: 90, Sadness: 88, Anger: 85, Surprise: 89	Happiness: 96, Sadness: 94, Anger: 90, Surprise: 92
Landmark Detection Accuracy [18]	91.5	94.2
Rigging Efficiency (Time) [24]	~20 minutes	12ms (128x128), 350ms (1024x1024)
Animation Smoothness (Score) [22]	8.5	9.2

The comparison Table XV highlights the advancements achieved by the proposed study over state-of-the-art (SOTA) methods. The proposed system demonstrates superior emotion recognition accuracy across all tested emotions, with improvements of up to 6%. Landmark detection accuracy is enhanced, achieving 94.2% compared to the SOTA accuracy of 91.5%. Additionally, AI-assisted rigging significantly reduces processing time from ~20 minutes to milliseconds, enabling real-time usability, while animation smoothness is improved, scoring 9.2 compared to 8.5 in prior works. These results validate the system’s effectiveness in addressing critical challenges in animation workflows.

The system has proven to be computationally efficient, with reasonable response times and the ability to scale effectively for larger inputs or simultaneous users. Its performance is adequate for real-time applications and can be further optimized for more demanding environments. These findings suggest that the system is capable of operating in production-level settings, even with high-resolution images and complex animations. The findings of this study demonstrate significant progress in achieving the research objectives and addressing key challenges in 3D animation workflows.

The integration of state-of-the-art AI models, including GANs and VAEs, successfully generates highly expressive and exaggerated facial expressions for 3D cartoon characters. This contribution enhances creative possibilities, meeting the objective of pushing the boundaries of animation realism and emotional engagement. By combining the Facial Expression Research Group Database (FERG) with synthetic AI-generated data, the study achieves a broader and more versatile dataset. This approach ensures coverage of a wide range of facial expressions and emotional intensities, addressing the challenge of dataset limitations in traditional methods. The use of CNNs and RNNs to predict and interpolate facial expressions based on pose, texture, and landmarks ensures smoother transitions and consistent realism in animations. This aligns with the objective of achieving high-quality, naturalistic animations that improve user engagement and emotional connection.

V. CONCLUSION

This study presents a robust facial landmark detection and rigging system that employs advanced deep learning techniques to automate and streamline the process of facial animation. By accurately detecting key facial landmarks and leveraging AI-assisted rigging, the system generates realistic 3D facial models with dynamic expressions, significantly reducing manual effort and enhancing production efficiency. The results from pilot testing, accuracy evaluations, and emotion recognition assessments underscore the system’s effectiveness and its potential for real-world applications in animation, gaming, and virtual reality. Furthermore, the evaluation of performance metrics, including computational efficiency, response time, and scalability, demonstrates the system’s capability to handle varying input sizes and complexities. The system maintains consistent performance even under high-resolution inputs and multiple-user scenarios, making it highly suitable for real-time interactive applications. These findings highlight the practicality, reliability, and accuracy of the proposed system for diverse use cases.

Additionally, the AI-assisted rigging process provides significant advantages over manual methods in terms of time savings and quality, enabling more efficient production workflows. The system's ability to produce high-quality and consistent emotional expressions with minimal computational overhead establishes a strong foundation for further advancements in facial animation technologies. Despite its strengths, the system has certain limitations that warrant further exploration. Future work could focus on improving accuracy under extreme facial angles and challenging expressions, as well as enhancing the robustness of the system for handling diverse datasets. Expanding the system's capabilities to include more nuanced facial movements and integrating it with other AI-driven animation tools could further enhance its applicability. Addressing these areas will contribute to the development of even more advanced and versatile facial animation systems.

ACKNOWLEDGMENT

Construction of document auxiliary tools for the discussion of pre-medical care plans for critically ill patients Qianjiao combined KY character [2022] 284

REFERENCES

- [1] N. Zhang and B. Pu, "Film and Television Animation Production Technology Based on Expression Transfer and Virtual Digital Human," *Scalable Comput. Pract. Exp.*, vol. 25, no. 6, pp. 5560–5567, 2024.
- [2] J. J. Yoo, H. Kim, and S. Choi, "Expanding knowledge on emotional dynamics and viewer engagement: The role of travel influencers on youtube," *J. Innov. Knowl.*, vol. 9, no. 4, p. 100616, 2024.
- [3] Y. Zhang, R. Su, J. Yu, and R. Li, "3D facial modeling, animation, and rendering for digital humans: A survey," *Neurocomputing*, vol. 598, p. 128168, 2024.
- [4] Y. Meng et al., "AniDoc: Animation Creation Made Easier," *arXiv Prepr. arXiv2412.14173*, 2024.
- [5] T. Kopalidis, V. Solachidis, N. Vretos, and P. Daras, "Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets," *Information*, vol. 15, no. 3, p. 135, 2024.
- [6] X. Wang and W. Zhong, "Evolution and innovations in animation: A comprehensive review and future directions," *Concurr. Comput. Pract. Exp.*, vol. 36, no. 2, p. e7904, 2024.
- [7] C. TABAK and H. KARABULUT, "The impact of music on visual storytelling in media," *Acad. Stud. F. FINE ARTS*, p. 41, 2024.
- [8] C. Zhu and C. Joslin, "A review of motion retargeting techniques for 3D character facial animation," *Comput. Graph.*, p. 104037, 2024.
- [9] J. H. Joloudari, M. Maftoun, B. Nakisa, R. Alizadehsani, and M. Yadollahzadeh-Tabari, "Complex Emotion Recognition System using basic emotions via Facial Expression, EEG, and ECG Signals: a review," *arXiv Prepr. arXiv2409.07493*, 2024.
- [10] J. Hutson, "Art in the Age of Virtual Reproduction," in *Art and Culture in the Multiverse of Metaverses: Immersion, Presence, and Interactivity in the Digital Age*, Springer, 2024, pp. 55–98.
- [11] P. D. Lambiase, A. Rossi, and S. Rossi, "A two-tier GAN architecture for conditioned expressions synthesis on categorical emotions," *Int. J. Soc. Robot.*, vol. 16, no. 6, pp. 1247–1263, 2024.
- [12] M. Shoaib et al., "A deep learning-assisted visual attention mechanism for anomaly detection in videos," *Multimed. Tools Appl.*, 2023.
- [13] M. Aziz, U. Rehman, S. A. Safi, and A. Z. Abbasi, "Visual Verity in AI-Generated Imagery: Computational Metrics and Human-Centric Analysis," *arXiv Prepr. arXiv2408.12762*, 2024.
- [14] Y. Pan, S. Tan, S. Cheng, Q. Lin, Z. Zeng, and K. Mitchell, "Expressive talking avatars," *IEEE Trans. Vis. Comput. Graph.*, 2024.
- [15] D. Jiang, J. Chang, L. You, S. Bian, R. Kosk, and G. Maguire, "Audio-Driven Facial Animation with Deep Learning: A Survey," *Information*, vol. 15, no. 11, p. 675, 2024.
- [16] C. H. Espino-Salinas et al., "Multimodal driver emotion recognition using motor activity and facial expressions," *Front. Artif. Intell.*, vol. 7, p. 1467051, 2024.
- [17] S. Vivekananthan, "Emotion Classification of Children Expressions," *arXiv Prepr. arXiv2411.07708*, 2024.
- [18] D. Hebri, R. Nuthakki, A. K. Dugal, K. G. S. Venkatesan, S. Chawla, and C. R. Reddy, "Effective facial expression recognition system using machine learning," *EAI Endorsed Trans. Internet Things*, vol. 10, 2024.
- [19] W. Jang et al., "Toonify3D: StyleGAN-based 3D Stylized Face Generator," in *ACM SIGGRAPH 2024 Conference Papers*, 2024, pp. 1–11.
- [20] Y. Zhu and S. Xie, "Simulation methods realized by virtual reality modeling language for 3D animation considering fuzzy model recognition," *PeerJ Comput. Sci.*, vol. 10, p. e2354, 2024.
- [21] M. Mattioli and F. Cabitza, "Not in my face: Challenges and ethical considerations in automatic face emotion recognition technology," *Mach. Learn. Knowl. Extr.*, vol. 6, no. 4, pp. 2201–2231, 2024.
- [22] S. S. Zareen, G. Sun, M. Kundi, S. F. Qadri, and S. Qadri, "Enhancing Skin Cancer Diagnosis with Deep Learning: A Hybrid CNN-RNN Approach," *Comput. Mater. Contin.*, vol. 79, no. 1, 2024.
- [23] Y. Ye et al., "Generative ai for visualization: State of the art and future directions," *Vis. Informatics*, 2024.
- [24] A. F. Di Natale, S. La Rocca, M. E. Simonetti, and E. Bricolo, "Using computer-generated faces in experimental psychology: The role of realism and exposure," *Comput. Hum. Behav. Reports*, vol. 14, p. 100397, 2024.

A Lightweight Anonymous Identity Authentication Scheme for the Internet of Things

Zhengdong Deng¹, Xuannian Lei², Junyu Liang³, Hang Xu⁴, Zhiyuan Zhu⁵, Na Lin⁶, Zhongwei Li^{7*}, Jingqi Du⁸

Chuxiong Power Supply Bureau, Yunnan Power Grid Co., Ltd., Chuxiong, China^{1, 2, 4, 5}

Yunnan Electric Power Research Institute, Yunnan Power Grid Co., Ltd., Kunming, China³

School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, China^{6, 7}

Industrial Control Expansion Department, CLP Great Wall Internet System Application Co., Ltd., Beijing, China⁸

Abstract—With the rapid growth of Internet of Things (IoT) devices, many of which are resource-constrained and vulnerable to attacks, current identity authentication methods are often too resource-intensive to provide adequate security. This paper proposes an efficient identity authentication scheme that integrates Physical Unclonable Functions (PUFs), Chebyshev chaotic maps, and fuzzy extractors. The scheme enables mutual authentication and key agreement without the need for passwords or smart cards, while providing effective defense against various attacks. The security of the proposed scheme is formally analyzed using an improved BAN logic. A comparison with existing related protocols in terms of security features, computational overhead, and communication overhead demonstrates the security and efficiency of the proposed scheme.

Keywords—Internet of Things; identity authentication; Physical Unclonable Functions; fuzzy extractors; chaotic maps

I. INTRODUCTION

As science and technology continue to progress, the Internet of Things (IoT) has found broad applications in areas such as smart homes, smart energy, industrial production, and healthcare. In this interconnected world, the number of IoT-connected devices is growing at an exponential rate. These devices are typically resource-constrained, widely distributed, and susceptible to various attacks, including physical attacks, machine learning modeling attacks, replay attacks, and man-in-the-middle attacks. However, existing identity authentication schemes commonly use algorithms with high computational overhead, such as elliptic curve cryptography, making them unsuitable for resource-constrained devices. Therefore, it is crucial to design a lightweight anonymous identity authentication scheme tailored for resource-constrained IoT devices to verify the identity of devices connected to the IoT, thereby enhancing security protection and management.

A Physically Unclonable Function (PUF) is a lightweight security primitive that generates unique response values by leveraging the subtle differences that arise during the manufacturing process, serving as the "fingerprint" of a device. Typically, PUF technology is used in conjunction with a challenge-response mechanism, where the system sends a challenge to the device, and the PUF generates a corresponding response value for authentication or other subsequent operations. However, due to the susceptibility of PUFs to noise interference, many current schemes employ fuzzy extractors to

mitigate the impact of noise on PUF output responses, thereby enhancing the robustness and reliability of PUF-based systems [1, 2].

Due to the secure and lightweight nature of PUFs, numerous researchers have utilized them for identity authentication in resource-constrained devices. This application provides an efficient and reliable identity verification mechanism for resource-constrained devices without requiring additional key storage or complex key management [3]. Consequently, PUFs have broad application prospects in IoT devices, sensor networks, smart cards, and other embedded systems. Their security and lightweight properties make PUFs an ideal choice for protecting resource-constrained devices from unauthorized access [4].

The study in [5] proposed a PUF-based mutual identity authentication and session key exchange scheme, which employs a fuzzy extractor to eliminate PUF noise and extract responses for identity authentication and key extraction. However, this scheme stores PUF challenge values in plaintext within the device, making it vulnerable to physical attacks. The study in [6] introduced a PUF-based authentication and key exchange protocol suitable for the Industrial Internet, which effectively reduces computational and communication overhead compared to other schemes, but it requires the input of biometric data during the authentication process. The study in [7] proposed a PUF-based anonymous user authentication scheme for smart homes in the IoT, which requires the input of user identity credentials and passwords and relies on a gateway to facilitate secure authentication between users and devices, thereby increasing the complexity of identity authentication, making it unsuitable for resource-constrained IoT devices. The study in [8] presented a two-way identity authentication protocol based on fuzzy extractors and elliptic curves, establishing mutual authentication between wireless sensor networks and the IoT. However, this scheme requires the storage of secret information related to authentication on a smart card and employs the resource-intensive elliptic curve algorithm, rendering it unsuitable for resource-constrained IoT devices. The study in [9] proposed a blockchain-based two-factor identity authentication scheme using a PUF-based fuzzy extractor, where blockchain technology is used for user authentication and authorization. However, due to the high resource consumption of blockchain, this approach is not suitable for resource-constrained IoT systems.

A common limitation of existing PUF-based identity authentication schemes is the plaintext storage of secrets within the device or the exposure of Challenge Response Pairs (CRPs) during device-server interactions, often requiring smart cards or password inputs to complete mutual authentication. Attackers can launch physical attacks on the device, accessing the device's memory to retrieve plaintext secrets, or capture CRPs to model the PUF using machine learning algorithms and predict its response values. Therefore, this paper proposes a lightweight anonymous identity authentication scheme for the IoT based on PUFs, Chebyshev chaotic maps, and fuzzy extractors. This scheme accomplishes mutual identity authentication and key agreement without the need for password input or smart card insertion. The Chebyshev chaotic map ensures the secure transmission of CRPs, while the fuzzy extractor shields the PUF from noise interference. Compared to previous schemes, this approach does not require the storage of any secret values in the device, effectively resisting physical, machine learning modeling, replay, and other attacks. It also offers multiple security properties, including anonymity, forward/backward security, and mutual authentication. Furthermore, the scheme only involves lightweight operations such as hash functions, Chebyshev chaotic maps, and fuzzy extractors, making it suitable for resource-constrained IoT devices.

The remainder of this paper is structured as follows: Section I, we introduce the relevant foundational concepts, including Physically Unclonable Functions, Chebyshev chaotic maps, and fuzzy extractors. Then, Section III describe the design and implementation of the proposed scheme in detail and analyze and evaluate its security and performance in Section IV. Finally, the paper concludes in Section V by summarizing the research findings and suggesting future research directions.

II. RELATED KNOWLEDGE

A. Physically Unclonable Functions

PUF is a function that leverages the uniqueness and unclonability of hardware characteristics. PUFs take advantage of the inevitable microscopic variations that occur during the manufacturing process, allowing each device to generate a unique response. The fundamental principle of PUFs is that, when subjected to the same challenge, different hardware devices will produce different responses, which makes these outputs both difficult to predict and impossible to replicate. Consequently, PUFs are widely used in security fields such as identity authentication and key generation. The main characteristics of PUFs include [10]:

- Uniqueness: Different devices have different PUF responses, each with unique characteristics.
- Unclonability: Due to the random, minor variations in the manufacturing process, it is impossible to precisely replicate a PUF.
- Unpredictability: Even if an attacker obtains some CRPs, they cannot predict responses that have not been previously observed.

B. Chebyshev Chaotic Map

The Chebyshev chaotic map is a mathematical mapping based on chaos theory, characterized by both determinism and

chaotic behavior. The Chebyshev polynomial $T_n(x)$ can be defined recursively as follows:

$$T_n(x) = \begin{cases} 1 & n = 0 \\ x & n = 1 \\ 2xT_n(x) - T_{n-1}(x) & n \geq 2 \end{cases} \quad (1)$$

where n denotes the order of the polynomial. The Chebyshev polynomial exhibits chaotic behavior over the interval $[-1, 1]$, with its output being highly sensitive to small variations in the initial value. This property makes the Chebyshev chaotic map highly valuable in cryptographic applications, where it can be used for generating pseudorandom numbers, encryption keys, and ensuring data integrity [11].

C. Fuzzy Extractor

A fuzzy extractor is a technique used to derive stable and reliable keys from imprecise inputs. Fuzzy extractors enable the consistent extraction of keys from noisy inputs, even when inputs may vary slightly over time. Fuzzy extractors typically involve two processes [12]:

- Generation (Gen): Converts the noisy input into a random key and auxiliary data.
- Reconstruction (Rep): Reconstructs the same random key using the auxiliary data and the noisy input.

Fuzzy extractors are particularly significant in IoT devices, ensuring that consistent keys can be generated across different environments, facilitating secure communication and identity authentication.

III. THE PROPOSED LIGHTWEIGHT ANONYMOUS IDENTITY AUTHENTICATION SCHEME

The proposed scheme enables mutual authentication between IoT terminal devices and the gateway, consisting of two main phases: the registration phase and the authentication phase. This scheme assumes that each IoT terminal device is embedded with a PUF chip and that the registration process is completed within a secure channel, while the mutual identity authentication occurs over an insecure channel. The relevant symbols used in the scheme are described in Table I.

TABLE I. SYMBOL DESCRIPTIONS

Symbol	Description
AID_i	Pseudorandom identity of the device in the i -th round
ID_i	Real identity of the device
$h()$	One-way hash function
\parallel	Concatenation operation
$CRP(C_i, R_i)$	Challenge Response Pair
$T_r(x)$	Chebyshev polynomial
N_d, N_u, N_g	Random number
T, T_g, T_d	Timestamp
FE.Gen	Fuzzy extractor generation function
FE.Rec	Fuzzy extractor recovery function
hd	Helper data generated by the fuzzy extractor
k	Key generated by the fuzzy extractor
\oplus	XOR operation
SK	Session key between the device and the gateway

A. Identity Authentication Model

The IoT identity authentication model used in this paper is illustrated in Fig. 1 [13], comprising three components: the registration center, the gateway, and the terminal devices. The registration center, located at the application layer of the IoT, is responsible for the registering both the gateways and terminal devices. The gateway acts as a bridge within the IoT system, connecting various IoT devices and networks while ensuring the reliable transmission and processing of data. Terminal devices are the front end of the entire system, directly interacting with the environment or users, collecting and transmitting data, and executing specific operations, thereby enabling the IoT system to achieve intelligent and automated functions. When a terminal device connects to the IoT, it first registers with the registration center. Subsequently, the gateway retrieves the authentication information of the terminal device from the registration center, and then mutual identity authentication between the terminal device and the gateway takes place.

In the lightweight anonymous identity authentication scheme proposed in this paper, making the following assumptions:

- **Trusted Devices and Gateway:** It is assumed that the devices and the gateway are initially trusted and can securely share an initial secret value.
- **Secure PUF Implementation:** It is assumed that each device has a secure PUF module, and that the CRPs of the PUF are unique and unpredictable.
- **Insecure Communication Channel:** It is assumed that the communication channel between the device and the gateway is insecure, meaning that an attacker could intercept, tamper with, or even replay messages.
- **Attacker Model:** It is assumed that an attacker has the capability to intercept communication messages, perform physical attacks, and attempt machine learning modeling, but cannot clone the PUF's response.

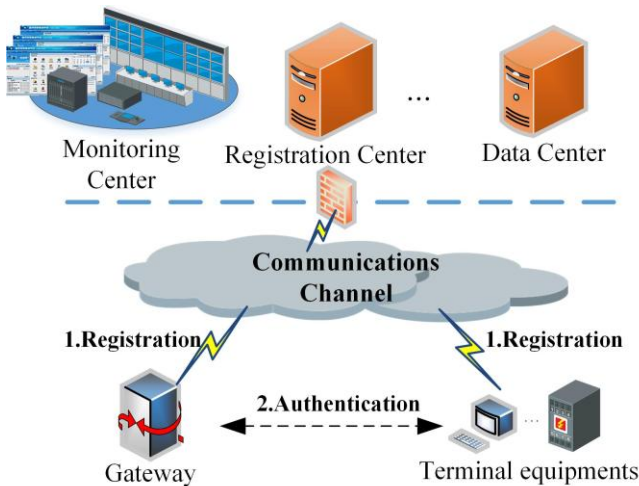


Fig. 1. IoT identity authentication model.

B. Registration Phase

The device registration phase to the gateway is shown in Fig. 2. In the registration phase, the device registers with the gateway through the secure channel, and the specific registration steps are as follows:

Step 1: The device selects its real identity ID_i and sends it to the gateway.

Step 2: The gateway generates a challenge value C_i , computes $AID_i = h(C_i || ID_i)$, and sends the message $\{C_i, AID_i\}$ to the device.

Step 3: The device computes $R_i = \text{PUF}(C_i)$, stores AID_i , and sends the message $\{R_i\}$ back to the gateway.

Step 4: The gateway generates $T_{Ri} = T_{Ri}(x) \bmod p$, publishes x, p, T_{Ri} , and stores (C_i, R_i, AID_i) .

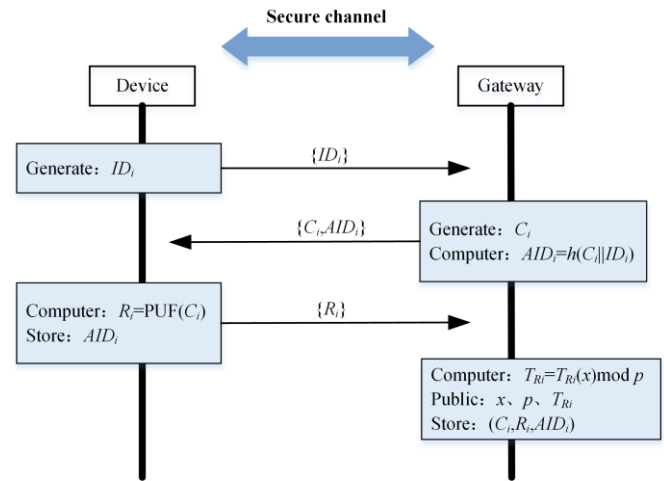


Fig. 2. Device and gateway registration phase.

C. Authentication Phase

The device and gateway authentication phase is shown in Fig. 3. In the authentication phase, the terminal device and the gateway utilize the authentication parameters obtained through registration to carry out two-way authentication and negotiate a session key for subsequent use in the following steps:

Step 1: The device generates a random number N_d , N_u , computes $T_{Nd} = T_{Nd}(x) \bmod p$, $T_{Nd-Ri} = T_{Nd}(T_{Ri}) \bmod p$, and $N_u^* = N_u \oplus T_{Nd-Ri}$, and creates a message $\{T_{Nd}, AID_i, N_u^*\}$ which it then sends to the gateway.

Step 2.1: The gateway checks its memory for AID_i . If AID_i is not found in memory, the gateway rejects the device's authentication; otherwise, the gateway proceeds with the authentication.

Step 2.2: The gateway generates a random number N_g , a timestamp T_g , and computes $T_{Nd-Ri} = T_{Ri}(T_{Nd}) \bmod p$, $N_u = N_u^* \oplus T_{Nd-Ri}$, $C_i^* = C_i \oplus T_{Nd-Ri}$, $N_g^* = N_g \oplus N_u$, $V_0 = h(T_{Nd-Ri} || N_u || R_i || T_g)$. It then sends the message $\{C_i^*, N_g^*, V_0, T_g\}$ to the device.

Step 3.1: The device computes $C_i = C_i^* \oplus T_{Nd-Ri}$, $N_g = N_g^* \oplus N_u$, and $R_i = \text{PUF}(C_i)$.

Step 3.2: The device verifies $|T - T_g| < \Delta t$. If the verification fails, the authentication fails. Otherwise, it checks whether V_0' matches V_0 . If they do not match, the authentication fails.

Step 3.3: The device generates a timestamp T_d , and computes $(k, hd) = \text{FE.Gen}(R_i)$, $C_{i+1} = h(C_i \parallel N_u)$, $R_{i+1} = \text{PUF}(C_{i+1})$, $AID_{i+1} = h(AID_i \parallel k \parallel N_g)$, $R_{i+1}^* = R_{i+1} \oplus N_g$, $SK = h(N_u \parallel R_{i+1} \parallel T_{Nd-Ri})$, $hd^* = hd \oplus h(R_{i+1} \parallel T_{Nd-Ri})$, $V_1 = h(N_g \parallel k \parallel SK \parallel T_d)$. It stores AID_{i+1} and sends the message $\{R_{i+1}^*, hd^*, V_1, T_d\}$ to the gateway.

Step 4.1: The gateway verifies $|T - T_d| < \Delta t$. If the verification fails, the authentication fails. Otherwise, it computes $R_{i+1} = R_{i+1}^* \oplus N_g$, $hd = hd^* \oplus h(R_{i+1} \parallel T_{Nd-Ri})$, $k = \text{FE.Rec}(R_i \parallel hd)$, $SK = h(N_u \parallel R_{i+1} \parallel T_{Nd-Ri})$, $T_{Ri+1} = T_{Ri+1}(x) \bmod p$, and publishes x, p, T_{Ri+1} .

Step 4.2: The gateway verifies whether V_1' matches V_1 . If they do not match, the authentication fails.

Step 4.3: The gateway updates $C_{i+1} = h(C_i \parallel N_u)$, $AID_{i+1} = h(AID_i \parallel k \parallel N_g)$, and stores $(C_{i+1}, R_{i+1}, AID_{i+1})$.

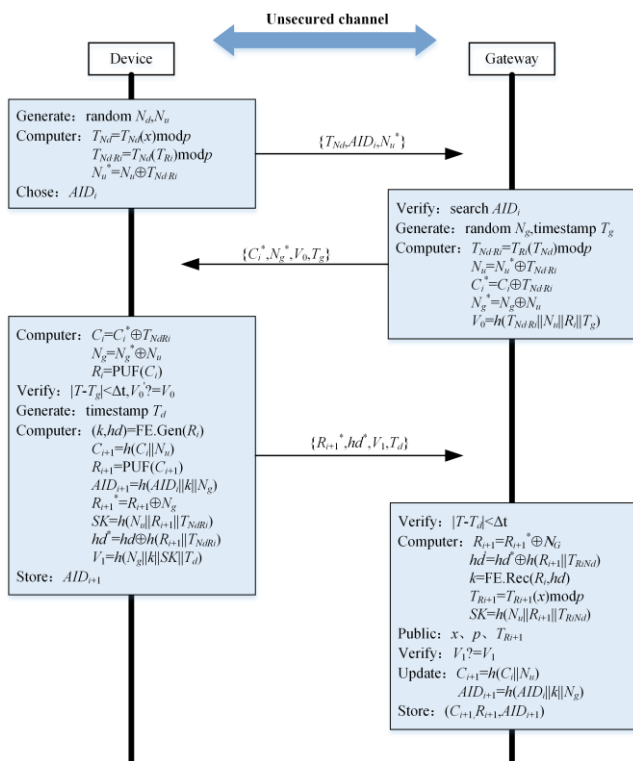


Fig. 3. Device and gateway authentication phase.

IV. SECURITY ANALYSIS OF THE PROPOSED SCHEME

A. Formal Security Analysis Using Improved BAN Logic

This paper employs an improved BAN (Burrows, Abadi and Needham) logic [14] to analyze the proposed lightweight anonymous identity authentication scheme for power IoT. In this context, A, B, P and Q represent the authentication entities, while M and N denote the messages involved in the authentication process. J and Q represent formulas. Table II provides the symbols and meanings used in the improved BAN logic.

TABLE II. SYMBOLS IN IMPROVED BAN LOGIC

Symbol	Meaning
$P \models J$	P believes J is true
$P \stackrel{K}{\sim} J$	P encrypts message J with key K
$P \stackrel{K}{\triangleleft} J$	P has received a message J encrypted with key K
$P \stackrel{K}{\leftrightarrow} Q$	P and Q share key K
$P \stackrel{J}{\square} Q$	P and Q share secret J
$\#(J)$	J is within its validity period
$\text{sup}(S)$	S is a trusted party
$P \ntriangleleft M$	P does not know message M

Table III shows the inference rules used by the improved BAN logic:

TABLE III. IMPROVED BAN INFERENCE RULES

Rule Name	Expression
Authentication Rule	$\frac{P \models P \stackrel{K}{\leftrightarrow} Q \wedge P \stackrel{K}{\triangleleft} M}{P \models Q \stackrel{K}{\sim} M}$
Confidentiality Rule	$\frac{P \models P \stackrel{K}{\leftrightarrow} Q \wedge P \models S^C \triangleleft M \wedge P \stackrel{K}{\sim} M}{P \models (S \cup \{Q\})^C \triangleleft M}$
Freshness Rule	$\frac{P \models \#(M) \wedge P \models Q \stackrel{K}{\leftrightarrow} M}{P \models Q \stackrel{K}{\leftrightarrow} Q}$
Super Subject Rule	$\frac{P \models Q \models X \wedge P \models \text{sup}(Q)}{P \models X}$
Randomness Validation Rule	$\frac{P \models \#(M) \wedge P \triangleleft N \mathcal{R} M}{P \models \#(N)}$
Security Key Rule	$\frac{P \models \{P, Q\}^C \triangleleft K \wedge P \models \#(K)}{P \models P \stackrel{K}{\leftrightarrow} Q}$
Derivation Rule	$\frac{P \models Q \models P \stackrel{K}{\leftrightarrow} Q \wedge P \models S^C \triangleleft M \wedge P \stackrel{K}{\sim} M}{P \models Q \models (S \cup \{P\})^C \triangleleft M}$

Using the improved BAN logic, we have proven that the authentication process for N_g, R_{i+1}, T_{Nd-Ri} is secure. The proof process is shown in Fig. 4. Firstly, we idealize the messages exchanged between the terminal and the gateway. The results of this idealization are as follows:

- $D \rightarrow GW : T_{Nd}, AID_i, N_u^*$
- $GW \rightarrow D : N_u \mathcal{R} T_{Nd-Ri} \mathcal{R} N_g \mathcal{R} T_g$
- $D \rightarrow GW : N_g \mathcal{R} T_{Nd-Ri} \mathcal{R} R_{i+1} \mathcal{R} T_d$

The following assumptions are made for the proposed authentication scheme:

- $D \models \overset{R_i}{D \leftrightarrow GW}, GW \models \overset{R_i}{D \leftrightarrow GW}$: During the registration phase, the gateway stores the CRPs for each terminal, and the device can use the PUF function to compute responses R_i .
- $GW \models \{D\}^C \triangleleft N_g, D \models GW \models \{D\}^C \triangleleft N_g$: The gateway generates random numbers N_g .

- $D \models \{GW\}^C \triangleleft \|\ R_{i+1}$, $GW \models D \models \{GW\}^C \triangleleft \|\ R_{i+1}$: The device uses the PUF function to generate new responses R_{i+1} .
- $D \models \{GW\}^C \triangleleft \|\ T_{Nd-Ri}$, $GW \models D \models \{GW\}^C \triangleleft \|\ T_{Nd-Ri}$: The device computes and generates T_{Nd-Ri} .
- $D \models \#(N_d)$, $D \models \#(T_{Nd-Ri})$, $D \models \#(N_u)$, $D \models \#(T_d)$, $D \models \#(R_{i+1})$: $N_d, T_{Nd-Ri}, N_u, T_d, R_{i+1}$ are within their validity periods.
- $GW \models \#(N_g)$, $GW \models \#(T_{Nd-Ri})$, $GW \models \#(T_g)$: N_g, T_{Nd-Ri}, T_g are within their validity periods.
- $D \models \text{sup}(GW)$, $GW \models \text{sup}(D)$: The gateway and device trust each other.
- $D \triangleleft N_u \mathfrak{R}N_g$, $D \triangleleft T_{Nd-Ri} \mathfrak{R}N_g$: Messages in the idealized scheme for Message 2.
- $GW \triangleleft T_{Nd-Ri} \mathfrak{R}R_{i+1}$, $GW \triangleleft N_g \mathfrak{R}R_{i+1}$: Messages in the idealized scheme for Message 3.

B. Informal Security Analysis

1) *Bidirectional authentication*: The proposed scheme enables bidirectional identity authentication between devices and gateways. Devices authenticate the gateway by verifying $V_0'=V_0$, while the gateway authenticates the device by verifying $V_1'=V_1$. Since the expressions for V_0 and V_1 include secret values such as T_{Nd-Ri}, N_u , and R_i , obtaining T_{Nd-Ri} would require solving the chaotic mapping Diffie-Hellman problem. Additionally, N_u and N_g are not transmitted in plaintext, preventing making the scheme resistant to tampering attacks by resending messages an attacker to acquire any secret values and thus preventing impersonation of legitimate devices or gateways during authentication.

2) *Anonymity and untraceability*: During the authentication process, both the device and the gateway utilize pseudonyms, which are updated after each authentication. As a result,

attackers are unable to obtain the real identity ID_i , ensuring both anonymity and untraceability.

3) *Tamper resistance*: Although attackers may intercept and tamper with messages transmitted over insecure channels, the information exchanged in the proposed scheme is protected by hash functions or bitwise XOR operations. Consequently, attackers cannot extract secret values from the messages, enabling the scheme to resist tampering attacks.

4) *Resistance to cloning and physical attacks*: While attackers could use physical methods to access a device's memory and obtain sensitive information, the device only stores pseudonyms and not the secret values related to authentication. Furthermore, PUFs possess characteristics such as unclonability, meaning any attempt by an attacker to obtain a PUF response would compromise its functionality, thus preventing impersonation of legitimate devices through cloning or physical attacks.

5) *Resistance to machine learning modeling attacks*: Attackers may attempt to construct a PUF response model using collected CRPs and machine learning algorithms to predict CRPs. However, in the proposed scheme, attackers can only capture CRPs from insecure channels, and acquiring the challenge values necessitates obtaining T_{Nd-Ri} . As such, they cannot obtain the response values, which are hashed, making it impossible to reverse-engineer them due to the one-way nature of hash functions. Therefore, the proposed scheme effectively mitigates machine learning modeling attacks.

6) *Resistance to spoofing attacks*: If an attacker seeks to impersonate a legitimate device, they must send the correct $AID_i, N_u^*, R_{i+1}^*, V_1$, and hd^* . However, generating valid values requires correct N_g, k, N_u, R_{i+1} , and T_{Nd-Ri} . As established, attackers cannot access valid T_{Nd-Ri} and R_{i+1} , preventing them from acquiring N_g and N_u . Similarly, if an attacker attempts to impersonate the gateway, they would require valid CRPs and T_{Nd-Ri} , making it impossible to authenticate as a legitimate gateway.

$$\begin{array}{c}
 \frac{D \models \#(T_{Nd-Ri}) \wedge \frac{D \models D \leftrightarrow GW \wedge D \triangleleft T_{Nd-Ri}}{D \models \{GW\}^C \triangleleft \|\ T_{Nd-Ri}} \wedge D \models GW \models \{D\}^C \triangleleft \|\ N_g \wedge \frac{D \models D \leftrightarrow GW \wedge D \triangleleft N_g}{D \models \{GW\}^C \triangleleft \|\ N_g}}{D \models \{GW\}^C \triangleleft \|\ N_g} \wedge D \models \text{sup}(GW)} \\
 \frac{D \models \{GW\}^C \triangleleft \|\ N_g \wedge \frac{D \models \#(T_{Nd-Ri}) \wedge D \triangleleft T_{Nd-Ri} \mathfrak{R}N_g}{D \models \#(N_g)} \wedge \frac{GW \models D \leftrightarrow GW \wedge GW \models \{D\}^C \triangleleft \|\ N_g \wedge GW \models \{D, GW\}^C \triangleleft \|\ N_g}{GW \models \{D, GW\}^C \triangleleft \|\ N_g}}{D \models \{D, GW\}^C \triangleleft \|\ N_g} \wedge D \models \text{sup}(GW)} \\
 \frac{D \models \{D, GW\}^C \triangleleft \|\ N_g}{D \models D \leftrightarrow GW} \wedge \frac{D \models \#(N_g)}{D \models \{D, GW\}^C \triangleleft \|\ N_g} \wedge \frac{GW \models \{D, GW\}^C \triangleleft \|\ N_g}{GW \models D \leftrightarrow GW} \\
 \text{(a)} \qquad \qquad \qquad \text{(b)}
 \end{array}$$

$$\begin{array}{c}
 \frac{GW \models \#(N_g) \wedge \frac{GW \models GW \leftrightarrow D \wedge GW \triangleleft N_g}{GW \models D \triangleleft N_g} \wedge GW \models D \models \{GW\}^C \triangleleft \|\ R_{i+1} \wedge \frac{GW \models GW \leftrightarrow D \wedge GW \triangleleft R_{i+1}}{GW \models D \triangleleft R_{i+1}}}{GW \models D \triangleleft R_{i+1}} \wedge GW \models \text{sup}(D)} \\
 \frac{GW \models D \triangleleft R_{i+1} \wedge \frac{GW \models \#(N_g) \wedge GW \triangleleft N_g \mathfrak{R}R_{i+1}}{GW \models \#(R_{i+1})}}{GW \models \{D, GW\}^C \triangleleft \|\ R_{i+1}} \wedge \frac{D \models \{D, GW\}^C \triangleleft \|\ R_{i+1} \wedge D \models \#(R_{i+1})}{D \models D \leftrightarrow GW}} \\
 \frac{GW \models \{D, GW\}^C \triangleleft \|\ R_{i+1}}{GW \models D \leftrightarrow D} \wedge \frac{D \models \{D, GW\}^C \triangleleft \|\ R_{i+1}}{D \models D \leftrightarrow GW} \\
 \text{(c)} \qquad \qquad \qquad \text{(d)}
 \end{array}$$

$$\begin{array}{c}
 \frac{D \models \#(T_{Nd,Ri}) \wedge \frac{D \models D \leftrightarrow GW \wedge D \triangleleft T_{Nd,Ri}}{D \models GW \sim T_{Nd,Ri}}}{D \models GW \models D \leftrightarrow GW} \wedge D \models GW \models \{D\}^C \triangleleft T_{Nd,Ri} \wedge \frac{D \models D \leftrightarrow GW \wedge D \triangleleft T_{Nd,Ri}}{D \models GW \sim T_{Nd,Ri}} \\
 \hline
 \frac{D \models GW \models \{D, GW\}^C \triangleleft T_{Nd,Ri}}{D \models \{D, GW\}^C \triangleleft T_{Nd,Ri}} \wedge D \models \text{sup}(GW) \\
 \hline
 \frac{D \models D \leftrightarrow GW}{D \models D \leftrightarrow GW}
 \end{array}
 \quad
 \begin{array}{c}
 \frac{GW \models GW \leftrightarrow D \wedge GW \models \{D\}^C \triangleleft T_{Nd,Ri} \wedge GW \sim T_{Nd,Ri} \wedge GW \models \#(T_{Nd,Ri})}{GW \models \{D, GW\}^C \triangleleft T_{Nd,Ri}} \\
 \hline
 \frac{GW \models GW \leftrightarrow D}{GW \models GW \leftrightarrow D}
 \end{array}$$

(e) (f)

Fig. 4. Security Proof of the Improved BAN Logic for $N_g, R_{i+1}, T_{Nd,Ri}$. (a) D believes that N_g is a shared secret between D and GW ; (b) GW believes that N_g is a shared secret between GW and D ; (c) GW believes that R_{i+1} is a shared secret between GW and D ; (d) D believes that R_{i+1} is a shared secret between D and GW ; (e) D believes that $T_{Nd,Ri}$ is a shared secret between D and GW ; (f) GW believes that $T_{Nd,Ri}$ is a shared secret between GW and D .

7) *Resistance to replay attacks*: The proposed scheme incorporates a timestamp mechanism, requiring verification of transmission delays before authentication. This prevents attackers from initiating replay attacks through message resending. Additionally, timestamps are included in V_0 and V_1 ; any attempt by an attacker to change the timestamp will result in authentication failure. Moreover, the secret values in V_0 and V_1 are updated after each authentication, effectively resisting replay attacks.

8) *Resistance to Denial-of-Service (DoS) attacks*: When attackers send excessive invalid information to disrupt communication between devices and gateways, the devices and gateways will first validate the transmission delays and then verify the values of V_0 or V_1 . Any failure to meet these criteria will result in a rejection of authentication.

9) *Forward and backward security*: In the proposed scheme, the session key negotiated is $SK = h(N_u \parallel R_{i+1} \parallel T_{Nd,Ri})$. Since N_u, R_{i+1} , and $T_{Nd,Ri}$ are updated after each authentication, even if an attacker acquires the current device's secret values and CRPs, they cannot trace past or future communications of the device, thus ensuring both forward and backward security.

V. PERFORMANCE ANALYSIS

A. Security Feature Analysis

Table IV compares the security features of the proposed scheme with those of existing solutions. In study [15], attackers can obtain CRPs through eavesdropping or spoofing, which makes the system vulnerable to machine learning modeling attacks. In contrast, the proposed scheme stores only pseudonymous identities on the device, preventing attackers from obtaining plaintext CRPs through physical attacks. Furthermore, the CRPs are protected by XOR or hash functions during the authentication process, which helps safeguard against machine learning modeling attacks. The study in [16] describes a system where authentication values are generated from secret values stored on the device or randomly generated by users. If this secret information is compromised, attackers could potentially impersonate legitimate devices or gateways. In the proposed scheme, however, attackers would need to access secret information such as N_u, N_g, R_{i+1} . These secrets are protected by Chebyshev polynomials or hash functions, making it difficult for attackers to access them and thus defending against spoofing and man-in-the-middle attacks.

B. Computational Overhead Analysis

Based on the execution times for various operations outlined in study [14], the following time parameters are considered: T_h for executing a hash function, T_{PUF} for executing a PUF, T_{che} for executing a Chebyshev polynomial, T_{Mul} for performing an elliptic curve point multiplication, $T_{FE,Gen}$ for generation with a fuzzy extractor, and $T_{FE,Rep}$ for recovery with a fuzzy extractor. The execution times for these operations are listed in Table V.

Table VI compares the computational overhead of the proposed scheme with those of other schemes in the literature (Fig. 5). As shown in the table, the schemes in studies [15] and [16] use resource-intensive elliptic curve point multiplication, resulting in the highest computational overheads of 3909.2284 μ s and 3549.8392 μ s, respectively. In contrast, the proposed scheme utilizes lightweight Chebyshev chaotic mappings, resulting in a total computational overhead of 1396.3521 μ s. This represents a reduction of 60.664% and 64.281%, respectively, compared to the computational overheads of the schemes proposed in the other references.

C. Communication Overhead Analysis

Before comparing the communication overhead, the lengths of the various variables are referenced from [14]. Table VII presents a comparison of the communication overhead between the proposed scheme and those in the literature. The table shows that the communication overhead of the proposed scheme is 1216 bits, which is lower than that of the schemes proposed in studies [15] and [16]. Therefore, the proposed scheme is well-suited for anonymous identity authentication and session key negotiation for resource-constrained terminal devices and gateways.

TABLE IV. COMPARISON OF SECURITY FEATURES

Security Attribute	Bai Haodong et al. [15]	Soni [16]	Proposed Scheme
Mutual Authentication	✓	✓	✓
Untraceability	✓	✓	✓
User Anonymity	✓	✓	✓
Forward/Backward Security	✓	✓	✓
DoS Attack	✓	✓	✓
Replay Attack	✓	✓	✓
Machine Learning Modeling Attack	×	✓	✓
Spoofing Attack	✓	×	✓
Man-in-the-Middle Attack	✓	×	✓
Mutual Authentication	✓	✓	✓

TABLE V. EXECUTION TIMES FOR VARIOUS OPERATIONS

Operation	Operation execution time	
	Device Side	Gateway Side
T_h	2.7324 μ s	0.1315 μ s
T_{PUF}	6.7 μ s	/
T_{che}	91.2600 μ s	10.6604 μ s
T_{Mul}	426.4887 μ s	103.8660 μ s
$T_{FE.Gen}$	278.0889 μ s	74.7562 μ s
$T_{FE.Rep}$	696.1048 μ s	157.4092 μ s

TABLE VI. COMPARISON OF COMPUTATIONAL OVERHEAD

Scheme	Device Side	Gateway Side	Total Time
Bai et al. [15]	$5T_h + T_{FE.Gen} + T_{FE.Rep} + 5T_{Mul} + 2T_{PUF}$ $\approx 3133.8492\mu s$	$4T_h + 4T_{Mul}$ $\approx 415.99\mu s$	3549.8392 μs
Soni et al. [16]	$11T_h + T_{FE.Rep} + 6T_{Mul}$ $\approx 3285.2434\mu s$	$6T_h + 6T_{Mul}$ $\approx 623.985\mu s$	3909.2284 μs
Proposed Scheme	$8T_h + 2T_{Che} + 2T_{PUF} + T_{FE.Gen}$ $\approx 495.8681\mu s$	$8T_h + 2T_{Che} + T_{FE.Rep}$ $\approx 900.484\mu s$	1396.3521 μs

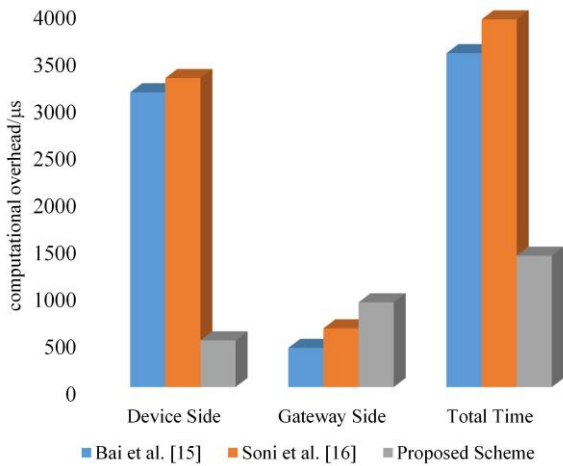


Fig. 5. Comparison of computation overhead.

TABLE VII. COMPARISON OF COMMUNICATION OVERHEAD

Scheme	Number of messages	Communication cost
Bai et al. [15]	3	1472bit
Soni et al. [16]	2	2304bit
Proposed Scheme	3	1216bit

The experimental results confirm that the proposed scheme offers significant improvements in both security and efficiency compared to existing solutions. As shown in Table IV, the scheme effectively mitigates the vulnerabilities of previous methods, such as machine learning attacks and impersonation, by storing only pseudonymous identities and using XOR/hash functions to protect CRPs. This is a clear advantage over reference [15], where CRPs can be intercepted, and reference [16], where compromised secrets may lead to spoofing.

In terms of computational overhead, our scheme, utilizing Chebyshev chaotic mappings, significantly reduces processing time by approximately 60-64% compared to the elliptic curve-based methods in studies [15] and [16]. This makes it more suitable for resource-constrained IoT devices. Furthermore, as seen in Table VII, the communication overhead of our scheme is lower than that of existing solutions, making it ideal for devices with limited bandwidth.

Overall, our scheme provides a balanced approach, offering robust security and efficiency, which is essential for resource-constrained IoT environments.

VI. CONCLUSION

This paper presents a lightweight identity authentication scheme designed for resource-constrained IoT devices, which has been verified for security using an improved BAN logic. The results indicate that the proposed scheme is capable of resisting attacks such as physical attacks, machine learning modeling attacks, replay attacks, and man-in-the-middle attacks. Compared to existing solutions, the computational overhead of the proposed scheme is only 1396.3521 μs , and the communication overhead is only 1216 bits, making it suitable for efficient and secure authentication in IoT environments. Future work will focus on further optimizing the performance of the scheme, reducing system overhead, and conducting more extensive testing and validation in complex application scenarios to enhance the overall security and reliability of the scheme.

REFERENCES

- [1] W. Che, F. Saqib, and J. P. Plusquellic, "PUF-based authentication," in *2015 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Austin, TX, USA, November 2015, pp. 337–344, doi: 10.1109/ICCAD.2015.7372589.
- [2] A. Braeken, "PUF based authentication protocol for IoT," *Symmetry*, vol. 10, no. 8, p. 352, 2018, doi: 10.3390/sym10080352.
- [3] J. Zou, B. Zhao, X. Li, Y. Liu, and J. Li, "tPUF-based secure access solution for IoT devices," *Computer Engineering and Applications*, vol. 57, no. 02, pp. 119–126, 2021.
- [4] J. Delvaux, R. Peeters, D. Gu, and I. Verbauwhede, "A survey on lightweight entity authentication with strong PUFs," *ACM Computing Surveys (CSUR)*, vol. 48, no. 2, pp. 1–42, 2015, doi: 10.1145/2818186.
- [5] Z. He, H. Li, M. Wan, and T. Wu, "A two-party authentication and session key exchange protocol based on PUF," *Computer Engineering and Applications*, vol. 54, no. 18, pp. 17–21, 2018.
- [6] Y. Xia, R. Qi, and S. Ji, "Research on lightweight key exchange protocol based on PUFs in industrial Internet of Things," *Computer Applications and Software*, vol. 39, no. 03, pp. 316–321, 2022.
- [7] Y. Cho, J. Oh, D. Kwon, S. Son, J. Lee, and Y. Park, "A secure and anonymous user authentication scheme for IoT-enabled smart home environments using PUF," *IEEE Access*, vol. 10, pp. 101330–101346, 2022, doi: 10.1109/ACCESS.2022.3208347.
- [8] A. K. Maurya and V. N. Sastry, "Fuzzy extractor and elliptic curve based efficient user authentication protocol for wireless sensor networks and Internet of Things," *Information*, vol. 8, no. 4, p. 136, 2017, doi: 10.3390/info8040136.
- [9] N. Singh and A. K. Das, "TFAS: two factor authentication scheme for blockchain enabled IoT using PUF and fuzzy extractor," *The Journal of Supercomputing*, vol. 80, no. 1, pp. 865–914, 2024, doi: 10.1007/s11227-023-05507-6.
- [10] Y. Gao, S. F. Al-Sarawi, and D. Abbott, "Physical unclonable functions," *Nature Electronics*, vol. 3, no. 2, pp. 81–91, 2020, doi: 10.1038/s41928-020-0372-5.

- [11] D. Dharminder and P. Gupta, "Security analysis and application of Chebyshev Chaotic map in the authentication protocols," *International Journal of Computers and Applications*, vol. 43, no. 10, pp. 1095–1103, 2021, doi: 10.1080/1206212X.2019.1682238.
- [12] C. Herder, L. Ren, M. Van Dijk, M. D. Yu, and S. Devadas, "Trapdoor computational fuzzy extractors and stateless cryptographically-secure physical unclonable functions," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 1, pp. 65–82, 2016, doi: 10.1109/TDSC.2016.2536609.
- [13] H. Ma, C. Wang, G. Xu, Q. Cao, G. Xu, and L. Duan, "Anonymous authentication protocol based on physical unclonable function and elliptic curve cryptography for smart grid," *IEEE Systems Journal*, vol. 17, no. 4, pp. 6425–6436, 2023, doi: 10.1109/JSYST.2023.3289492.
- [14] X. Jin, N. Lin, Z. Li, W. Jiang, Y. Jia, and Q. Li, "A lightweight authentication scheme for Power IoT based on PUF and Chebyshev Chaotic Map," *IEEE Access*, vol. 12, pp. 83692–83706, 2024, doi: 10.1109/ACCESS.2024.3413853.
- [15] H. Bai and X. Jia, "A smart grid equipment authentication scheme based on physically unclonable functions," *Journal of South-Central Minzu University (Natural Science Edition)*, vol. 42, no. 3, pp. 382–386, 2023, doi: 10.20056/j.cnki.ZNMDZK.20230313.
- [16] P. Soni, J. Pradhan, A. K. Pal, and S. H. Islam, "Cybersecurity Attack-Resilience Authentication Mechanism for Intelligent Healthcare System," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 1, pp. 830–840, 2022, doi: 10.1109/TII.2022.3179429.

Comparative Analysis of Feature Selection Based on Metaheuristic Methods for Human Heart Sounds Classification Using PCG Signal

Motaz Farooq A Ben Hamza, Nilam Nur Amir Sjarif

Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur 54100, Malaysia

Abstract—Cardiovascular disease is a critical threat to human health, as most death cases are due to heart disease. Although several doctors employ stethoscopes to auscultate heart sounds to detect abnormalities, the accuracy of the approach is considerably dependent upon the experience and skills of the physician. Consequently, optimal methods are required to analyse and classify heart sounds with Phonocardiogram (PCG) signal-based machine learning methods. The current study formulated a binary classification model by subjecting PCG signals to hyper-filtering with low-pass and cosine filters. Subsequently, numerous features are extracted with the Wavelet Scattering Transform (WST) method. During the feature selection stage, several metaheuristic methods, including Harris Hawks Optimisation (HHO), Dragonfly Algorithm (DA), Grey Wolf Optimiser (GWO), Salp Swarm Algorithm (SSA), and Whale Optimisation Algorithm (WOA), are employed to compare the attributes separately and determine the ideal characteristics for improved classification accuracy. Finally, the selected features were applied as input for the Bidirectional Long Short-Term Memory (Bi-LSTM) algorithm, simplifying the classification process for distinguishing normal and abnormal heart sounds. The present study assessed three PCG datasets: PhysioNet 2016, Yaseen Khan 2018, and PhysioNet 2022, documenting 94.85%, 100%, and 66.87% accuracy rates with 127-SSA, 168-HHO, and 163-HHO, respectively. Based on the results of the PhysioNet 2016 and 2022 datasets, the proposed method with hyperparameters demonstrated superior performance to those with default parameters in categorising normal and abnormal heart sounds appropriately.

Keywords—Cardiovascular Diseases (CVDs); Phonocardiogram (PCG) signal processing; Wavelet Scattering Transform (WST); Metaheuristic Methods; Harris Hawks Optimisation (HHO); Dragonfly Algorithm (DA); Grey Wolf Optimiser (GWO); Salp Swarm Algorithm (SSA); Whale Optimisation Algorithm (WOA); Bidirectional Long Short-Term Memory (Bi-LSTM)

I. INTRODUCTION

Diagnosis with heart sounds has drawn much attention in the biomedical research area because of the crucial nature of the heart and the high mortality rates associated with cardiovascular diseases. Heart sounds are the mechanical activities of the heart, which vary according to pathological conditions affecting it [1]. Currently, the two major expensive technologies for detecting heart diseases are echocardiography and cardiac Magnetic Resonance Imaging (MRI) [2]. Auscultation is a basic diagnostic technique commonly used to assess heart function and quality [3]. It involves the listening of

heart sounds directly by placing a stethoscope at various points in the chest. Nonetheless, it is subjective and highly dependent on the acuteness of hearing and experience of the physician. Therefore, there is a need to develop a system that can objectively process heart sounds, enabling faster and more accurate diagnoses.

There are two major sounds produced by a normally functioning heart. The first heart sound (S1) and the second heart sound (S2). Closing of the mitral and tricuspid valves causes the first sound, S1, while the second sound, S2, is a result of the aortic and pulmonary valves closing at the end of the systolic phase. Sound identification can be one essential part of the heart disease diagnosis since the nature and other characteristics of the heart murmurs can vary from one to another as per concise disease conditions. In most instances, differences in heart sound patterns between healthy and unhealthy states are distinguished on the basis of changes in intensity, timing, location, etc., among other factors [4].

Several researchers have worked on the study and categorisation of Phonocardiogram (PCG) signals using various machine learning techniques [5]; particularly, two main areas have been focused on: Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) [6], [7]. However, these studies often require extraction of a wide array of features, which can complicate the training phase when dealing with heart sound signals. Also, these studies involve issues of existing noise in PCG signals and imbalanced datasets, leading to the extraction of unnecessary features [8]. Therefore, the contribution of this study is as follows:

- To introduce low-pass and cosine hyper-filters during preprocessing to reduce background noise from PCG signals.
- To extract various features from the PCG signals with Wavelet Scattering Transform (WST).
- To develop metaheuristic methods with hyperparameters to select optimal features for refined features from an initial set of features, therefore reducing computational complexity and improving classification performance by diminishing the search space.
- To execute high-performance heart sound classification with the Bidirectional Long Short-Term Memory (Bi-LSTM) classifier.

Section II of this study discusses related articles on heart sound classification. Details on the WST characteristics, the proposed methodology with several metaheuristic methods, and the Bi-LSTM algorithm are included in Section III. In Section IV, the results, including the dataset employed for comparison with existing literature, are discussed. The conclusions of this study are mentioned in Section V. Recommendations for future research are also included.

II. RELEVANT WORK

In the biomedical field, several techniques have been proposed to classify normal and abnormal heart sounds, emphasising Phonocardiogram (PCG) signal processing. Generally, the signals are preprocessed to filter out extraneous high-frequency noises. Subsequently, characteristic features are extracted and utilised as input data to construct or develop mathematical models intended for diagnosing heart conditions [8], [9].

Preprocessing is fundamental in PCG categorisation. The step involves three primary tasks: baseline wander removal, background noise reduction, and normalisation. Baseline wander is a common issue in low-frequency PCG recordings. This issue shifts the baseline reference level, which hinders accurate signal property extraction [10], [11]. Background noise is prevalent in high-frequency PCG recordings [12] and can degrade signal quality. Normalisation scales sample values in the dataset to a standard range (typically 0-1 or -1 to 1), ensuring that amplitude discrepancies do not distort feature extraction and classification processes [13]. This standardisation prevents bias in the model due to variations in recording amplitudes [14].

Potes et al. [15] classified heart sound recordings from the PhysioNet Challenge 2016 into normal and abnormal classes with an ensemble classifier. The report applied the Butterworth bandpass filter and extracted 124 features from PCG signals with Mel Frequency Cepstral Coefficients (MFCCs). The approach achieved an 86.02% accuracy rate without employing any feature reduction technique. The study also achieved first place in the PhysioNet Challenge 2016.

In 2017, Kay and Agarwal [16] applied a hidden semi-Markov model for segmentation before extracting temporal and spectral features utilising continuous WST and MFCCs from the PhysioNet Challenge 2016 dataset. Subsequently, Principal Component Analysis (PCA) was employed to reduce data dimensionality. Finally, the information was fed into an Artificial Neural Network (ANN), yielding an 85.2% accuracy. In another study, Bao et al. revealed that the Bi-LSTM classifier performed better than CNN on an identical dataset. Accuracy, sensitivity, and specificity rates of 92.64%, 84.77%, and 95.14%, respectively, were noted [17].

Li et al. [18] utilised the Twin SVM (TWSVM) classifier to compare the performance of multi-dimensional scaling and PCA for feature selection on the PhysioNet Challenge 2016 dataset. The Multi-Dimensional Scaling (MDS) outperformed PCA with 98.58%, 98.58%, 98.57%, and 99% in respective accuracy, sensitivity, specificity, and F1 score. Meanwhile, Alshamma et al. (2019) applied a high-pass filter on each PCG signal before employing a normalisation method based on zero

mean and standard deviation. Subsequently, different K-Nearest Neighbours (KNN) and Support Vector Machines (SVM) classifiers were compared with the PhysioNet Challenge 2016 dataset. The fine-KNN classifier documented interesting results with a 93.5% accuracy [11].

In 2020, Singh and Majumder obtained short and non-segmented signals from the PhysioNet Challenge 2016 dataset. High noise frequencies were filtered with the Butterworth low-pass filter before filtering 27 features with MFCCs. Post-training with an ensemble classifier, the features recorded 92.47% accuracy, 94.08% sensitivity, and 91.95% specificity [19].

An energy envelopogram was employed to extract PCG signals from the PhysioNet Challenge 2016 dataset (Potdar, 2021) [20]. The traits were procured with Discrete Wavelet Transform (DWT), selected utilising PCA, and fed into a fine-tuned Bayesian-optimised SVM algorithm. On the other hand, Milani et al. (2021) applied Springer's segmentation algorithm to preprocess the PhysioNet Challenge 2016 dataset, extracting the attributes with MFCCs. Subsequently, Linear Discriminant Analysis (LDA) was applied to further diminish dimensionality. The ANN-based classification recorded a 93.33% accuracy [12].

In 2022, Zhang et al. extracted numerous features in two-dimensional convolution from the PhysioNet Challenge 2016 dataset. Subsequently, the Particle Swarm Optimisation (PSO) was subjected to feature selection before being trained with the convolutional Bi-LSTM network. The report documented accuracy, sensitivity, and specificity of 91.93%, 91.58%, and 92.27%, respectively [14].

A recent study [21] employed a VGG16 model during the spectrogram trait extraction of 1,330 PCG signals from the PhysioNet Challenge 2016 dataset, achieving an 88.84% accuracy. Nevertheless, the authors concluded that the accuracy level was insufficient for reliable patient diagnosis, requiring a more substantial and balanced dataset for performance enhancement. Although the original dataset was extensive, its imbalance might result in biased results. Consequently, implementing more sophisticated balancing techniques rather than simple down-sampling could improve the outcomes. Moreover, the report did not remove noise, a critical aspect of real-world PCG analysis.

Son and Kwon (2018) [22] constructed the Yaseen Khan 2018 dataset, consisting of 1,000 PCG signals. The report classified heart sounds via three algorithms: SVM, KNN, and Deep Neural Network (DNN). The MFCCs and DWT were also applied during feature extraction. The SVM exhibited the best performance, recording accuracy, sensitivity, and specificity of 97.9%, 98.2%, and 99.4%, respectively.

Flores-Alonso et al. applied a smoothing filter, normalised, and segmented the noisy Yaseen Khan 2018 dataset. The report integrated features for classification utilising CNN and Multi-Layer Perceptron (MLP), including MFCCs, DWT, and Continuous Wavelet Transform (CWT). The report recorded an accuracy of 99.8% [23]. In another study, [24] utilised 957 PCG signals from an identical dataset. Nonetheless, the study

excluded samples under 2 s. Multiple features were extracted with MFCCs and DWT. The data obtained was then employed to train classification models with five machine learning algorithms: Random Forests (RF), KNN, SVM, Naive Bayes (NB), and MLP. A remarkably high accuracy of 99.89% in diagnosing normal and abnormal heart sounds with the RF algorithm was procured.

McDonald et al. [25] focused on segmentation techniques in PCG signals, separating data into S1, S2, systole, and diastole. The study also applied hidden semi-Markov models to the PhysioNet Challenge 2022 dataset. Furthermore, a classification model was developed utilising a Bi-GRU algorithm. The 60.2% accuracy documented led the study to win the competition. Meanwhile, Singh et al. [26] applied Mel-spectrograms to extract features from the same dataset. The extracted features were classified with the U-Net method. The report achieved 56.8%, 54.77%, 59.29%, and 58.33% accuracy, sensitivity, specificity, and F1 score, respectively.

During the past five years, most research employed the Butterworth filter for background noise reduction [19], [27] and segmentation techniques [12], [20] for preprocessing. Nonetheless, segmenting non-linear PCG datasets remains challenging as it might lead to valuable information loss through abstract feature extractions [19], [28]. Moreover, a Wavelet Transform (WT) method has been broadly applied to extract PCG signal features. Nevertheless, the technique often extracts redundant traits, further degrading model performance [29], resulting in numerous research studies focusing on traditional feature extraction selection methods, such as LDA and PCA [12], [16]. The recent review by study [8] suggested using the WT method with metaheuristic methods for feature analysis from PCG signals, which allows the LSTM algorithm to generate the optimal model that can achieve a high classification accuracy.

Based on the literature reviewed in this section, heart sound classification was performed through various approaches. An increasing interest in feature selection and optimisation techniques to improve classification accuracy has also been observed. Although substantial strides have been achieved, handling and minimising non-linear PCG dataset feature redundancies are still challenging.

Consequently, advanced preprocessing and feature selection methodologies necessitate further exploration to improve the accuracy and reliability of heart sound classification systems. Preprocessing heart sounds to eliminate noise is a promising avenue for future studies. Furthermore, combining multiple extracted features into a single representation and training the model on the enriched data might yield superior results.

III. PROPOSED METHODOLOGY

The current study reduced PCG signal noise in the preprocessing with Butterworth and cosine filters. Subsequently, the signal attributes were automatically extracted utilising WST. The characteristics were optimally selected through several metaheuristic methods, including Harris Hawks Optimisation (HHO), Dragonfly Algorithm (DA), Grey Wolf Optimiser (GWO), Salp Swarm Algorithm

(SSA), and Whale Optimisation Algorithm (WOA), contributing to feature selection.

The results revealed that the proposed method could achieve excellent classification accuracy with a few feature sets. The suggested model also involved three processes: preprocessing, feature extraction and selection, and classification, as shown in Fig. 1. Each procedure is detailed in subsections A to D.

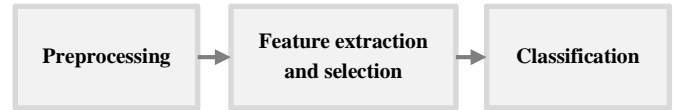


Fig. 1. The general process diagram of the current study.

A. Preprocessing of the PCG Signal

The preprocessing phase is vital in improving Phonocardiogram (PCG) signal quality before extracting and classifying any features. The following lists the steps necessary during the procedure.

1) *Baseline wander removal*: Baseline wander refers to low-frequency noise typically arising from respiratory movements or patient motion. The noises might distort PCG signals. Consequently, a high-pass Butterworth filter with a 0.5 Hz cut-off frequency was employed to overcome the matter. The filter effectively attenuated frequencies associated with baseline drift. Moreover, heart sound signals predominantly detected between 20 and 150 Hz were preserved. The findings indicated that removing low-frequency noise is critical to avoid interference during the feature extraction phase.

2) *Background noise removal*: The PCG signals are susceptible to various noise sources, including ambient, lung, and muscle contraction sounds. Accordingly, this study applied a two-stage filtering process to enhance the Signal-to-Noise Ratio (SNR). Firstly, a Butterworth low-pass filter at a 150 Hz cut-off frequency was employed to eliminate high-frequency noises. Subsequently, the Adaptive Noise Cancellation (ANC) utilised allowed noise component estimations within the PCG signal by referencing a separate noise signal and subtracting them from the PCG data. The two-phase approach effectively isolated relevant heart sound components from the background noise, facilitating accurate feature extraction and classification.

3) *Data normalisation*: The PCG signal amplitudes vary significantly depending on the recording environment, device, and patient physiology. Consequently, this study employed data normalisation to address the issue. The procedure scaled signal amplitudes to a common range, typically between 0 and 1 or -1 and 1. The min-max normalisation in the present study was calculated according to following the equation.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Where x' refers to normalised PCG signals, x denotes the original PCG signal, $max(x)$ represents the maximum PCG value, and $min(x)$ is the minimum PCG value.

B. Feature Extraction

Feature extraction transforms preprocessed PCG signals into classification-appropriate representations. Various methods, including MFCCs, DWT, and WST, have been employed in previous studies.

1) *Mel-frequency Cepstral Coefficient (MFCC)*: The MFCCs capture spectral PCG signal properties by emphasising perceptually relevant frequency bands. The process involves the following steps [19]:

a) *Framing* - Dividing the filtered PCG signal into overlapping 5-s frames.

b) *Spectrum estimation* - Calculating the amplitude spectrum of each frame.

c) *Windowing* - Multiplying each frame by a Hamming window to reduce spectral leakage through signal end section attenuation to zero.

d) *Fourier transform* - Applying the Fast Fourier Transform (FFT) to convert the time-domain signal into the frequency domain.

e) *Mel Filterbank* - Passing the spectrum through a series of triangular bandpass filters spaced according to the Mel scale.

f) *Logarithm* - Computing the Mel Filterbank output algorithm to mimic the human auditory system perception.

g) *Discrete Cosine Transform (DCT)* - Applying the DCT to the log-filterbank outputs to decorrelate the coefficients. Resultantly, 27 MFCCs in the time-frequency domain for each PCG signal were obtained. Fig. 2 demonstrates the MFCCs extraction phases, which were determined based on following the equation.



Fig. 2. The MFCC method for feature extraction.

$$MFCC_n = \sum_{k=1}^K \log(S_k) \cdot \cos\left(\frac{\pi n(k-0.5)}{K}\right) \quad (2)$$

Where S_k is the log power at each Mel frequency k , K refers to the total number of Mel frequency bands, and N denotes the number of MFCCs to retain.

2) *Discrete Wavelet Transform (DWT)*: The current study utilised DWT to analyse the non-stationary nature of the PCG signals by iteratively decomposing them into time-frequency domains through wavelet filters [30]. The technique fuses low- and high-pass filters, providing approximation and detail coefficients from the PCG signal [31].

The DWT enables high-frequency energy in systolic and diastolic heart sound alteration representations [32]. The approach captures high-frequency components (details) and low-frequency components (approximations), crucial for detecting anomalies in heart sounds. Meanwhile, following the

equation determines wavelet coefficients by projecting $x(t)$ signals onto scaled and shifted wavelet $\psi(t)$ function versions.

$$DWT(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^*\left(\frac{t-b}{a}\right) dt \quad (3)$$

Where $DWT(a, b)$ represents the wavelet coefficient at scale a and position b , $x(t)$ is the input signal, ψ^* denotes the complex conjugate of the wavelet function ψ , a represents the scaling parameter, and b is the translation parameter.

In this study, the approximation coefficient ($c, A-n$) and the detail coefficient ($c, D-n$) of each level, n , was established with following the equation. The DWT applied in the present study had 15 decomposition levels for each cycle. The signal energy, entropy, and standard deviation were procured as 16-time features, while the waveform length and wavelet variance estimates were obtained as 16-frequency features.

$$A = cA_n \sum_{k=0}^n cD_n \quad (4)$$

Where A represents the wavelet coefficient values, while n indicates the approximate level number.

Utilising the DWT method, 80 features were extracted from the 15 approximation coefficients for each cycle per the guidelines outlined by [31]. The coefficients were calculated based on the Shannon energy. Fig. 3 represents a generalised flow of the wavelet decomposition from the PCG signal.

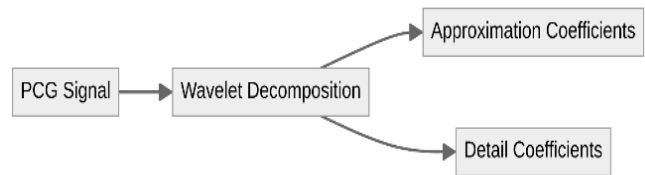


Fig. 3. The wavelet decomposition from the PCG signal.

3) *Wavelet Scattering Transform (WST)*: The WST is an advanced technique that captures stable, hierarchical time-frequency features from PCG signals. The method involves the following steps [18][33]:

- Convolution of PCG signals with a wavelet family of varying scales.
- The application of non-linear modulus operators to the convolved signals to obtain scattering coefficients.
- The coefficients are averaged over time to produce the final scattering form, retaining translation-invariant and stable attributes.

In WST, the $x(t)$ signal is convolved with a wavelet function, ψ_1 , before being further convolved with additional wavelet functions, ψ_2, \dots, ψ_J . Eq. (5) is the mathematical representation of WST. The phase enables multi-scale and multi-resolution feature extractions as illustrated in Fig. 4.

$$S_J(x(t)) = |x * \psi_1| * \psi_2 * \dots * \psi_J \quad (5)$$

Where $S_J(x(t))$ denotes the wavelet scattering coefficients at level J , $x(t)$ represents the input signal, and $\psi_1, \psi_2, \dots, \psi_J$ are the wavelet functions at different scales.



Fig. 4. The wavelet scattering transforms on PCG signal.

C. Feature Selection

Feature selection aims to eliminate duplicate or non-informative features, which reduces the excess amount of input features, leading to achieving an optimal model. Hence, feature selection is critical in ensuring that only the most relevant information is passed to the classifier [8]. In the current study, five metaheuristic algorithms for feature selection were employed: HHO, DA, GWO, SSA, and WOA.

1) *Harris Hawks Optimisation (HHO)*: The HHO is a swarm intelligence algorithm published by [34] in 2019. The model was inspired by nature, mimicking the cooperative hunting strategy of Harris hawks. Balancing exploration and exploitation, the algorithm simulates different prey-hunting phases, such as surprise pounce and soft besiege.

The current study employed HHO to determine the most suitable optimal features. The approach is characterised by rapid large solution space explorations with substantial accuracy and rapid convergence, maintaining exploration and exploitation equilibrium. Post-feature extraction, the initial parameters were applied to establish a random position for HHO within the search bounds.

The HHO algorithm determines the best position for prey. The exploration phase is denoted by prey energy values over 1 ($E > 1$), where HHO is still in the searching mode. Meanwhile, a slight energy drop below 1 ($E < 1$) denotes the model entering the exploitation stage. During the soft siege ($E \geq 0.5$), HHO continually updates the position of the prey demonstrating energy to escape. Nevertheless, HHO dives around the target in a hard siege once the energy falls below 0.5 with a significant escape probability.

Features within the optimal range and low error rate are selected at each stage. During the final cycle, the selected PCG attributes are optimised and employed as input for the classification model. The HHO process [35] is illustrated in Fig. 5.

The technique is particularly effective for selecting features from the WST as it efficiently handles the hierarchical and multi-scale data. Feature selection was calculated according to following the equation.

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1 \cdot |X_{rand}(t) - 2r_2 \cdot X(t)| & E \geq 0.5 \\ X_{prey}(t) - X(t) - r_3 \cdot |X_{prey}(t) - X(t)| & E < 0.5 \end{cases} \quad (6)$$

Where $X(t+1)$ refers to the next position of the hawk, X_{rand} and X_{prey} represent random and prey positions, while r_1 , r_2 , and r_3 are random numbers.

2) *Dragonfly Algorithm (DA)*: Introduced in 2016 by [36], DA is a metaheuristic optimisation algorithm based on the static and dynamic swarming behaviours of dragonflies. The model utilises five key factors: separation, alignment, cohesion, attraction to food, and distraction from enemies.

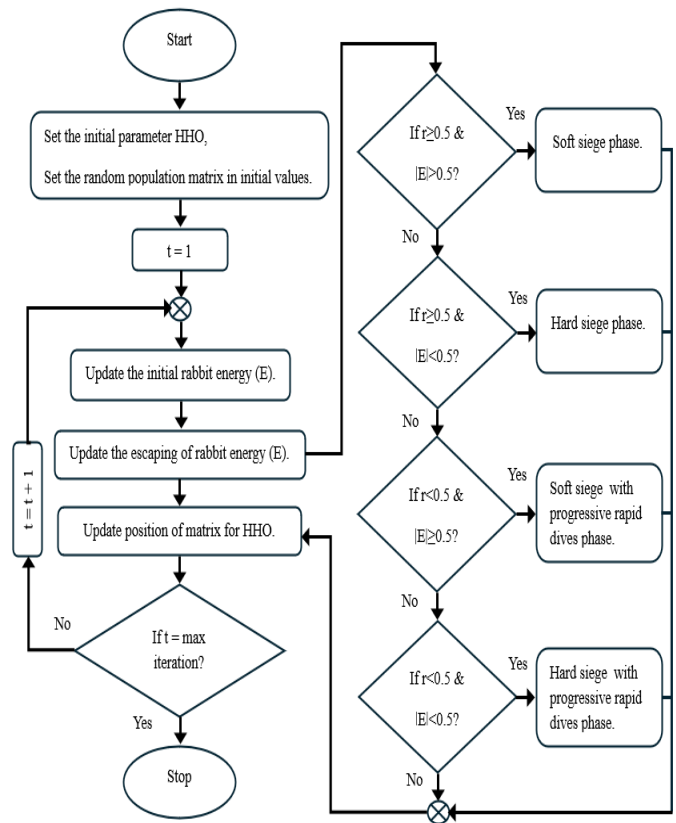


Fig. 5. The HHO working principle.

The DA defines a neighbourhood of radius r according to how it works. Neighbourhood sizes are scaled up based on a linear relation with the iteration counter during exploration to exploitation transitions. Consequently, static swarms are transformed into dynamic swarming. In the final optimisation stage, each dragonfly solution is joined to form an active unified swarm, which converges towards the best global solution at the end of the convergence [37], [38], [39]. The feature selection for the technique is represented by following the equation. The DA was also applied in this study during the feature selection stage (see Fig. 6).

$$\Delta X = \sum_{i=1}^N (S_i + A_i + C_i + F_i + E_i) \quad (7)$$

Where S is separation, A denotes alignment, C represents cohesion, F is food attraction, and E denotes enemy distraction. The working principle of DA.

3) *Gray Wolf Optimiser (GWO)*: Mirjalili et al. [40] proposed GWO, a swarm intelligence-based metaheuristic algorithm, in 2014. The algorithm was based on the grey wolf leadership hierarchy and hunting strategy. Based on the grey wolves hierarchical system [35], the alpha (α), or the most dominant wolf in the pack, leads the other wolves during food hunting and finding. During the absence of the alpha wolf, the beta (β) becomes the pack leader. In the hierarchy, the power levels of the delta (δ) and omega (ω) groups are significantly less apparent than their nearest rival (see Fig. 7).

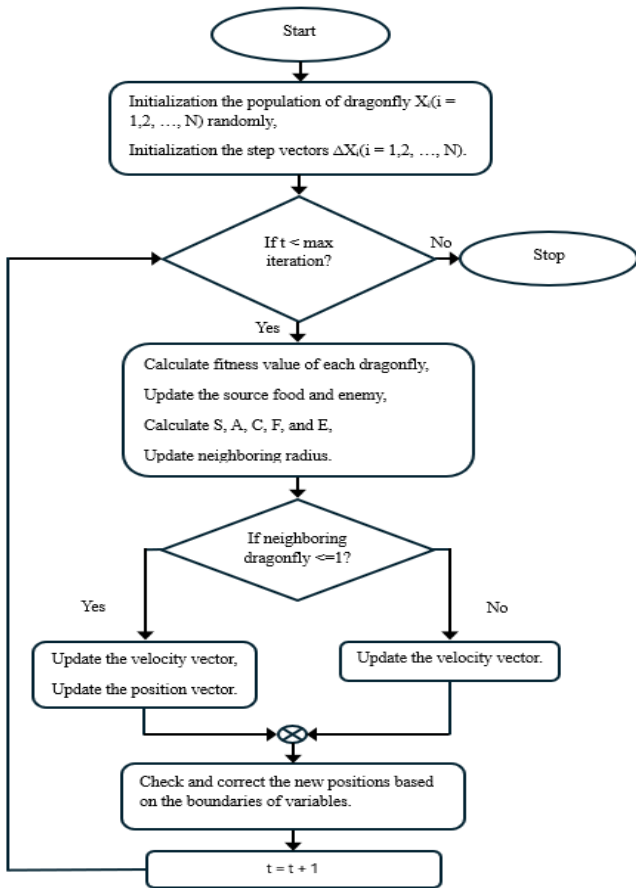


Fig. 6. The working principle of DA.

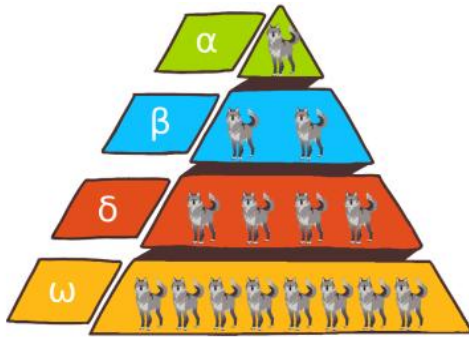


Fig. 7. The GWO hierarchical levels.

The mechanism of the GWO algorithm in the sociological agenda is complex social intelligence. Grey wolves exhibit remarkable hunting strategies by chasing, encircling, and attacking their prey [41], [42], [43]. Successful pursuits lead them to the optimal solution through successive phases with distinctive efficiencies, encouraging others to adopt similar cooperative actions [35]. Fig. 8 outlines the GWO feature selection process, while the following equation was employed to determine feature selection.

$$\vec{X}(t + 1) = \frac{\vec{X}_\alpha + \vec{X}_\beta + \vec{X}_\delta}{3} \quad (8)$$

Where $X(t + 1)$ refers to the next position of the wolf, and $\vec{X}_\alpha, \vec{X}_\beta, \vec{X}_\delta$ are the positions of the top three wolves.

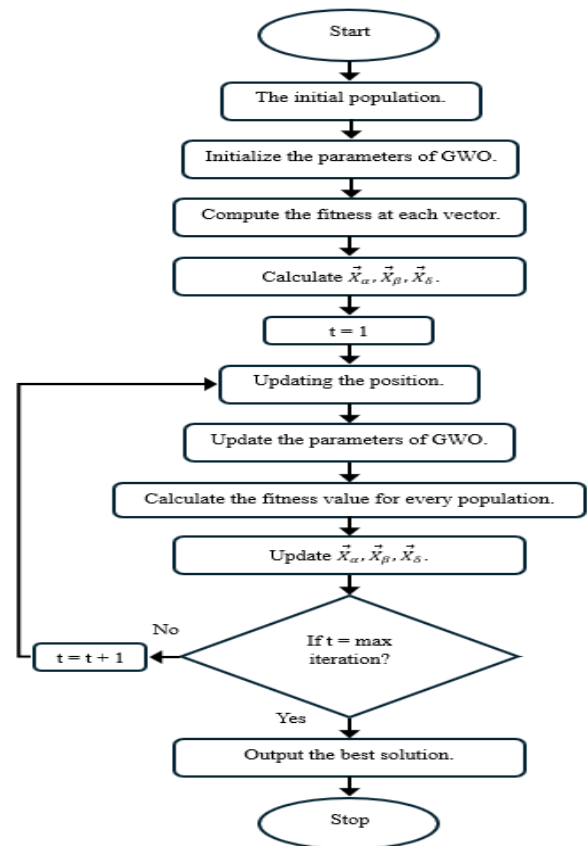


Fig. 8. The GWO working principle.

4) *Salp Swarm Algorithm (SSA)*: In 2017, SSA, a novel nature-inspired optimiser, was proposed by Mirjalili et al. [44]. The model mimics the collective behaviour of salps, which are marine wildlife. The SSA is versatile, efficient, straightforward, and applicable to parallel and serial modes. Furthermore, the algorithm has a single adaptively decreasing parameter, contributing to optimal diversification and intensification tendencies balancing.

Salps move dynamically and update their positions through mutual interactions to avoid being trapped in local optima. The SSA behaviour is recognised by the salp chain algorithm, searching and selecting optimal food sources effectively. The swarm aims to identify and locate a specific food source within the search space.

The salps in the SSA approach are categorised as either "leaders" or "followers" based on their position in the chain. Follower salps rely on the actions of their leader for guidance. The flowchart of the SSA processes is demonstrated in Fig. 9 [45]. In this study, feature selection utilising SSA was established according to following the equation.

$$X_{i,j}(t + 1) = \begin{cases} X_j(t) + c_1 \cdot (U_j - L_j) + L_j & i = 1 \\ \frac{X_{i,j}(t) + X_{i-1,j}(t)}{2} & i > 1 \end{cases} \quad (9)$$

Where X_j is the food source in the j dimension, U_j and L_j refer to the upper and lower bounds of the j dimension, respectively, and c_1 denotes a random number.

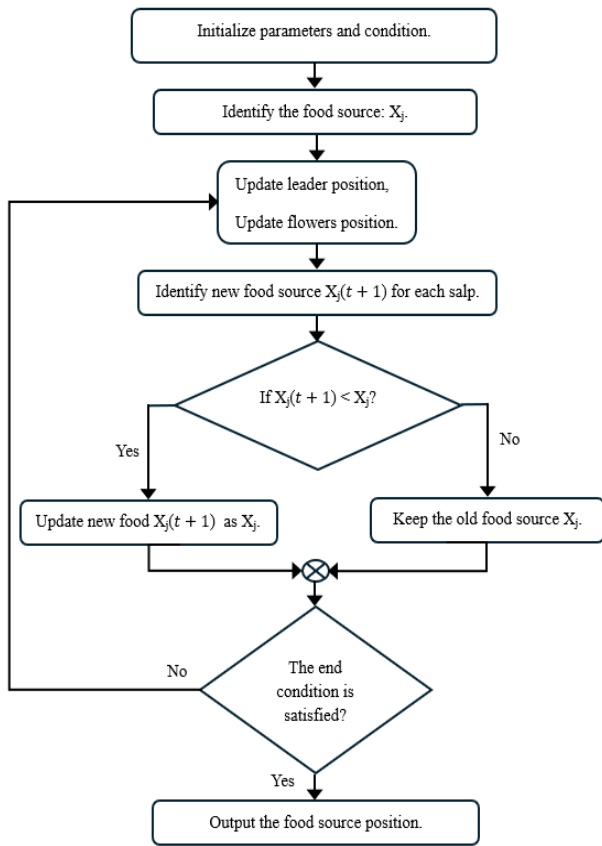


Fig. 9. The working principle of SSA.

5) *Whale Optimisation Algorithm (WOA)*: The WOA algorithm was developed by Mirjalili and Lewis [46] in 2016. The model was the first metaheuristic approach applicable to comprehensive optimisations. The WOA algorithm involves two primary phases based on the bubble-net hunting strategy of humpback whales.

Firstly, humpback whales hunt and encircle their prey. Similarly, the WOA algorithm assumes that the best solution is unknown, thus identifying the optimal candidate solution as the target prey or close to it. Subsequently, additional search agents adjust their positions to match the best search agent, shrinking the search space effectively.

Fig. 10 depicts the second phase of the nut strategy, the bubble-net attacking stage. During the step, the whales reduce the predator chain around the prey and move in a spiral attacking pattern. Meanwhile, random humpback whales calculate new positions instead of relying solely on the globally best-known position in the update phase [47], [48]. Eq. (10) was employed to determine the feature selection in this study, and the selection process is illustrated in Fig. 11.

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - A \cdot D & p < 0.5 \\ D' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) & p \geq 0.5 \end{cases} \quad (10)$$

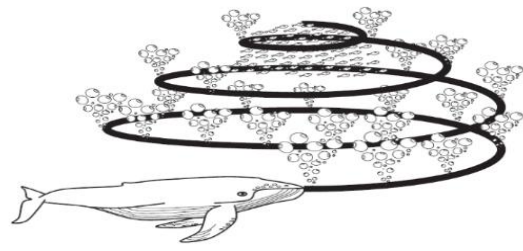


Fig. 10. The WOA behaviour.

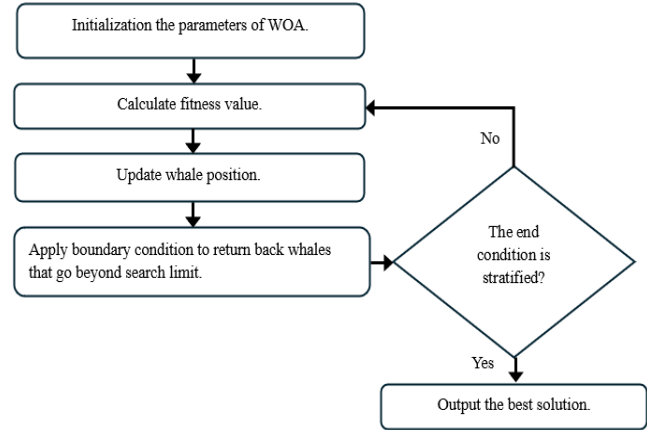


Fig. 11. The working principle of WOA.

Where $X(t+1)$ is the updated position of the whale, $X^*(t)$ represents the position of the best solution (prey), A and D are coefficient vectors calculated as $A = 2a \cdot r - a$ and $D = |C \cdot X^*(t) - X(t)|$, $D' = |X^*(t) - X(t)|$ denotes the distance to the prey, p is a random number between 0 and 1 that determines the exploitation or exploration phase, b is attributed to a constant defining the shape of the logarithmic spiral, and l represents a random number between -1 and 1 .

D. Classification

During the classification model design, the selected features were classified with a 17-layered Recurrent Neural Network (RNN), which included Bi-LSTM. The architecture was designed to capture dependencies in sequential heart sound data (past and future), enabling the model to differentiate normal and abnormal sounds with significant accuracy.



Fig. 12. The Bi-LSTM block diagram.

The model design is illustrated in Fig. 12. The description of each layer in the proposed model is presented as follows:

- The input layer receives selected features.
- The input features were converted into dense vectors in the embedding layer.
- Bi-LSTM Layer: Processes input in the forward and backward directions to capture temporal dependencies.

- The dropout layers were applied following the Bi-LSTM and dense layers to prevent overfitting.
- A total of 13 dense layers followed were included to transform the features.
- A final dense layer, the output layer, contains softmax activation for normal or abnormal heart sound categorisation.
- The forward and backward Bi-LSTM passes are presented by equations (11) and (12), respectively. This study utilised the Bi-LSTM layer model due to its bidirectional dependencies in PCG signals, essential for accurately categorising heart sounds. Its 17-layer structure also allows the model to learn complex patterns, improving classification performance. The proposed model architecture is illustrated in Fig. 13.

$$\vec{h}_t = \sigma(W_h \cdot [\vec{h}_{t-1}, x_t] + b_h) \quad (11)$$

$$\hat{h}_t = \sigma(W_h \cdot [\hat{h}_{t+1}, x_t] + b_h) \quad (12)$$

Where \vec{h}_t and \hat{h}_t represent hidden states in the forward and backward directions, respectively.

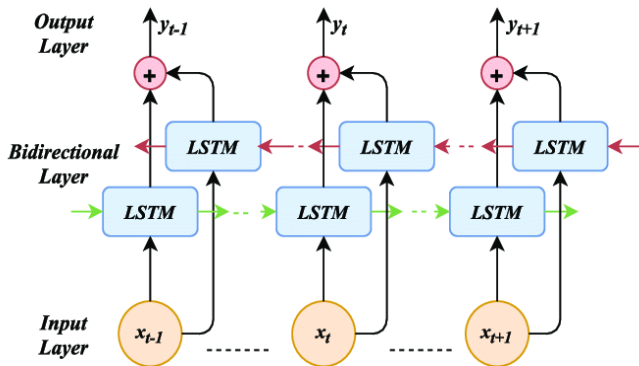


Fig. 13. The Bi-LSTM-incorporated RNN architecture.

IV. EXPERIMENTS AND RESULTS

The dataset and evaluation criteria employed in this study are discussed in this section. The experiments were conducted according to the specified criteria, and the results obtained are also included.

A. Dataset

The current study explored three datasets commonly applied in related research: PhysioNet 2016 [49] and 2022 [50] and Yaseen Khan 2018 [22]. The PhysioNet 2016 dataset consists of 3,240 PCG signals from global volunteers. The duration of the heart sounds in the dataset ranged between 5 and 120 sec (s) recorded at 2,000 Hz. Furthermore, the dataset includes normal and abnormal heart sounds.

The Yaseen Khan 2018 dataset comprises 1,000 2-s PCG signals. The sounds were sampled at 8,000 Hz with binary and multi-class labels. The third dataset, PhysioNet 2022, contains 3,163 PCG signals of 5 to 65 s recorded at 4,000 Hz. In this study, each dataset was split with the k-fold and stratified k-fold (K = 5 and 10) Cross-Validation (CV) methods.

B. Performance Evaluation Metrics

The accuracy, sensitivity, specificity, and F1 score were the evaluation criteria of the proposed model. Each criterion was determined according to Eq. (13)–(16). Meanwhile, Fig. 14 demonstrates the confusion matrix-based evaluation procedure applied in the present study [50].

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \quad (13)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (14)$$

$$Specificity = \frac{TN}{TN+FP} \quad (15)$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (16)$$

		Abnormal	Normal	
Output Class	Abnormal	True Positive (TP)	False Positive (FP)	Precision = $\frac{TP}{(TP + FP)}$
	Normal	False Negative (FN)	True Negative (TN)	Negative Predictive Value = $\frac{TN}{(TN + FN)}$
		Sensitivity (recall) = $\frac{TP}{(TP + FN)}$	Sepecificity = $\frac{TN}{(TN + FP)}$	Accuracy = $\frac{TP + TN}{(TP + TN + FP + FN)}$
		Target Class		

Fig. 14. Evaluation measures using confusion matrix.

C. Experimental Results

A comprehensive explanation of the results is included in this section, documenting the performance of the proposed method. The suggested model was implemented with MATLAB 2023, while training and testing were conducted on a system equipped with an Intel Core i7 processor (2.6 GHz), 32 GB RAM, and NVIDIA RTX 960M GPU (4 GB).

During the initial experiments (baseline), 3,240, 1,000, and 3,163 PCG signals from PhysioNet 2016, Yaseen Khan 2018, and PhysioNet Challenge 2022, respectively, were assessed. The datasets were split with hold-out CV, with 85% utilised as training, while 15% were employed during the testing stage.

A 5-s time window applied on the PhysioNet 2016 dataset produced 10,000 samples at 2,000 Hz frequency, while the PhysioNet Challenge 2022 generated 20,000 records at 4 kHz. Meanwhile, 8,000 samples recorded at 8 kHz were procured from the Yaseen Khan 2018 with a 1-s interval.

Several methods, including MFCC, DWT, and WST, were employed to extract raw features from each dataset at varying window sizes (5, 10, and 8 WS). The present study also utilised the Bi-LSTM algorithm with 17 layers for training, applying K-fold and stratified (S) K-fold (K = 5 and 10) CV approaches to mitigate data imbalance. Classification performance was then evaluated based on accuracy, sensitivity (Sen%), specificity (Spec%), and F1 score (F1%). Whereas, WL stands for window length, WS stands for window size, and L stands for the number of decomposition levels of DWT.

Table I summarises the time and frequency domains of the PCG signals acquired from the three datasets evaluated. The findings indicated that each dataset required separate training. The WST technique extracted numerous features from the PCG signals, achieving the highest baseline accuracy. Nonetheless, the accuracy of the model required improvement due to

background noise in the PCG signals and irrelevant extracted features, which affect classification precision. Moreover, the stratified 10-fold CV recorded a notable accuracy rate when the datasets were individually assessed. Consequently, the approach was utilised in the proposed experiment, applying different window sizes for each dataset (see Table II).

TABLE I. THE BASELINE EXPERIMENTAL RESULTS FOR HEART SOUND ANALYSIS

Dataset	WL-feature extraction method		CV	Acc%	Sen%	Spe%	F1%
PhysioNet 2016, 3,240 PCG signals	5s-WL, 27-MFCC		5-folds	84.74	60.90	92.53	66.29
			10-folds	88.29	67.90	94.95	74.07
			S-5-folds	89	76.33	93.14	77.37
			S-10-folds	87.32	68.72	93.40	72.76
	5s-10WS-DWT	265 (L52)	5-folds	91.09	72.18	95.74	76.19
		115 (L22)	10-folds	91.41	70.62	96.53	76.48
		75 (L14)	S-5-folds	91.54	72.18	96.30	77.12
		80 (L15)	S-10-folds	91.25	76.04	95.00	77.45
	5s-10WS	263-WST	5-folds	92.48	75.93	96.56	79.97
			10-folds	91.76	71.14	96.84	77.34
S-5-folds			92.53	75.10	96.82	79.88	
S-10-folds			92.69	74.79	97.10	80.17	
Yaseen Khan 2018, 1,000 PCG signals	1s-WL, 27-MFCC		5-folds	99.66	100	97.61	99.80
			10-folds	99.66	99.61	100	99.80
			S-5-folds	99.66	99.61	100	99.80
			S-10-folds	99.66	99.61	100	99.80
	1s-8WS-DWT	30 (L5)	5-folds	100	100	100	100
		40 (L7)	10-folds	100	100	100	100
		45 (L8)	S-5-folds	100	100	100	100
		30 (L5)	S-10-folds	100	100	100	100
	1s-8WS	261-WST	5-folds	99.91	99.90	100	99.95
			10-folds	99.91	99.90	100	99.95
S-5-folds			100	100	100	100	
S-10-folds			100	100	100	100	
PhysioNet 2022, 3,163 PCG signals	5s-WL, 27-MFCC		5-folds	50.61	30.36	69.58	37.28
			10-folds	50.56	49.57	51.49	49.23
			S-5-folds	53.79	59.97	48.01	55.66
			S-10-folds	51.64	36.51	65.80	42.20
	5s-5WS-DWT	20 (L3)	5-folds	57.17	47.43	66.04	51.36
		30 (L5)	10-folds	59.49	59.11	59.83	58.18
		200 (L39)	S-5-folds	60.59	78.23	44.51	65.43
		65 (L12)	S-10-folds	61.56	65.30	58.14	61.83
	5s-5WS	348-WST	5-folds	57.84	58.23	57.50	56.84
			10-folds	61.43	60.44	62.33	59.91
S-5-folds			59.87	61.15	58.70	59.23	
S-10-folds			62.82	65.13	60.72	62.55	

TABLE II. THE WINDOW SIZE RECOMMENDATION FOR THE DATASETS

Dataset	Time and frequency domains	Feature extraction method	Window size	Classifier	CV
PhysioNet 2016	5 sec, 2,000 Hz	WST	10	Bi-LSTM algorithm (17 layers)	Stratified 10-fold
Yaseen Khan 2018	2 sec, 8,000 Hz		8		
PhysioNet 2022	5 sec, 4,000 Hz		5		

TABLE III. THE METAHEURISTIC RESULTS UNDER DEFAULT PARAMETERS

Dataset	Preprocessing	WL-feature extraction and selection method	Acc%	Sen%	Spe%	F1%
PhysioNet 2016, 3,240 PCG signals	Butterworth filter (low-pass), cosine filter (low-pass), normalisation (-1, 1)	132-HHO	92.44	77.39	96.15	80.19
		141-DA	92.55	75.41	96.76	80.00
		122-GWO	92.32	78.02	95.84	80.06
		126-SSA	92.24	75.72	96.30	79.41
		132-WOA	92.59	78.95	95.94	80.81
Yaseen Khan 2018, 1,000 PCG signals	Butterworth filter (low-pass), cosine filter (low-pass), normalisation (-1, 1), and zero padding	168-HHO	100	100	100	100
		175-DA	100	100	100	100
		155-GWO	100	100	100	100
		151-SSA	100	100	100	100
		59-WOA	100	100	100	100
PhysioNet 2022, 3,163 PCG signals	Butterworth filter (low-pass), cosine filter (high-pass), normalisation (-1, 1)	195-HHO	60.67	69.20	52.90	62.66
		193-DA	59.62	60.61	58.70	58.87
		185-GWO	57.17	56.54	57.74	55.73
		175-SSA	61.68	55.39	67.41	57.96
		181-WOA	60.59	57.87	63.06	58.34

During the evaluations, each dataset was assessed independently and split with holdout CV (see Table III). Low-pass Butterworth and cosine filters were applied during preprocessing to reduce high-frequency intensities from the PhysioNet 2016 and Yaseen Khan 2018 datasets at 200 Hz and 800 Hz, respectively, cut-off frequencies. Meanwhile, the PhysioNet 2022 dataset was subjected to low-pass Butterworth and high-pass cosine filters within the 15–400 Hz cut-off frequency.

After preprocessing, each dataset was normalised to a range of -1 to 1. A 5-s duration was applied to the PhysioNet 2016 and 2022 datasets, while Yaseen Khan 2018 was subjected to 2-s intervals with zero padding. Subsequently, the features were extracted with the WST. Resultantly, 263 WST features with 10-WS were procured from the PhysioNet 2016 dataset, 330 WST features with 8-WS from Yaseen Khan 2018, and 348 WST features with 5-WS from PhysioNet 2022. Metaheuristic techniques were employed during feature selection, including HHO, DA, GWO, SSA, and WOA. Default parameters and hyperparameters were applied to enhance classification accuracy.

Initially, all metaheuristic approaches were employed under default parameters of 10 maximum iterations, a 10-population size, and a 5 K-value for the KNN classifier. Holdout CV of

80% training and 20% testing, 0.99 alpha (α), and 0.01 beta (β) were also applied per the information reported in previous studies.

The 17-layer Bi-LSTM architecture developed in this study was utilised to train the classification model. A stratified 10-fold CV was also applied to overcome the class imbalance. The metaheuristic methods were then evaluated with test data with classification accuracy as the key performance metric. The results are listed in Table III.

Under default parameters, the proposed model documented significant results with 100% accuracy when applied to the Yaseen Khan 2018 dataset utilising all metaheuristic techniques. Conversely, the accuracy of the PhysioNet 2016 and 2022 datasets was not considerably higher than the baseline. Consequently, the current study employed hyperparameters for all metaheuristic methods in the second task of the proposed model as indicated in Table IV.

Preprocessing methods under hyperparameter settings and training step with the Bi-LSTM classifier (17 layers) were applied to the PhysioNet 2016 and 2022 datasets to create the classification model in this study. A stratified 10-fold CV was employed during training. Fig. 15 and Fig. 16 demonstrate the findings from the PhysioNet 2016 and 2022 datasets, respectively.

TABLE IV. THE HYPERPARAMETERS-USED FOR THE PHYSIONET 2016 AND 2022 DATASETS

Parameters	PhysioNet 2016	PhysioNet 2022
Cross-Validation	Holdout (85% training, 15% testing)	Holdout (80% training, 20% testing)
Fitness Value	$\beta = 1e-10 - 1e-1$	$\beta = 0.01 - 0.1$
Population Size	10	
Classifier	KNN	
k-value	11	5
Max iterations (M-iters)	10	

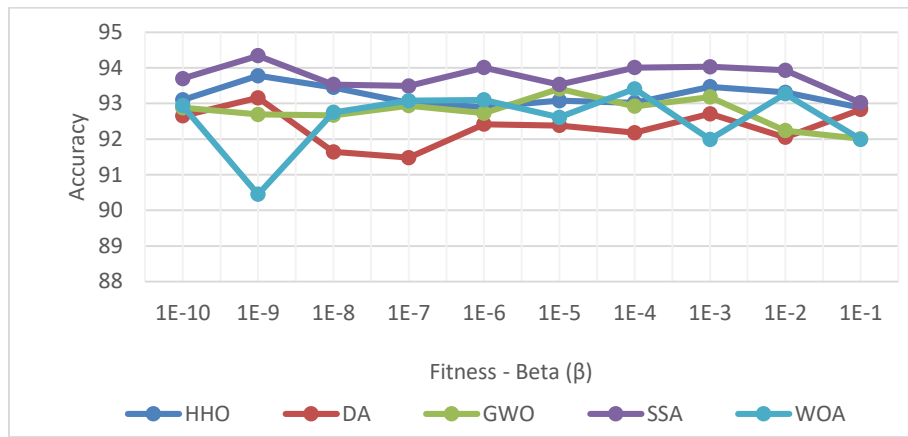


Fig. 15. The best fitness value and accuracy from the PhysioNet 2016 dataset when hyperparameter settings were applied for HHO, DA, GWO, SSA, and WOA.



Fig. 16. The optimal fitness value and accuracy of the PhysioNet 2022 dataset were evaluated with HHO, DA, GWO, SSA, and WOA under hyperparameter settings.

Based on Fig. 15 and Fig. 16, improved classification accuracy was observed when the hyperparameters for the Physionet 2016 and 2022 datasets were adjusted. The datasets recorded excellent accuracy rates at SSA and HHO tuned to 94.34% and 66.83%, respectively. The data were considered more accurate than the baseline and default parameters. The findings are summarised in Table V. Fig. 17 to Fig. 19 and Table VI illustrate the results of real-world experiments through majority voting of the datasets employed.

According to Table VI, a 94.85% final accuracy rate was recorded by the PhysioNet 2016 dataset in real-world evaluations without voting ties (NoUniqueMode = 0). During the assessment, WST (263 extracted features, 5 sec) and SSA (127 selected features) with the Bi-LSTM (17 layers) algorithm were applied (see Fig. 17).

TABLE V. SUMMARY OF THE METAHEURISTIC RESULTS UNDER HYPERPARAMETERS

Dataset	WL-feature extraction and selection method	Optimal-fitness value (β)	k/M-iters	Acc%	Sen%	Spe%	F1%	
PhysioNet 2016, 3,240 PCG signals	5s-10WS-263WST	142-HHO	1E-9	11/10	93.78	81.87	96.71	83.88
		137-DA	1E-9		93.16	81.77	95.97	82.54
		128-GWO	1E-5		93.41	81.45	96.35	83.01
		127-SSA	1E-9		94.34	82.29	97.30	85.17
		147-WOA	1E-4		93.41	79.47	96.84	82.66
PhysioNet 2022, 3,163 PCG signals	5s-5WS-348WST	163-HHO	0.04	5/10	66.83	60.70	72.41	63.57
		178-DA	0.04		62.02	57.61	66.04	59.12
		183-GWO	0.09		60.50	55.92	64.67	57.45
		173-SSA	0.07		64.43	62.74	65.96	62.71
		181-WOA	0.03		64.89	61.59	67.90	62.58

TABLE VI. SUMMARY OF THE FINAL EXPERIMENTAL RESULTS OF THE PROPOSED METHOD

Dataset	Feature extraction and selection method	Acc%	Sen%	Spe%	F1%
PhysioNet 2016, 3,240 PCG signals	263-WST, 127-SSA	94.85	83.33	97.69	86.48
Yaseen Khan 2018, 1,000 PCG signals	330-WST, All metaheuristic methods	100	100	100	100
PhysioNet 2022, 3,163 PCG signals	348-WST, 163-HHO	66.87	60.61	72.58	63.57

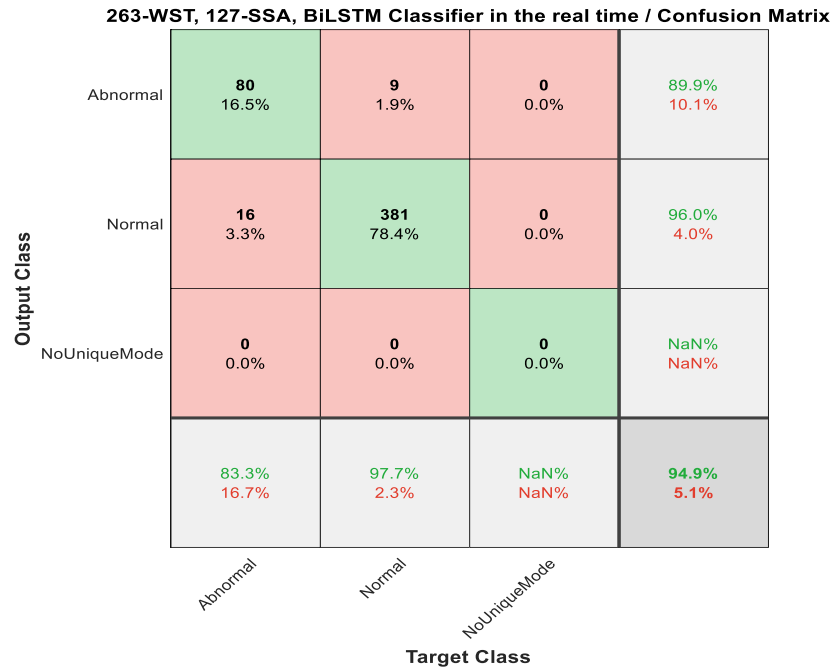


Fig. 17. The final experimental results for the PhysioNet 2016 dataset evaluated under hyperparameters by 127-SSA.

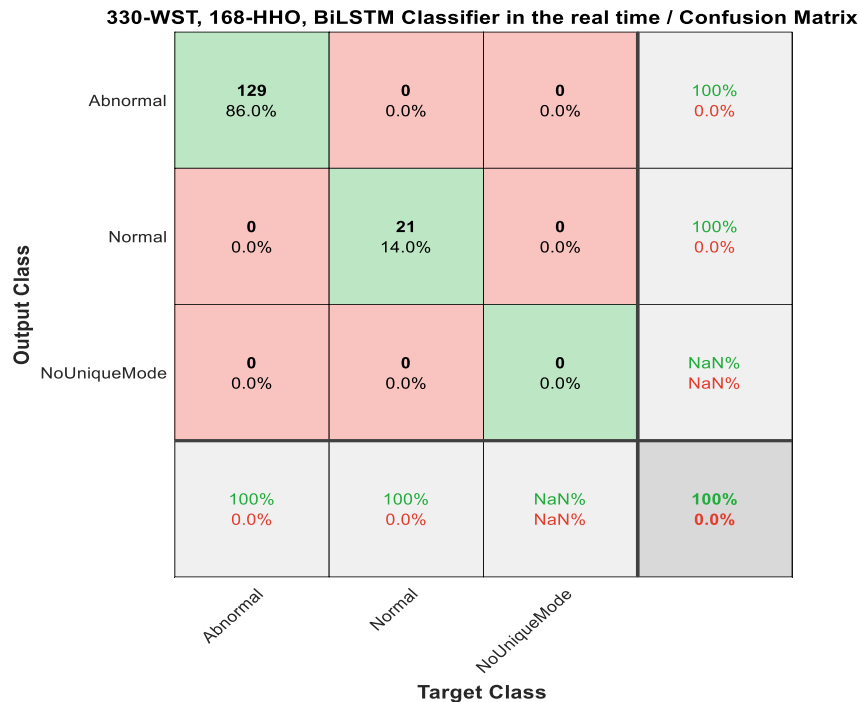


Fig. 18. The final experimental findings for the Yaseen Khan 2018 dataset with default parameters in all metaheuristic methods.

348-WST, 163-HHO, BiLSTM Classifier in the real time / Confusion Matrix

Output Class	Abnormal	137 28.9%	68 14.3%	0 0.0%	66.8% 33.2%
	Normal	89 18.8%	180 38.0%	0 0.0%	66.9% 33.1%
	NoUniqueMode	0 0.0%	0 0.0%	0 0.0%	NaN% NaN%
		60.6% 39.4%	72.6% 27.4%	NaN% NaN%	66.9% 33.1%
	Abnormal	Normal	NoUniqueMode		
	Target Class				

Fig. 19. The final experimental results for the PhysioNet 2022 dataset assessed with hyperparameter settings by 163-HHO.

In the Yaseen Khan 2018 real-world evaluation, the 2-s proposed model with WST and all metaheuristic methods achieved substantial results with 100% accuracy without any prediction ties. The results are summarised in Fig. 18. Meanwhile, the PhysioNet 2022 dataset achieved a final accuracy of 66.87% in the real-world assessment with the WST (348 extracted features, 5 sec) and HHO (163 selected features) methods. The data was considered more accurate than the baseline and default parameter values. The superiority was due to reduced redundant features without any ties in the voting classes, as illustrated in Fig. 19.

D. Discussion

Overall, HHO, DA, GWO, SSA, and WOA algorithms were more sensitive and negatively affected in imbalanced datasets. The metaheuristic techniques revealed similar efficiency and convergence speed limitations and global solution procurement issues. Furthermore, randomisation is crucial during the exploration and exploitation phases. Accordingly, increased randomisation would lead to classification accuracy from elevated computational time.

Based on the results, the SSA approach improved the accuracy and performance of the classification model in the PhysioNet 2016 dataset. The method exhibited superior exploration abilities for features at low frequencies than the other metaheuristic methods employed in this study.

All metaheuristic techniques utilised in the current study achieved a considerable accuracy rate when applied to the Yaseen Khan 2018 dataset. Meanwhile, the HHO approach performed better than GWO, SSA, and WOA on the PhysioNet 2022 dataset. The findings might be due to the significant exploration capabilities for features at high frequencies of HHO.

In this study, the performance of the proposed method utilising the metaheuristic techniques was compared to recently published reports that aimed to diagnose heart sounds utilising PCG signals (see Table VII). The highest accuracy rates achieved with the PhysioNet 2016 dataset were reported by [11] and [12] at 93.5% and 93.33%, respectively. Nonetheless, [11] primarily focused on extracting 527 features without applying any technique to eliminate irrelevant features. Whereas [12] procured 130 features with MFCCs and the traditional LDA technique to select optimal attributes.

For the 1,000 PCG Yaseen Khan 2018 dataset, [18] and [47] documented the highest scores, 99.80% and 99.90%, respectively. Nevertheless, [23] only employed MFCC, DWT, and CWT to extract several features. Conversely, the proposed method was based on the WST. All metaheuristic approaches also recorded a significant accuracy rate of 100% when applied to the same dataset.

Reports on applying feature selection methods to PCG signals, particularly metaheuristic approaches, are limited. For instance, [14] extracted deep features and employed PSO for improvement before utilising the Bi-LSTM algorithm, recording a 91.93% accuracy rate. Consequently, the present study aimed to close the knowledge gap, focusing on selecting optimal features based on metaheuristic approaches to improve the performance of the model.

This study achieved the best performance with 94.85%, 83.33%, 97.69%, and 86.48% accuracy, sensitivity, and specificity rates and F1 score. The WST based on the SSA method was employed. The data supported the superiority of the suggested model.

TABLE VII. RESULTS COMPARISON BETWEEN THE PROPOSED METHOD AND PREVIOUS STUDIES

Author	Dataset used	Feature extraction method	Feature selection method	Classifier	CV method	Acc (%)	Sen (%)	Spe (%)	F1 (%)
[11]	PhysioNet 2016 dataset (3,240 PCG signals)	Multi-domain features (527 features)	Not used	Fine KNN	Hold-out	93.5	-	-	-
[12]	PhysioNet 2016 dataset (3,126 PCG signals)	MFCCs (130 features)	LDA	ANN	Hold-out	93.33	-	-	-
[14]	PhysioNet 2016 dataset (3,240 PCG signals)	Deep features	PSO	Bi-LSTM	5-fold	91.93	91.58	92.27	-
[16]	PhysioNet 2016 dataset (3,153 PCG signals)	CWT and MFCC (675 features)	PCA	ANN	10-fold	85.2	-	-	-
[19]	PhysioNet 2016 dataset (3,240 PCG signals)	MFCCs (27 features)	Not used	Ensemble classifier	5-fold	92.47	94.08	91.95	-
[21]	PhysioNet 2016 (1,330 PCG signals)	Spectrogram feature	Not used	VGG16	-	88.84	87.23	86.55	87.87
[22]	Own dataset	MFCCs + DWT (43 features)	Not used	SVM	5-fold	97.9	98.2	99.4	99.7
[23]	Yaseen Khan 2018 dataset (1,000 PCG signals)	MFCC, DWT, and CWT	Not used	CNN and MLP	Hold-out	99.8	99.8	99.8	-
[24]	Yaseen Khan 2018 dataset (957 PCG signals)	Multiple features using MFCC and DWT methods	Not used	RF	-	99.89	99.90	99.60	99.90
[25]	PhysioNet 2022 dataset (3,163 PCG signals)	Spectrogram feature	Not used	Bi-GRU	5-fold	60.2	-	-	54.9
[26]	PhysioNet 2022 dataset (3,163 PCG signals)	Mel-spectrograms	Not used	U-Net	Hold-out	56.80	54.77	59.29	58.33
[28]	PhysioNet 2016 (2,435 PCG signals)	Merage short- and long-term features (33 features)	Not used	Subspace K-Nearest Neighbor	Hold-out	92.7	96	82	-
[51]	Yaseen Khan 2018 dataset (1,000 PCG signals)	Deep features	Not used	Vision Transformer (ViT)	10-fold	99.90	99.90	99.90	99.90
[52]	PhysioNet 2016 dataset (3,153 PCG signals)	Multiple features (515 features)	Selected features randomly (400 features)	SVM	10-fold	88	88	87	-
[53]	PhysioNet 2016 dataset (3,126 PCG signals)	Zero crossing rate (ZCR), discrete fourier transform (DFT), and MFCCs (315 features)	Genetic Algorithm (GA) (15 features)	LightGBM (Light Gradient Boosting)	10-fold	92.3	91.06	93.54	-
[54]	Yaseen Khan 2018 dataset (1,000 PCG signals)	Time-varying spectral feature (35 features)	Not used	KNN	5-fold	99.6	99.79	98.83	99.75
[55]	Yaseen Khan 2018 dataset (1,000 PCG signals)	Chirplet Transform (CT) (300 features)	Not used	Composite Classifier	Hold-out	98.33	-	-	-
[56]	PhysioNet 2022 dataset (3,163 PCG signals)	MFCCs (25 features)	Not used	Ensemble classifier	10-fold	56.8	-	-	52.8
Propose Methods	PhysioNet 2016 dataset (3,240 PCG signals)	WST (263 features)	SSA (127 features)	Bi-LSTM (17 layers)	Stratified 10-fold	94.85	83.33	97.69	86.48
	Yaseen Khan 2018 (1,000 PCG signals)	WST (330 features)	HHO, DA, GWO, SSA, and WOA			100	100	100	100
	PhysioNet 2022 (3,163 PCG signals)	WST (348 features)	HHO (163 features)			66.87	60.61	72.58	63.57

The study in [25] documented an accuracy of 60.2% for the PhysioNet Challenge 2022 dataset with spectrogram features. The less accurate results were from the substantially noisy PCG signals, rendering the extraction of vital features and improvement of the classification model challenging. Nonetheless, the proposed design achieved an optimal solution by obtaining an accuracy rate of 66.87% when applied to the

same dataset. Furthermore, the current study effectively determined the sensitivity, specificity, and F1 scores, yielding 60.61%, 72.58%, and 63.57%, respectively. The suggested method diminished the number of extracted features, positively affecting convergence speed during the training process and achieving excellent performance.

V. CONCLUSION AND FUTURE DIRECTIONS

Medically, diagnosing heart sounds faces significant obstacles, particularly PCG signal processing, due to noise, imbalanced datasets, and the considerable extracted feature search space. Numerous researchers have addressed the issues and developed various solutions. Nonetheless, related reports on applying metaheuristic techniques to PCG signals are scarce. The non-linear nature of the data and its complicated relationships have contributed to the knowledge gap.

The current study proposed a model employing WST to address the challenges posed by noise signals and irrelevant extracted features. Hyper-filters were applied to mitigate the impact of noise, while metaheuristic optimisation techniques, HHO, DA, GWO, SSA, and WOA, were utilised to select the most informative WST features. The selected features then served as input to a Bi-LSTM algorithm to produce the classification model. Moreover, a stratified 10-fold cross-validation was implemented to mitigate the effects of imbalanced datasets and overfitting. Resultantly, the metaheuristic methods documented potential, exhibiting 100% accuracy with the Yaseen Khan 2018 dataset with default parameters. Meanwhile, the classification accuracy of the suggested model on the PhysioNet 2016 and 2022 datasets under hyperparameter settings was 94.85% and 66.87% with SSA and HHO, respectively.

Overall, the proposed method documented superior results to previous research. The suggested model might also improve clinical finding reliability. However, the limitations are still having in this study, such as noise in PCG signals, imbalanced datasets, and unnecessary features. Consequently, future studies should enhance the attribute by focusing on several key areas. For instance, utilising a more significant PCG signal dataset and refining the preprocessing techniques by applying deep filters to reduce noise. The imbalanced classes issue could also be mitigated by enhancing the stratified K-fold CV process. Furthermore, extracting multiple features with CWT, DWT, and WST techniques could be considered. Finally, an improved model performance might be achieved by applying hybrid metaheuristic optimisation methods, such as HHO-SSA, to select optimal features efficiently.

REFERENCES

- [1] J. Chen, Z. Guo, X. Xu, G. Jeon, and D. Camacho, "Artificial intelligence for heart sound classification: A review," *Expert Syst.*, vol. 41, no. 4, pp. 1–20, Apr. 2024, doi: 10.1111/exsy.13535.
- [2] A. Javeed, S. S. Rizvi, S. Zhou, R. Riaz, S. U. Khan, and S. J. Kwon, "Heart risk failure prediction using a novel feature selection method for feature refinement and neural network for classification," *Mob. Inf. Syst.*, vol. 2020, pp. 1–11, Aug. 2020, doi: 10.1155/2020/8843115.
- [3] H. Yadav et al., "CNN and Bidirectional GRU-based heartbeat sound classification architecture for elderly people," *Mathematics*, vol. 11, no. 6, pp. 1–25, Mar. 2023, doi: 10.3390/math11061365.
- [4] W. Chen, Q. Sun, X. Chen, G. Xie, H. Wu, and C. Xu, "Deep learning methods for heart sounds classification: A systematic review," *Entropy*, vol. 23, no. 6, pp. 1–18, May 2021, doi: 10.3390/e23060667.
- [5] Z. Ren, Y. Chang, T. T. Nguyen, Y. Tan, K. Qian, and B. W. Schuller, "A comprehensive survey on heart sound analysis in the deep learning era," *arXiv Prepr. arXiv2301.09362*, pp. 1–16, May 2024, [Online]. Available: <https://arxiv.org/abs/2301.09362>
- [6] F. Li, Z. Zhang, L. Wang, and W. Liu, "Heart sound classification based on improved mel-frequency spectral coefficients and deep residual learning," *Front. Physiol.*, vol. 13, pp. 1–16, Dec. 2022, doi: 10.3389/fphys.2022.1084420.
- [7] N. B. Aji, K. Kurnianingsih, N. Masuyama, and Y. Nojima, "CNN-LSTM for heartbeat sound classification," *JOIV Int. J. Informatics Vis.*, vol. 8, no. 2, pp. 735–741, May 2024, doi: 10.62527/joiv.8.2.2115.
- [8] M. F. A. Ben Hamza and N. N. A. Sjarif, "A comprehensive overview of heart sound analysis using machine learning methods," *IEEE Access*, vol. 12, pp. 117203–117217, Jul. 2024, doi: 10.1109/ACCESS.2024.3432309.
- [9] M. U. Khan, S. Zuriat-e-Zehra Ali, A. Ishtiaq, K. Habib, T. Gul, and A. Samer, "Classification of multi-class cardiovascular disorders using ensemble classifier and impulsive domain analysis," in 2021 Mohammad Ali Jinnah University International Conference on Computing (MAJICC), IEEE, Jul. 2021, pp. 1–8. doi: 10.1109/MAJICC53071.2021.9526250.
- [10] R. M. Potdar, M. R. Meshram, and R. Kumar, "Optimal Parameter Selection for DWT based PCG Denoising," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 10, pp. 7521–7532, Apr. 2021, doi: 10.17762/turcomat.v12i10.5658.
- [11] O. Alshamma, F. H. Awad, L. Alzubaidi, M. A. Fadhel, Z. M. Arkah, and L. Farhan, "Employment of multi-classifier and multi-domain features for PCG recognition," in 2019 12th International Conference on Developments in eSystems Engineering (DeSE), IEEE, Oct. 2019, pp. 321–325. doi: 10.1109/DeSE.2019.00066.
- [12] M. G. M. Milani, P. E. Abas, L. C. De Silva, and N. D. Nanayakkara, "Abnormal heart sound classification using phonocardiography signals," *Smart Heal.*, vol. 21, pp. 1–18, Jul. 2021, doi: 10.1016/j.smhl.2021.100194.
- [13] Y. He, W. Li, W. Zhang, S. Zhang, X. Pi, and H. Liu, "Research on segmentation and classification of heart sound signals based on deep learning," *Appl. Sci.*, vol. 11, no. 2, pp. 1–15, Jan. 2021, doi: 10.3390/app11020651.
- [14] L. Zhang, C. P. Lim, Y. Yu, and M. Jiang, "Sound classification using evolving ensemble models and particle swarm optimization," *Appl. Soft Comput.*, vol. 116, pp. 1–28, Feb. 2022, doi: 10.1016/j.asoc.2021.108322.
- [15] C. Potes, S. Parvaneh, A. Rahman, and B. Conroy, "Ensemble of feature based and deep learning based classifiers for detection of abnormal heart sounds," in 2016 Computing in Cardiology Conference (CinC), Vancouver, BC, Canada: IEEE, Sep. 2016, pp. 621–624. doi: 10.22489/CinC.2016.182-399.
- [16] E. Kay and A. Agarwal, "Drop connected neural networks trained on time-frequency and inter-beat features for classifying heart sounds," *Physiol. Meas.*, vol. 38, no. 8, pp. 1–15, Mar. 2017, doi: 10.17863/CAM.12452.
- [17] X. Bao, Y. Xu, and E. N. Kamavuako, "The effect of signal duration on the classification of heart sounds: A deep learning approach," *Sensors*, vol. 22, no. 6, pp. 1–14, Mar. 2022, doi: 10.3390/s22062261.
- [18] J. Li, L. Ke, Q. Du, X. Ding, X. Chen, and D. Wang, "Heart sound signal classification algorithm: A combination of wavelet scattering transform and twin support vector machine," *IEEE Access*, vol. 7, pp. 179339–179348, Dec. 2019, doi: 10.1109/ACCESS.2019.2959081.
- [19] S. A. Singh and S. Majumder, "Short unsegmented PCG classification based on ensemble classifier," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 28, no. 2, pp. 875–889, Mar. 2020, doi: 10.3906/elk-1905-165.
- [20] R. M. Potdar, M. R. Meshram, and R. Kumar, "Optimization of automatic PCG analysis and CVD diagnostic system," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 11, pp. 3738–3751, May 2021, [Online]. Available: <https://turcomat.org/index.php/turkbilmart/article/view/6456>
- [21] K. E. K. Blitti, F. G. Tola, P. Wangdi, D. Kumar, and A. Diwan, "Heart sounds classification using frequency features with deep learning approaches," in 2024 IEEE Applied Sensing Conference (APSCON), IEEE, Jan. 2024, pp. 1–4. doi: 10.1109/APSCON60364.2024.10465862.
- [22] Yaseen, G.-Y. Son, and S. Kwon, "Classification of heart sound signal using multiple features," *Appl. Sci.*, vol. 8, no. 12, pp. 1–14, Nov. 2018, doi: 10.3390/app8122344.

- [23] S. I. Flores-Alonso, B. Tovar-Corona, and R. Luna-García, “Deep learning algorithm for heart valve diseases assisted diagnosis,” *Appl. Sci.*, vol. 12, no. 8, pp. 1–18, Apr. 2022, doi: 10.3390/app12083780.
- [24] S. Swaminathan, S. M. Krishnamurthy, C. Gudada, S. K. Mallappa, and N. Ail, “Heart sound analysis with machine learning using audio features for detecting heart diseases,” *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.*, vol. 16, no. 2, pp. 131–147, May 2024.
- [25] A. McDonald, M. JF Gales, and A. Agarwal, “Detection of heart murmurs in phonocardiograms with parallel hidden semi-markov models,” in *Conference: 2022 Computing in Cardiology (CinC)*, Tampere, Finland, Sep. 2022, pp. 1–4. doi: 10.22489/CinC.2022.020.
- [26] G. Singh, A. Verma, L. Gupta, A. Mehta, and V. Arora, “An automated diagnosis model for classifying cardiac abnormality utilizing deep neural networks,” *Multimed. Tools Appl.*, vol. 83, no. 13, pp. 39563–39599, Apr. 2024, doi: 10.1007/s11042-023-16930-5.
- [27] F. Li, H. Tang, S. Shang, K. Mathiak, and F. Cong, “Classification of heart sounds using convolutional neural network,” *Appl. Sci.*, vol. 10, no. 11, pp. 1–17, Jun. 2020, doi: 10.3390/app10113956.
- [28] M. Guven and F. Uysal, “A new method for heart disease detection: long short-term feature extraction from heart sound data,” *Sensors*, vol. 23, no. 13, pp. 1–15, Jun. 2023, doi: 10.3390/s23135835.
- [29] M. Wang, B. Guo, Y. Hu, Z. Zhao, C. Liu, and H. Tang, “Transfer learning models for detecting six categories of Phonocardiogram recordings,” *J. Cardiovasc. Dev. Dis.*, vol. 9, no. 3, pp. 1–17, Mar. 2022, doi: 10.3390/jcdd9030086.
- [30] S. Ismail, I. Siddiqi, and U. Akram, “Localization and classification of heart beats in phonocardiography signals —a comprehensive review,” *EURASIP J. Adv. Signal Process.*, vol. 2018, no. 1, pp. 1–27, Dec. 2018, doi: 10.1186/s13634-018-0545-9.
- [31] R. Khushaba, “Feature extraction using multisignal wavelet transform decom.” *GitHub*, Aug. 2020. [Online]. Available: <https://github.com/RamiKhushaba/getmswtfeat>
- [32] N. Dia, J. Fontecave-Jallon, P.-Y. Gumery, and B. Rivet, “Heart rate estimation from phonocardiogram signals using non-negative matrix factorization,” in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, May 2019, pp. 1293–1297. doi: 10.1109/ICASSP.2019.8682343.
- [33] B. Soro and C. Lee, “A wavelet scattering feature extraction approach for deep neural network based indoor fingerprinting localization,” *Sensors*, vol. 19, no. 8, pp. 1–12, Apr. 2019, doi: 10.3390/s19081790.
- [34] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, “Harris hawks optimization: Algorithm and applications,” *Futur. Gener. Comput. Syst.*, vol. 97, pp. 1–36, Aug. 2019, doi: 10.1016/j.future.2019.02.028.
- [35] M. Al-Kaabi, V. Dumbrava, and M. Eremia, “Multi criteria frameworks using new meta-heuristic optimization techniques for solving multi-objective optimal power flow problems,” *Energies*, vol. 17, no. 9, pp. 1–37, May 2024, doi: 10.3390/en17092209.
- [36] S. Mirjalili, “Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems,” *Neural Comput. Appl.*, vol. 27, no. 4, pp. 1053–1073, May 2016, doi: 10.1007/s00521-015-1920-1.
- [37] M. Alshinwan et al., “Dragonfly algorithm: a comprehensive survey of its results, variants, and applications,” *Multimed. Tools Appl.*, vol. 80, no. 10, pp. 14979–15016, Apr. 2021, doi: 10.1007/s11042-020-10255-3.
- [38] P. T. Prinson and A. Geetha, “Dragonfly algorithm and variants for feature selection: A review,” in *2023 International Conference on Quantum Technologies, Communications, Computing, Hardware and Embedded Systems Security (iQ-CCHES)*, IEEE, Sep. 2023, pp. 1–5. doi: 10.1109/iQ-CCHES56596.2023.10391693.
- [39] Y. Meraihi, A. Ramdane-Cherif, D. Acheli, and M. Mahseur, “Dragonfly algorithm: A comprehensive review and applications,” *Neural Comput. Appl.*, vol. 32, no. 21, pp. 16625–16646, Nov. 2020, doi: 10.1007/s00521-020-04866-y.
- [40] S. Mirjalili, S. M. Mirjalili, and A. Lewis, “Grey wolf optimizer,” *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014, doi: 10.1016/j.advengsoft.2013.12.007.
- [41] S. N. Makhadmeh et al., “Recent advances in grey wolf optimizer, its versions and applications: Review,” *IEEE Access*, vol. 12, pp. 22991–23028, Feb. 2024, doi: 10.1109/ACCESS.2023.3304889.
- [42] H. Faris, I. Aljarah, M. A. Al-Betar, and S. Mirjalili, “Grey wolf optimizer: a review of recent variants and applications,” *Neural Comput. Appl.*, vol. 30, no. 2, pp. 413–435, Jul. 2018, doi: 10.1007/s00521-017-3272-5.
- [43] Y. Liu, A. As’arry, M. K. Hassan, A. A. Hairuddin, and H. Mohamad, “Review of the grey wolf optimization algorithm: variants and applications,” *Neural Comput. Appl.*, vol. 36, no. 6, pp. 2713–2735, Feb. 2024, doi: 10.1007/s00521-023-09202-8.
- [44] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, “Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems,” *Adv. Eng. Softw.*, vol. 114, pp. 1–29, Dec. 2017, doi: 10.1016/j.advengsoft.2017.07.002.
- [45] L. Abualigah, M. Shehab, M. Alshinwan, and H. Alabool, “Salp swarm algorithm: a comprehensive survey,” *Neural Comput. Appl.*, vol. 32, no. 15, pp. 11195–11215, Aug. 2020, doi: 10.1007/s00521-019-04629-4.
- [46] S. Mirjalili and A. Lewis, “The whale optimization algorithm,” *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016, doi: 10.1016/j.advengsoft.2016.01.008.
- [47] Ş. Ay, E. Ekinici, and Z. Garip, “A comparative analysis of meta-heuristic optimization algorithms for feature selection on ML-based classification of heart-related diseases,” *J. Supercomput.*, vol. 79, no. 11, pp. 11797–11826, Jul. 2023, doi: 10.1007/s11227-023-05132-3.
- [48] D. Guha, P. K. Roy, and S. Banerjee, “Whale optimization algorithm applied to load frequency control of a mixed power system considering nonlinearities and PLL dynamics,” *Energy Syst.*, vol. 11, no. 3, pp. 699–728, Aug. 2020, doi: 10.1007/s12667-019-00326-2.
- [49] C. Liu et al., “An open access database for the evaluation of heart sound algorithms,” *Physiol. Meas.*, vol. 37, no. 12, pp. 2181–2213, Nov. 2016, doi: 10.1088/0967-3334/37/12/2181.
- [50] M. A. Reyna et al., “Heart murmur detection from phonocardiogram recordings: The George B. Moody PhysioNet Challenge 2022,” *PLOS Digit. Heal.*, vol. 2, no. 9, pp. 1–22, Sep. 2023, doi: 10.1371/journal.pdig.0000324.
- [51] S. Jamil and A. M. Roy, “An efficient and robust phonocardiography (PCG)-based valvular heart diseases (VHD) detection framework using vision transformer (ViT),” *Comput. Biol. Med.*, vol. 158, pp. 1–15, May 2023, doi: 10.1016/j.compbiomed.2023.106734.
- [52] H. Tang, Z. Dai, Y. Jiang, T. Li, and C. Liu, “PCG classification using multidomain features and SVM classifier,” *Biomed Res. Int.*, vol. 2018, no. 2, pp. 1–14, Jul. 2018, doi: 10.1155/2018/4205027.
- [53] N. K. Sawant, S. Patidar, N. Nesaragi, and U. R. Acharya, “Automated detection of abnormal heart sound signals using Fano-factor constrained tunable quality wavelet transform,” *Biocybern. Biomed. Eng.*, vol. 41, no. 1, pp. 1–44, Jan. 2021, doi: 10.1016/j.bbe.2020.12.007.
- [54] P. Upreti and M. E. Yuksel, “Accurate classification of heart sounds for disease diagnosis by a single time-varying spectral feature: Preliminary results,” in *Conference: 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT)*, IEEE, Apr. 2019, pp. 1–4. doi: 10.1109/EBBT.2019.8741730.
- [55] S. K. Ghosh, R. N. Ponnalagu, R. K. Tripathy, and U. R. Acharya, “Automated detection of heart valve diseases using chirplet transform and multiclass composite classifier with PCG signals,” *Comput. Biol. Med.*, vol. 118, pp. 1–17, Mar. 2020, doi: 10.1016/j.compbiomed.2020.103632.
- [56] Z. Imran, E. Grooby, C. Sitaula, V. Malgi, S. Aryal, and F. Marzbanrad, “A fusion of handcrafted feature-based and deep learning classifiers for heart murmur detection,” in *Conference: 2022 Computing in Cardiology (CinC)*, Tampere, Finland: IEEE, Dec. 2022, pp. 1–4. doi: 10.22489/CinC.2022.310.

Developing an Integrated Platform to Track Real Time Football Statistics for Somali Football Federation (SFF)

Bashir Abdinur Ahmed*, Husein Abdirahman Hashi, Abdifatah Abdilatif Ahmed, Abdikani Mahad Ali
Department of Computer Application, Jamhuriya University of Science and Technology, Mogadishu, Somalia

Abstract—The integration of technology in sports has revolutionized how stakeholders interact with and perceive the game. This thesis presents the development of an integrated platform aimed at tracking real-time football statistics for the Somali Football Federation (SFF). Football, being one of the most popular sports globally, relies heavily on accurate and up-to-date statistical data for player performance analysis, team strategies, and fan engagement. The SFF, like many other federations, faces challenges in collecting, managing, and utilizing football statistics effectively. The advent of digital technologies and the internet has revolutionized data collection and dissemination methods across various fields, including sports. Traditional methods of data collection and analysis, which are often manual and time-consuming, can no longer meet the demands of modern football analytics. The platform encompasses a mobile application for fans, an admin panel for administrators, and a backend system for data management. Leveraging modern technologies such as Flutter for mobile development, Node.js and MySQL for backend services, and React for the admin interface, the system ensures comprehensive coverage of match events, player statistics, and tournament standings. Real-time updates facilitated by Socket.IO enhance user engagement and decision-making capabilities for coaches and administrators.

Keywords—Real-time football statistics; Integrated sports platform; Somali Football Federation (SFF); user engagement; sports technology

I. INTRODUCTION

The Somali Football Federation (SFF) is a national administrative body under the Confederation of African Football (CAF) that oversees and regulates football activities and competitions within the Federal Republic of Somalia. This includes managing the first, second, and third divisions, as well as the Inter-regional Cup. Additionally, the SFF is responsible for the management of the Somalia national football team. The SFF's primary mission is to promote and develop football throughout the country [1].

In recent years, there has been a significant rise in the adoption of data analytics within the sports industry. This surge is attributed to sports teams' growing recognition of the immense value data can offer in enhancing performance and gaining a competitive advantage [2].

In their seminal work on the evolving landscape of sports, [3] underscored the remarkable transformation of sports into a dynamic, multifaceted competition, particularly evident in football's burgeoning role as a significant branch of the business world. They highlighted the intricate strategic manoeuvres that now characterize the sport, emphasizing that even pivotal

moments within a match often elude both human perception and the most advanced camera technologies.

Despite the growing popularity and adoption of data analytics in sports, the SFF currently lacks a comprehensive real-time football match statistics system. In the absence of such a system, the federation faces several challenges in effectively harnessing the power of data analytics to enhance its analytical capabilities and improve decision-making processes [4]. Traditional methods of match analysis rely on post-match statistics, which are often time consuming and do not provide immediate insights into critical match events.

For the SFF, embracing real-time statistical analysis is not just about keeping pace with global trends; it's about unlocking new opportunities for growth and competitive advantage. Integrating such analytical approaches can significantly improve the tactical planning and performance analysis of Somali football teams. The insights gained from real-time data can help coaches make informed decisions, tailor training programs to address specific weaknesses, and develop strategies that exploit the opposition's vulnerabilities. Furthermore, as outlined by study [4] the development of technologies for tracking and analysis in sports can play a crucial role in enhancing team performance, which is particularly relevant for federations looking to optimize their resources and talent.

This study outlines the development of a real-time football statistics platform personalized to the Somali Football Federation (SFF). It explores the challenges faced by the federation, reviews advancements in sports, and details the design and implementation of a comprehensive system that integrates real-time updates, data management, and user-friendly interfaces. The platform aims to enhance decision-making, and fan engagement while addressing the lack of existing solutions for Somali football statistics.

II. LITERATURE REVIEW

The transition to real-time analysis in football marks a significant milestone in the sport's analytical journey. This shift was propelled by technological advancements, enabling the collection and analysis of data in real-time during matches. Systems like the one introduced by [5] for multi-view event detection in soccer games exemplify the capabilities of real-time analytics to provide immediate insights into player movements, ball trajectories, and game dynamics.

The development of sports analytics is underpinned by several key theories that guide the collection, analysis, and interpretation of data. These theories include statistical models for predicting outcomes, optimization theories for team

*Corresponding Author.

composition and strategy, and performance analysis frameworks for evaluating player efficiency. For instance, the application of Poisson and negative binomial distributions to goal distributions in football, as discussed by [6], exemplifies the use of statistical models in sports analytics.

The literature review highlights several research gaps in the realm of real-time statistical analysis in football. One notable gap is the limited exploration of how these advanced analytics can be tailored and applied within the context of developing football nations, such as Somalia. Additionally, there is a need for real-time display of data which we hope to rectify [7].

This Research project aims to develop and implement a real-time football match statistics system to augment the analytical capabilities of the Somali Football Federation. This system will provide live updates on critical match events, including goals, cards, referees, and stadium information, thereby enhancing the viewing experience for fans and stakeholders alike. By leveraging real-time data insights, the federation seeks to empower coaches and decision-makers with the tools needed to make informed strategic choices and optimize player development programs [8].

Through the implementation of this real-time statistics system, the SFF aspires to set a new standard for football administration in the region, fostering greater engagement and interest in football across Somalia. This initiative represents a significant step towards embracing technology and innovation to propel Somali football into a new era of success and progress [9].

Researchers could focus on developing cost-effective and scalable analytics solutions that are accessible to football federations with limited resources. Investigating the integration of cultural and contextual factors into analytics models could also provide valuable insights, ensuring that the data interpretation is relevant and actionable for specific football environments [10]. Moreover, longitudinal studies on the impact of real-time analytics on player performance and injury prevention could significantly contribute to the field.

The study in [11] explore the use of machine learning in predicting football league standings and player performance, demonstrating how data-driven insights can enhance team strategies and performance optimization. This approach, applicable to real-time match statistics, can support the Somali Football Federation in making informed, data-driven decisions to improve player development and game strategies.

In addition to predictive models, the statistical dynamics of football have been extensively studied to understand the underlying mechanisms of the game. The study in [12] analyze the statistical dynamics of football, focusing on team behavior during matches. Their study applies statistical physics to understand game tactics, providing insights that can be integrated into real-time analytics systems. This can enhance decision-making for football federations like the SFF, helping improve tactical strategies during live matches.

Real-time football analytics have advanced with the development of sensors, machine learning, and computer vision technologies, enabling the collection of large amounts of data during matches. Technologies like wearable devices and video

analysis tools have transformed event tracking. Research on wearable technology shows how real-time biometric data can monitor player fitness, fatigue, and performance [13].

Studies by [14] have demonstrated the impact of well-designed visualizations in improving both player performance and fan engagement. This area could benefit from specific case studies on the design and usability of real-time systems tailored for football federations with varying levels of technical literacy and infrastructure, like the Somali Football Federation.

Another potential research gap could focus on the challenges of implementing real-time football analytics in countries with limited resources and infrastructure, such as Somalia. Research by study [15] highlights the barriers faced by sports organizations in developing nations, including financial constraints, technical know-how, and unreliable internet connectivity. Solutions such as low-cost data collection tools, offline-capable platforms, and mobile-based applications could be explored to bridge these gaps and make real-time football analytics more accessible.

These gaps and future research directions offer an exciting opportunity for the SFF and the academic community to contribute to the advancement of football analytics. By addressing these areas, there is potential to enhance the strategic application of analytics in football, promoting a more informed, effective, and competitive approach to the sport globally.

III. METHODOLOGY

Our research goal is to develop a comprehensive mobile application using Flutter, with a Node.js backend supported by MySQL, alongside a React-based admin panel. The primary objective is to create a platform that facilitates the management of Somali football statistics. This entails incorporating features such as real-time updates on match events, including goals, cards, and substitutions. Furthermore, the application will provide information on teams, referees, stadiums, match schedules, results, and standings, all presented in the Somali language to cater to the target audience. Notably, the absence of Somali football statistics APIs necessitates the dynamic entry of data through the React admin panel.

This platform prioritizes data security by implementing encrypted communication channels using TLS for real-time event broadcasting via Socket.IO. Authentication is managed through JWT to ensure secure access to sensitive data. Furthermore, player and fan data privacy is safeguarded by anonymizing personal data during statistical analysis and requiring user consent in data-sharing activities.

A. System Description

The system designed for the Somali Football Federation (SFF) is an integrated platform dedicated to tracking and managing real-time football statistics. This platform is tailored to meet the specific needs of Somali football, providing timely updates on key match events such as goals, cards, and substitutions. It also offers detailed information about teams, referees, stadiums, match schedules, results, and standings, ensuring a comprehensive view of the football landscape in Somalia. To cater to the local audience, the platform presents all data and content in the Somali language.

One of the system's core features is its ability to deliver real-time updates, making it a valuable tool for both fans and officials who require immediate access to match statistics. The platform's user interface is designed to be intuitive and accessible, with a focus on providing an engaging experience for users. Additionally, the system includes a React-admin panel, which allows authorized personnel to dynamically enter and manage data. This feature is particularly important due to the absence of existing APIs for Somali football statistics, ensuring that the platform remains accurate and up-to-date.

B. System Features

The application boasts a rich array of features designed to cater to the needs of Somali football enthusiasts. These features include:

Real-Time Goals: Users receive live updates on goals scored during matches.

Match Results: This feature enables users to quickly access match outcomes and review past performances.

Standings: Providing a snapshot of team rankings, this feature fosters healthy competition and engagement among fans of the teams.

Match Schedule: Users can plan and stay informed about upcoming matches.

Total Cards: This feature provides users with insights into the level of competitiveness and discipline exhibited by teams.

Substitutions: Users can track tactical changes made by teams during matches, gaining valuable insights into game strategies and player dynamics.

Referee Info: Transparency is maintained through the provision of information about the referees' officiating matches.

Stadium Information: This feature offers comprehensive details about the stadiums where matches are held.

Other tons of features that ensures comprehensive football platform.

C. System Requirements

The system requires hardware material and software programs, the most important requirement to run the platform are as follows:

1) Hardware requirements

a) Server: The server hosting the Node.js backend and MySQL database requires adequate CPU and RAM to support the application's backend operations effectively.

b) Mobile devices (Emulators): The mobile devices or emulators must be compatible with the Flutter framework.

c) Admin panel users: Standard computing devices, such as desktops or laptops, are required for accessing the React admin panel. These devices should have modern web browsers installed to ensure compatibility and smooth operation of the admin panel interface.

2) Software requirements

a) Mobile devices: The mobile devices must support the operating systems Android and iOS to run the Flutter application seamlessly.

b) Server: The server needs to have the Node.js runtime environment installed to execute the backend logic efficiently. Additionally, a MySQL database management system is essential for storing and managing the application data effectively.

c) Admin panel users: Admin panel users must have access to modern web browsers, such as Google Chrome, Mozilla Firefox, or Safari, to access and interact with the React admin panel seamlessly.

IV. SYSTEM ANALYSIS AND DESIGN

A. Current System and Drawbacks

Currently, the SFF relies on manual processes and social media posts, primarily Facebook, to update and inform stakeholders about match events such as goals, cards and substitutions. This approach has significant limitations:

- **Delayed Information:** Updates are not in real-time, which affects decision-making during matches.
- **Manual Entry:** The process is labour-intensive, time-consuming and prone high potential for human error.
- **Limited Coverage:** Big football apps like LiveScore and BeSoccer do not fully support Somali football, often displaying only match results with no detailed statistics.

B. Proposed System

The proposed system will leverage modern technology to provide a real-time football match statistics platform tailored to the needs of the SFF. The system will consist of a mobile application built with Flutter, a Node.js backend for data processing and APIs, and a React-based admin panel for managing data.

- **Real-Time Updates:** Live tracking of match events including goals, substitutions, and cards.
- **Comprehensive Data Management:** Admin panel for adding teams, scheduling matches, and managing tournament data.
- **User-Friendly Interface:** Mobile application in the Somali language for accessibility.
- **Socket.io Integration:** Real-time updates through WebSocket for immediate event broadcasting.

This personalized platform not only moderates the challenges of existing systems but also raises a stronger connection between football fans and their team.

C. System Design

Here there is level 1 diagram of the system to understand the system interaction, and how the layers of the platform work each other from fans using the application to the database and from the managers using the react admin panel to the database.

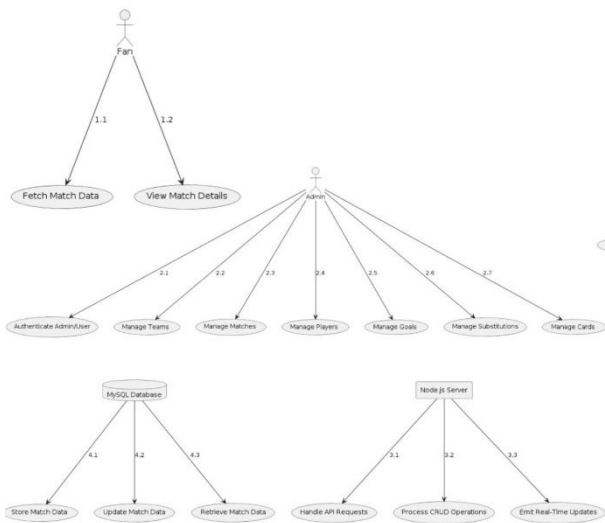


Fig. 1. Use case diagram.

Fig. 1 illustrates the system's use case diagram, which provides a detailed representation of the interactions between various user roles and the system components.

D. Database Design

The database used in this system is MySQL, chosen for its reliability and robustness in handling structured data.

The key tables in the database include Users, Teams, Matches, Players, Tournaments, Goals, Substitutions, and Many more. Each table is designed to store specific information relevant to the system's operation.

The ER Diagram below concludes main tables of the database and their relationship.

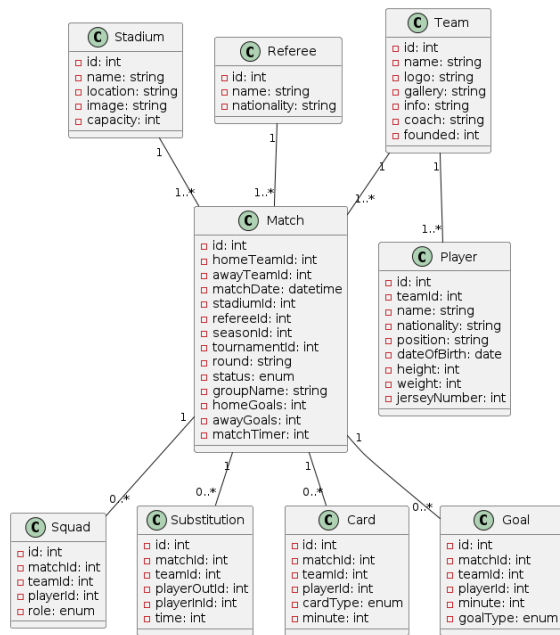


Fig. 2. Entity relationship diagram.

Fig. 2 presents the Entity-Relationship Diagram, showcasing the database structure and relationships among key entities like users, teams, and matches.

V. IMPLEMENTATION AND TESTING

Implemented the system using tools and technologies below.

- Mobile Application (Flutter and Dart) using IDE of Visual Studio Code with Packages like Provider, HTTP, Socket.IO.
- Backend Server (Node.js) using Express.js framework with MySQL database and tools like Socket.IO, JWT for authentication.
- Admin Panel (React).
- Hosting Localhost for development, cloud-based server for production in future.
- Development Tools o Version Control: Git, GitHub.

A. Testing Environment

Testing was conducted in several stages, including integration testing, application testing, and user acceptance testing (UAT). This testing environment was meticulously designed to replicate real-world conditions to ensure the accuracy and reliability of the tests.

B. System Snapshots

1) *Admin panel*: This figure shows that the admin panel is responsible for managing matches, where the management team can perform full CRUD operations. The same applies to other entities such as tournaments, seasons, teams, players, referees, etc.

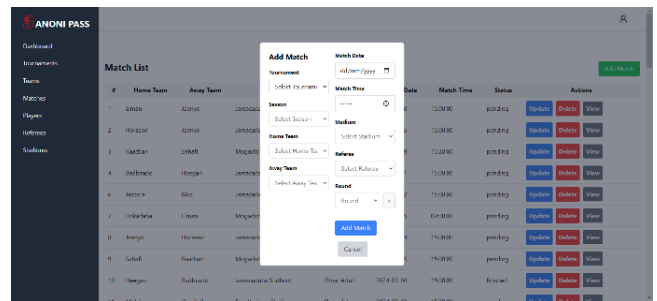


Fig. 3. Managing matches (Create new match).

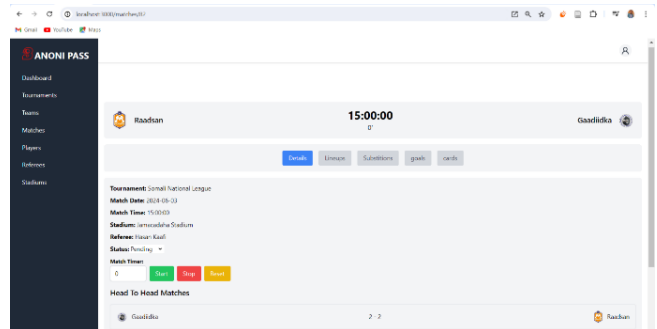


Fig. 4. Match event management.

This screen is responsible for managing match entities like goals, substitutions, cards, squads and also the status and the timer of the match for the platform (see Fig. 3 and Fig. 4).

2) *Mobile application:* This mobile application is designed to provide users with a seamless and engaging experience for tracking football matches and related data (Fig. 5).

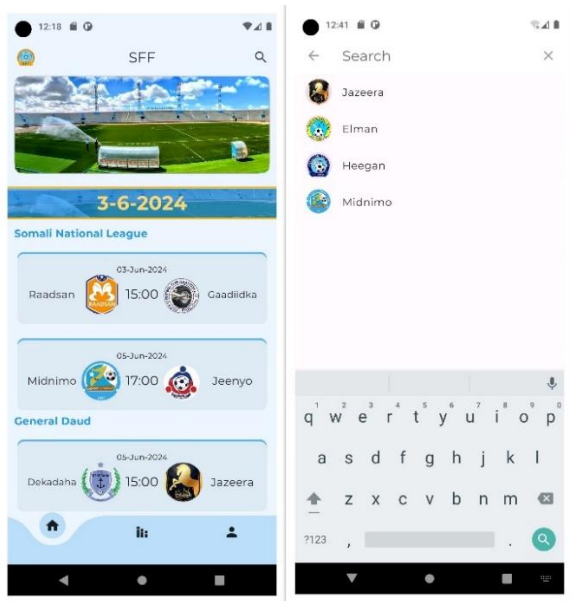


Fig. 5. Home screen (Recent matches).

This home dashboard serves as the main screen of the mobile application, displaying a list of matches organized by date, allowing users to browse upcoming and past matches. Users can select a desired date to filter and view relevant matches. Furthermore, the dashboard provides a search functionality, enabling users to easily search for specific teams and players.

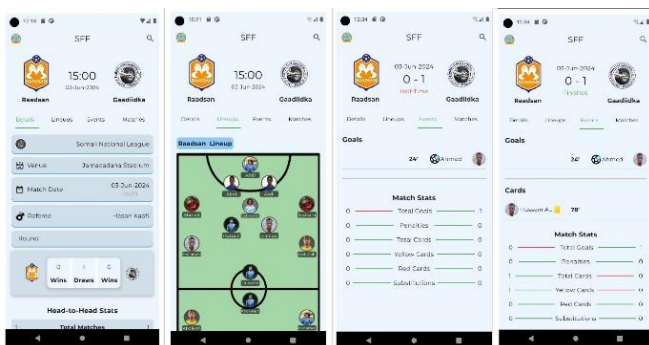


Fig. 6. Match details screen.

The match details (Fig. 6) screen provides a comprehensive overview of ongoing and completed matches. It includes real-time match information such as the teams' lineups (both starting lineup and bench), goals, cards, substitutions, and detailed match statistics that update in real-time. The interface also features head-to-head comparisons between the teams, allowing users to analyze historical encounters, past performance trends, and key statistics. This screen ensures users have all the essential details in one place, offering a dynamic and interactive experience for monitoring matches.

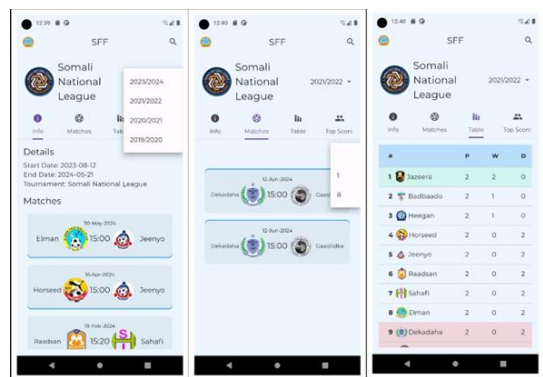


Fig. 7. Tournament details.

The Tournament Details (Fig. 7) Screen, filtered by season, displays a comprehensive list of matches for that particular tournament. As well as, this screen provides detailed tables showcasing the current team standings and top scorers of the tournament. This feature allows users to track team progress and individual player performance throughout the season, offering a clear and organized view of the tournament's key statistics and highlights.

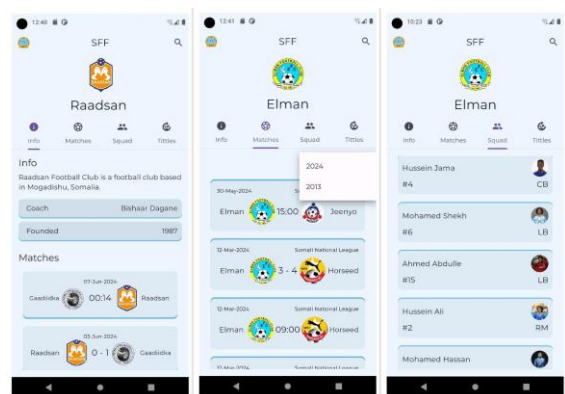


Fig. 8. Team details.

The Team (Fig. 8) Info screen offers an in-depth overview of the selected team, displaying key details such as the team's name, logo, founding year, coach, and other relevant information. The Matches section presents a list of games played by the team, conveniently grouped by year, with a dropdown menu allowing users to filter matches by a specific year. The Squad section showcases a list of players on the team, including each player's image, name, shirt number, and additional relevant information. The Team Titles section highlights the team's achievements and titles, giving users a clear view of the team's historical successes and accolades.

This section detailed the tools and technologies used. Snapshots of the mobile application, admin panel, backend, and database provided a visual representation of the system, illustrating its key features and interfaces.

C. Testing and Validation

Usability tests involved 100 users, including SFF administrators and football fans. Feedback from these sessions improved the navigation of the mobile app and the data-entry workflow of the React admin panel.

This system provides an opportunity to bridge technological gaps in Somali football by fostering transparency and engagement. However, challenges such as intermittent internet connectivity, low digital literacy, and limited funding remain. Solutions include integrating offline functionalities for the admin panel, offering training for administrators, and exploring partnerships with local ISPs to subsidize operational costs.

Validation metrics included:

- Socket.IO latency measured under various network conditions, achieving an average of 1.8 seconds.
- Data accuracy verified by cross-referencing system outputs with manual records during pilot matches.
- Admin panel functionality enabled full CRUD operations with a success rate of 100%.

VI. CONCLUSION AND FUTURE WORK

A. Conclusion

The development of the real-time football match statistics system for Somali Football Federation (SFF) has brought several key accomplishments. The system was developed to overcome the limitations of the existing manual processes and social media-based updates. One of the main achievements is its ability to provide live updates on match events, including the match timer, goals, substitutions, and cards, using Socket.IO technology. The system also covers a wide range of data, such as team and player details, match schedules, statistics, referee and stadium information, and tournament standings, including season-specific details.

Additionally, the system is designed with easy-to-use interfaces. The mobile app, built with Flutter, offers a smooth experience for fans in the Somali language, while the admin panel, developed in React, helps administrators manage data efficiently. Together, these features make the platform simple to use and effective for both users and administrators.

B. Future Work

The successful implementation of the real-time football match statistics system lays the foundation for future enhancements and expansions Including:

- **Analytics Incorporating:** Advanced analytics and machine learning capabilities could provide deeper insights into player performance, team strategies, and match outcomes.
- **Expanded Features:** Introducing new features such as video highlights push notifications for match events, social sharing options, and interactive visualizations of match statistics could further enrich the user experience and engagement with the system.
- **Machine Learning Integration** Integrate machine learning algorithms to provide personalized recommendations for users, such as suggested matches to watch or teams to follow based on their preferences and viewing history.

- **Integration with Other Platforms:** Integrating the system with other popular football platforms and social media channels could increase visibility and reach.

ACKNOWLEDGMENT

We sincerely thank our university and supervisor, for their invaluable guidance, financial assistance, and insightful perspectives throughout this research process.

REFERENCES

- [1] "Somali Football Federation," Jan. 19, 2024. Available: https://en.wikipedia.org/w/index.php?title=Somali_Football_Federation&oldid=1197242818
- [2] H. Elkins et al., "Implementing data analytics for U.Va. Football," in 2017 Systems and Information Engineering Design Symposium (SIEDS), Charlottesville, VA, USA: IEEE, Apr. 2017, pp. 202–207. doi: 10.1109/SIEDS.2017.7937717.
- [3] Y. Qiu et al., "Real-time analysis scheme of football matches based on face recognition and multiple object tracking," in 2023 5th International Conference on Robotics and Computer Vision (ICRCV), Nanjing, China: IEEE, Sep. 2023, pp. 159–163. doi: 10.1109/ICRCV59470.2023.10329030.
- [4] T. V. D. Grün, N. Franke, D. Wolf, N. Witt, and A. Eidloth, "A Real-Time Tracking System for Football Match and Training Analysis," in Microelectronic Systems, A. Heuberger, G. Elst, and R. Hanke, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 199–212. doi: 10.1007/978-3-642-23071-4_19.
- [5] M. Leo, N. Mosca, P. Spagnolo, P. L. Mazzeo, and A. Distanto, "Real-time multi-view event detection in soccer games," in 2008 Second ACM/IEEE International Conference on Distributed Smart Cameras, Palo Alto, CA, USA: IEEE, Sep. 2008, pp. 1–10. doi: 10.1109/ICDSC.2008.4635729.
- [6] J. Greenhough, P. C. Birch, S. C. Chapman, and G. Rowlands, "Football goal distributions and extremal statistics," *Phys. Stat. Mech. Its Appl.*, vol. 316, no. 1–4, pp. 615–624, Dec. 2002, doi: 10.1016/S0378-4371(02)01030-0.
- [7] P. Chaiwuttisak, "Measuring efficiency of Thailand's football premier leagues using data envelopment analysis," 2018.
- [8] E. Wheatcroft, "Forecasting football matches by predicting match statistics," *Journal of Sports Analytics*, vol. 7, no. 2, pp. 77–97, 2021. doi: 10.3233/JSA-200462.
- [9] A. García-Aliaga, M. Marquina, J. Coterón, A. Rodríguez-González, and S. Luengo-Sánchez, "In-game behaviour analysis of football players using machine learning techniques based on player statistics," *International Journal of Sports Science & Coaching*, vol. 16, no. 1, pp. 148–157, 2021. doi: 10.1177/1747954120959762.
- [10] A. Gangal, A. Talnikar, A. Dalvi, V. Zope, and A. Kulkarni, "Analysis and Prediction of Football Statistics using Data Mining Techniques," *International Journal of Computer Applications*, vol. 132, no. 5, pp. 8–11, 2015. doi: 10.5120/ijca2015907263.
- [11] V. C. Pantzalis and C. Tjortjijis, "Sports Analytics for Football League Table and Player Performance Prediction," in 2020 11th International Conference on Information, Intelligence, Systems and Applications (IISA), 2020, pp. 1–8. doi: 10.1109/IISA50023.2020.9284352.
- [12] R. S. Mendes, L. C. Malacarne, and C. Anteneodo, "Statistics of football dynamics," *The European Physical Journal B*, vol. 57, pp. 357–363, 2007. doi: 10.1140/epjb/e2007-00177-4.
- [13] A. Ç. Seçkin, B. Ateş, and M. Seçkin, "Review on Wearable Technology in Sports," *Appl. Sci.*, vol. 13, no. 18, p. 10399, 2023, doi: 10.3390/app131810399.
- [14] J. R. Dmello, "The impact of data analytics on player performance in professional sports: A systematic review," *Int. Res. J. Mod. Eng. Technol. Sci.*, vol. 5, no. 4, pp. 6308, Apr. 2023, doi: 10.56726/IRJMETS37597.
- [15] Y. Qi, S. M. Sajadi, S. Baghaei, R. Rezaei, and W. Li, "Digital technologies in sports: Strategies for safeguarding athlete wellbeing and competitive integrity in the digital era," *Technol. Soc.*, vol. 77, p. 102496, 2024, doi: 10.1016/j.techsoc.2024.102496.

Elevator Abnormal State Detection Based on Vibration Analysis and IF Algorithm

Zhaoxiu Wang

Department of Electronics and Information, Zhangzhou Institute of Technology, Zhangzhou, 363000, China

Abstract—Elevators play a crucial role in daily life, and their safety directly impacts the personal and property safety of users. To detect abnormal states of elevators and ensure people's personal safety, the acceleration signal of elevators is decomposed and Weiszfeld algorithm is used to estimate gravity acceleration. In addition, the study also introduces Kalman filtering to reduce error accumulation. To estimate the operating position of elevators, a method based on information fusion is studied and designed to construct a mapping relationship between elevator vibration energy and position, and to locate the height of elevator faults. Finally, an anomaly detection model combining vibration analysis and the Isolated Forest algorithm is developed. The results showed that the main distribution range of acceleration values in the horizontal direction was between $0.02 \text{ m}^2/\text{s}$ and $-0.02 \text{ m}^2/\text{s}$. The average estimation error and root mean square error of the research designed elevator position estimation method were 0.109 m and 0.113 m , respectively, which could solve the problem of accumulated position errors. The abnormal vibration energy and height corresponding to different operating conditions of elevators were different. The normal value ratios of the anomaly detection model under different sliding windows were 99.91% and 99.57% , respectively. The anomaly detection model designed for research has good performance and can provide technical support for the detection of elevator operation status.

Keywords—Vibration analysis; IF algorithm; elevator; abnormal; detection

I. INTRODUCTION

As a vertical transportation tool, elevators are used in various places in daily life, such as office buildings, residences, hotels, large libraries, and industrial and mining enterprises [1]. Elevators are frequently used in people's daily lives, and once any elevator malfunctions, it can pose a serious threat to people's safety. Therefore, timely and accurate detection of elevator abnormal states is crucial. The commonly used methods for elevator status detection include grey prediction model, genetic algorithm, and particle swarm algorithm [2-3].

Skog I designed a new non-invasive elevator fault detection method and corresponding efficient algorithm for detecting elevator faults. This method modeled the traffic load on the elevator through a non-homogeneous Poisson process and

described the process using a generalized linear model. The results showed that the method achieved an accuracy of 0.82 with a recall probability of 0.80 [4]. Oya J R G et al. designed a system based on time-domain reflectometry technology to detect elevator belt faults, and constructed a receiver based on compressive sensing to improve positioning capability. The results showed that this method could recover time-domain sparse signals and effectively detect elevator belt faults [5]. Ippili S et al. constructed a one-dimensional convolutional neural network based on sound signals for early identification of faults in rotating machinery, and used this network to process the original time signals. The results showed that this method had good performance in the early identification of rotating machinery faults, and its performance was significantly better than the accelerometer based method [6]. Mian T et al. designed a multi-sensor fault diagnosis system based on infrared thermal imaging and vibration sensors for diagnosing faults in rotating machines. In addition, the study also utilized deep convolutional neural networks, support vector machines, and principal component analysis. The results showed that this method could effectively diagnose faults in rotating machines under all working conditions [7].

However, these methods also have certain issues, such as the stronger dependence of model-based methods on specific parameters of elevator systems compared to data-driven methods on data training. To detect abnormal states in elevators, an anomaly detection model based on vibration analysis of horizontal vibration signals and Isolation Forest (IF) algorithm is designed. Methods are also designed to reduce the accumulation of position errors and estimate elevator dynamic characteristics. The research aims to improve the accuracy of elevator anomaly detection, avoid serious elevator accidents, and ensure people's personal safety. The innovation of the research is reflected in the combination of vibration analysis and IF algorithm, which reduces the cumulative position error and improves the efficiency of elevator anomaly detection. To better demonstrate the advantages of the design method proposed in this article, the study will compare it with existing methods in terms of performance indicators, scalability, and limitations. The comparison results are shown in Table I.

TABLE I. COMPARISON WITH RELATED WORK

Serial number	Performance index		Scalability	Limitation
	F1	Accuracy		
[4]	0.800	0.820	F1 drops to 0.685 on larger datasets	Model complexity and difficulty in parameter estimation
[5]	0.847	0.886	When facing multiple types of elevator belts, a significant amount of customized training is required	Limited measurement accuracy and signal quality issues
[6]	0.929	0.936	Cannot be compatible with new data types	Insufficient utilization of sequence order information
[7]	0.935	0.941	Difficulty in expanding sensor types	Easy overfitting and high computational resource consumption
Manuscript	0.988	0.992	Low cost, strong universality and scalability	Not much consideration has been given to the fault detection of elevator door systems

II. METHODS AND MATERIALS

To detect abnormal states in elevators, an anomaly detection method based on vibration analysis and IF algorithm is studied and designed. Due to the use of horizontal acceleration signals for anomaly detection, the study also designs a method for estimating elevator gravity acceleration and dynamic characteristics. In addition, the study also designs a method for estimating the operating position of elevators.

A. Design of Estimation Method for Elevator Gravity Acceleration and Dynamic Characteristics

To detect abnormal states in elevators, a gravity acceleration and dynamic feature estimation method is first designed to reduce error accumulation and improve detection accuracy. Secondly, a method for estimating the operating position of elevators is studied and designed to facilitate the construction of the mapping relationship between elevator vibration energy and position in the future. Finally, an anomaly detection method based on vibration analysis and IF algorithm is studied and designed. The study uses a three-axis acceleration sensor (MPU6050) to collect three-dimensional acceleration signals from the elevator, and decomposes the acceleration three-dimensional vector based on the gravity acceleration vector calibrated by the sensor. After that, the components in the direction of gravity acceleration can be obtained. The Weiszfeld algorithm is adopted for the estimation of elevator gravity acceleration. The Weiszfeld algorithm is a classic iterative algorithm for solving single facility site selection problems, which can obtain the optimal solution of the problem, and the essence of this algorithm is a steepest descent method. The Weiszfeld algorithm, as a repeated weighted least squares method, has the advantage of being able to handle weighted point sets and gradually converge to the optimal solution during the iteration process [8-9]. In addition, unlike some methods that rely on specific prior knowledge or assumptions, the Weiszfeld algorithm does not require extensive knowledge of the elevator's operating status, system parameters, etc. when estimating elevator gravity acceleration, and has a wider applicability [10-11]. The solution for this algorithm is shown in Eq. (1).

$$\begin{cases} x_{k+1} = \frac{\sum_{i=1}^a \frac{b_i}{E_i(x_k, y_k, z_k)}}{\sum_{i=1}^a \frac{1}{E_i(x_k, y_k, z_k)}} \\ y_{k+1} = \frac{\sum_{i=1}^a \frac{c_i}{E_i(x_k, y_k, z_k)}}{\sum_{i=1}^a \frac{1}{E_i(x_k, y_k, z_k)}} \\ z_{k+1} = \frac{\sum_{i=1}^a \frac{d_i}{E_i(x_k, y_k, z_k)}}{\sum_{i=1}^a \frac{1}{E_i(x_k, y_k, z_k)}} \end{cases} \quad (1)$$

In Eq. (1), (x_k, y_k, z_k) represents the median center solved by Weiszfeld algorithm after the k th iteration, and x , y , and z are vectors on the x-axis, y-axis, and z-axis, respectively. a represents the number of measured gravitational acceleration vectors, and (b_i, c_i, d_i) is the gravitational acceleration vector obtained from the i measurement. $E_i(x_k, y_k, z_k)$ represents the distance between (b_i, c_i, d_i) and the median center obtained from the k iteration. $(x_{k+1}, y_{k+1}, z_{k+1})$ represents the new median center. The calculation of $E_i(x_k, y_k, z_k)$ is shown in Eq. (2).

$$E_i(x_k, y_k, z_k) = \|(b_i, c_i, d_i) - (x_k, y_k, z_k)\| \quad (2)$$

The operation process of an elevator can be mainly divided into four states: stationary, accelerating, uniform, and decelerating, and the acceleration in all four ideal states remains constant [12]. To estimate the kinematic characteristics of elevators, the Kalman filtering method is used in the study. The advantage of the Kalman filtering method is that it can update the state estimation based on previous estimates and current measurements, and has strong robustness and adaptability [13-14]. The representation of elevator dynamic characteristics is shown in Eq. (3).

$$\chi_k = F_k \chi_{k-1} + \omega_k \quad (3)$$

In Eq. (3), χ_k represents the system state vector, F_k represents the state transition function, and ω_k represents the process noise vector. The expression of χ_k is shown in Eq. (4).

$$\chi_k \triangleq \begin{bmatrix} \chi'_k \\ \chi''_k \\ \chi'''_k \end{bmatrix} \quad (4)$$

In Eq. (4), \triangleq represents the identity equation, while χ'_k , χ''_k , and χ'''_k represent velocity, acceleration, and jerk, respectively. The expression of F_k is shown in Eq. (5).

$$F_k \triangleq \begin{bmatrix} 1 & g & \frac{g^2}{2} \\ 1 & g & \\ & 1 & g \\ & & 1 \end{bmatrix} \quad (5)$$

In Eq. (5), g represents the sampling time interval. To output the system state vector, the Kalman filtering method

needs to perform optimal estimation based on the prediction step and update step. Optimization estimation mainly includes four steps. The first step is to solve g and adjust F_k and process noise covariance Q_k based on g . The second step is to predict the state of the next time step and estimate the system covariance \bar{P}_k . The third step is to solve the Kalman gain K_k , and the fourth step is to solve the novel Y_k . The fifth step is to adjust the state prediction value and prediction covariance, and refresh the system state. The processing of the covariance matrix P_k is shown in Eq. (6).

$$P_k = (I - K_k H_k) \times \bar{P}_k \times (I - K_k H_k)^T + K_k R_k \times K_k^T \quad (6)$$

In Eq. (6), I is the identity matrix, H_k represents the measurement function, and R_k represents the observed noise variance. Since the measured values of acceleration sensors during elevator operations are typically non-zero, threshold values and corresponding constraints are introduced in the study. The automatic correction process of gravity acceleration is shown in Fig. 1.

From Fig. 1, the automatic correction process of gravity acceleration mainly consists of six steps. The first step is to determine whether the elevator is stationary. If it is stationary, slide the window to collect data, otherwise the process ends. The second step is to filter out outliers, and the third step is to use the Weiszfeld algorithm. The fourth step is to update the gravitational acceleration, and the fifth step is to determine whether the elevator is moving. If it is running, the process ends; otherwise, the sliding window continues to collect data.

B. Design of Elevator Operation Position Estimation Method

To locate the operating position of the elevator, the study first models the elevator floor information through acceleration sensors and Simultaneous Localization and Mapping (SLAM) algorithm. Secondly, the study uses a pressure sensor to solve the operating height of the elevator. Finally, a method for estimating the operating position of elevators based on information fusion is studied and designed. The elevator displacement solved by acceleration sensors has the advantage of high short-term accuracy, but there is also a drawback of fast error accumulation. Therefore, the SLAM algorithm is introduced to compensate for this deficiency. The advantage of the SLAM algorithm is its ability to integrate multiple sources

of information and incorporate Kalman filtering technology [15]. By using the Kalman filter in the SLAM algorithm, it is possible to obtain the floor spacing and floor height. The expression for the distance of elevator operation between two stops is shown in Eq. (7).

$$S_k = F_k S_{k-1} + B_k u_k + \omega_k \quad (7)$$

In Eq. (7), u_k is the control vector, B_k is the control matrix, and S_k represents the displacement of the elevator from rest. The expression of B_k is shown in Eq. (8).

$$B_k = \begin{pmatrix} g \\ \frac{g^2}{2} \end{pmatrix} \quad (8)$$

The initialization process of elevator floor information mainly consists of seven steps, including acceleration data collection, elevator motion judgment, calculation of elevator displacement and displacement error, judgment of whether the elevator is going up, judgment of whether the elevator is stationary, updating map information, and judgment of whether the set number of times has been reached. To solve the operating height of the elevator, a pressure sensor is used in the study. The advantage of air pressure sensors is that they can directly obtain real-time altitude information of elevators, and the error accumulation is slow [16-17]. The difference γ between the starting and measuring heights is solved as shown in Eq. (9).

$$\gamma = \left(\frac{JL}{WN} \right) \ln \left(\frac{\rho_0}{\rho} \right) \quad (9)$$

In Eq. (9), ρ_0 and ρ represent the atmospheric pressure at the starting and measuring heights, respectively, and N represents the molar mass of air. J represents the universal gas constant, L represents the measured air temperature, and W represents the gravitational acceleration of the Earth's surface. The solution for the current altitude ϕ is shown in Eq. (10) [18-19].

$$\phi = 44330 \times \left(1 - \left(\frac{\rho}{\rho_0} \right)^{\frac{1}{5.255}} \right) \quad (10)$$

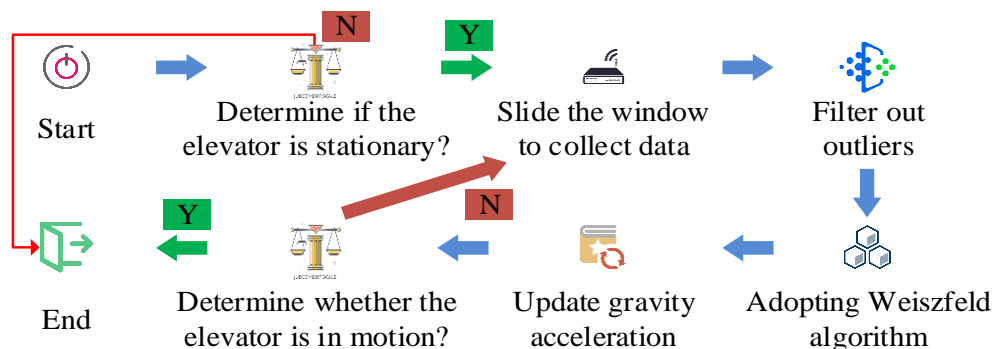


Fig. 1. Automatic correction process of gravity acceleration.

The solution of γ can be simplified as shown in Eq. (11).

$$\gamma = \phi - \phi_0 \quad (11)$$

In Eq. (11), ϕ_0 represents the altitude of the reference position. To preprocess the obtained height data and reduce the impact of noise, a first-order exponential smoothing method is used in the study. The advantage of this method is that it makes extrapolation predictions more realistic and has high practicality and effectiveness in time series prediction [20]. The current time step estimation value η_t is solved as shown in Eq. (12).

$$\eta_t = \beta\sigma_t + (1 - \beta)\eta_{t-1} \quad (12)$$

In Eq. (12), η_{t-1} represents the estimated value of the previous time step, β represents the smoothing coefficient, and $\beta \in (0, 1)$ and σ_t are the measured values of the current time step. t stands for Time. To estimate the operating position of the elevator, a combination of acceleration sensors and air pressure sensors is studied, and floor information is also introduced. Unscented Kalman Filter (UKF) is used to apply this information. Therefore, the solution for the operating height θ_τ of the elevator at τ time is shown in Eq. (13).

$$\theta_\tau = \theta_{\tau-1} + B_k u_k + \omega_k \quad (13)$$

The state transition function of the elevator system is expressed as Eq. (14).

$$\theta_\tau = f(\theta_{\tau-1}) = \begin{cases} \theta_{\tau-1} + B_k u_k + \omega_k & \text{Motion} \\ f\theta_j, \theta_{\tau-1} \in U(f\theta_j, \delta_{f\theta_j}) & \text{Static} \end{cases} \quad (14)$$

In Eq. (14), $f\theta_j$ represents the height of the car relative to the reference point, j is the number of heights, U is the symbol for the set, and $\delta_{f\theta_j}$ represents the prior measurement error. The core of UKF is the traceless transformation, as shown in Fig. 2.

In Fig. 2, \mathcal{G} represents a set of sigma points, and ψ represents a new set of points after nonlinear changes. The green and orange dots in the blue background represent the mean and covariance of the transformation point set, respectively, and are considered as new predicted values. To select sigma points, a symmetric sampling strategy is adopted in the study. The elevator position tracking process based on UKF is shown in Fig. 3.

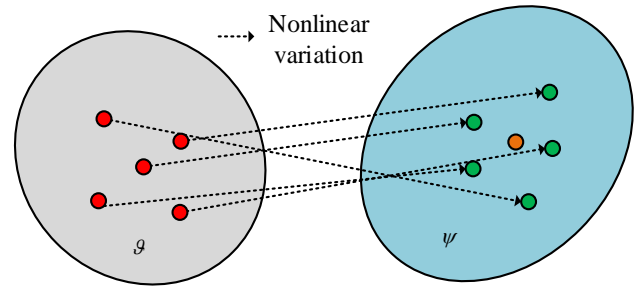


Fig. 2. Diagram of unscented transformation.

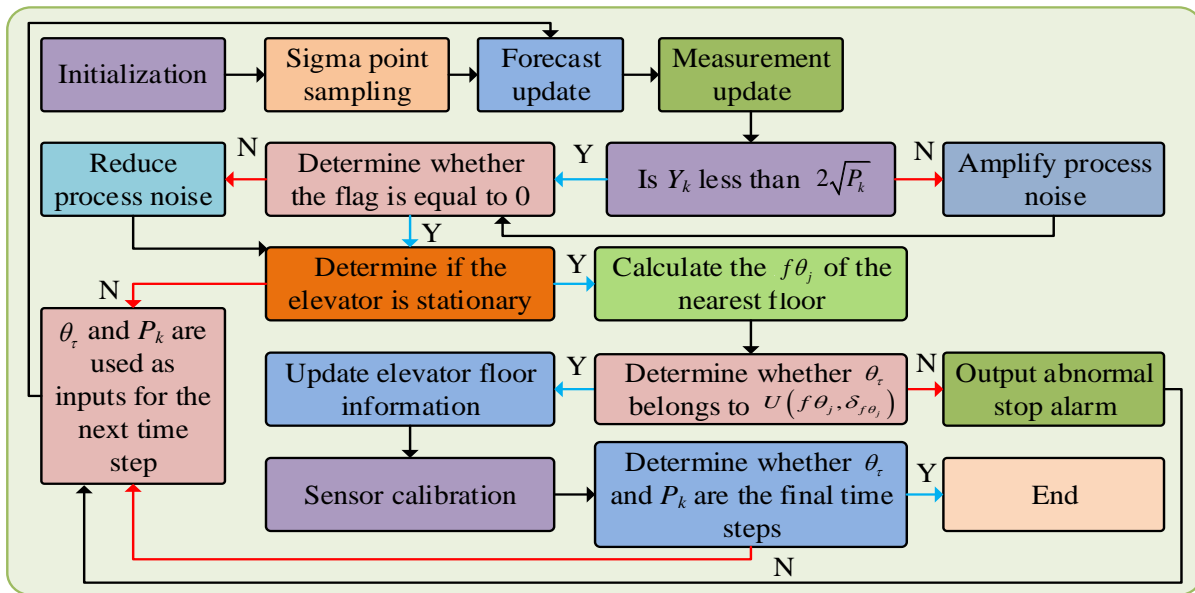


Fig. 3. Elevator position tracking process based on UKF.

From Fig. 3, the first step of the elevator position tracking process based on UKF is initialization, and the second step is sigma point sampling. The third step is to predict updates, and the fourth step is to measure updates. The fifth step is to determine whether Y_k is smaller than $2\sqrt{P_k}$. If it is less than,

proceed to step six; otherwise, amplify the process noise before proceeding to step six. The sixth step is to determine whether the flag is equal to 0. If it is equal, proceed to step seven; otherwise, reduce the process noise before proceeding to step seven. The seventh step is to determine whether the elevator is

stationary. If it is stationary, calculate the $f\theta_j$ of the nearest floor. Otherwise, use it as input for the next time step of θ_τ and P_k . The eighth step is to determine whether θ_τ belongs to $U(f\theta_j, \delta_{f\theta_j})$. If it belongs, update the elevator floor information. Otherwise, output an abnormal stop alarm. The ninth step is to perform sensor calibration, and the tenth step is to determine whether θ_τ and P_k are the final time steps. If so, the process ends. Otherwise, θ_τ and P_k are inputted for the next time step, and then return to the second step.

C. Design of Elevator Abnormal State Detection Method

To detect abnormal states in elevators, the study first constructs a mapping of elevator vibration energy and position based on vibration analysis. Secondly, the study uses the IF algorithm to train the elevator anomaly detection model. To determine the operating status of the elevator, the study considers the vibration of the elevator car as an important feature. To determine whether the data is abnormal, a baseline is constructed for elevator normal operation. In addition, the study uses horizontal acceleration signals to detect abnormal vibrations. The steps of the baseline generation method are shown in Fig. 4.

From Fig. 4, the first step in baseline generation is data collection, and the second step is to determine whether the baseline has been generated. If it has been generated, end the process; otherwise, perform data preprocessing and feature extraction. The third step is to obtain horizontal vibration energy, and the fourth step is to obtain cluster data. The fifth step is to calculate the moving average, and the sixth step is to calculate the moving average error. The seventh step is to determine whether the slope of the moving average error is approximately equal to 0. If it is, a baseline is generated and the process ends. Otherwise, the process returns to the first step.

The solution for the moving average MA_n is shown in Eq. (15) [21].

$$MA_n = \frac{\sum_{r=1}^n |\Pi_r|^2}{n} \quad (15)$$

In Eq. (15), n represents the number of collected signals, Π_r represents the root mean square of the r horizontal vibration signal, and r is the signal number. To clarify the abnormal state of the elevator, the intrinsic feature scale decomposition method is used to decompose the horizontal vibration signal, and envelope spectrum analysis is used to detect the impact signal to eliminate false alarms. To suppress the endpoint effect, the method of mirror extension is used in the study. To construct a mapping between elevator vibration energy and position to detect the vibration state of the guide rail on the highest and lowest floors, a horizontal acceleration signal is used and solved using the established elevator acceleration and position estimation method. To further detect outliers, the study adopts the IF algorithm and trains the outlier detection model through the IF algorithm. The IF algorithm, as an unsupervised method, has the advantages of low computational cost, linear time complexity, and does not rely on abnormal samples [22]. In addition, compared with similar outlier detection methods, the IF algorithm does not require calculating the distance or density between data points like some distance or density-based methods, has lower time complexity, and does not rely on data distribution assumptions. It is relatively insensitive to noise and outliers in the data, and can effectively identify true outliers, making it less susceptible to noise interference and misjudgment [23-24]. The IF algorithm uses the idea of ensemble learning and requires the construction of isolated trees. The construction process of an isolated tree is shown in Fig. 5.

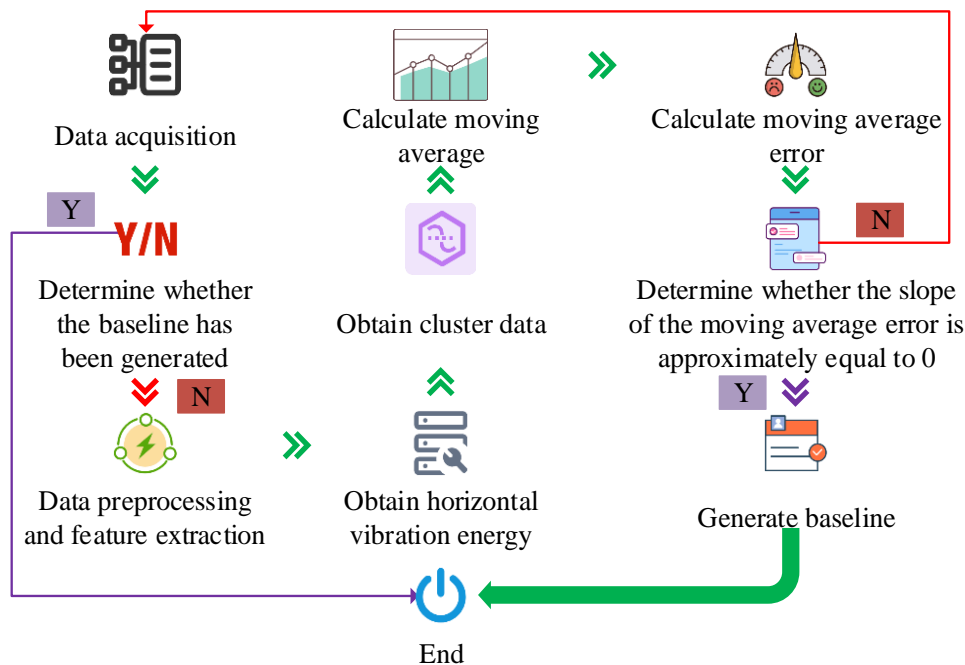


Fig. 4. Steps of baseline generation method.

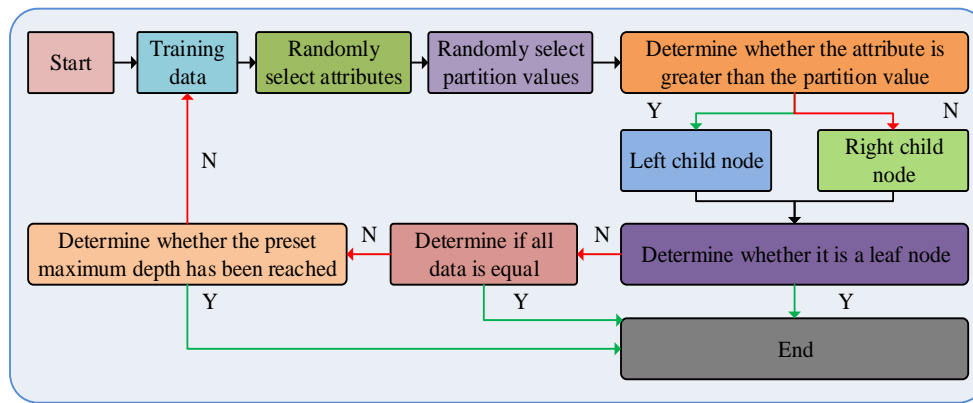


Fig. 5. The construction process of isolated trees.

From Fig. 5, the first step in constructing an isolated tree is to train the data, and the second step is to randomly select the attribute ε . The third step is to randomly select the partition value μ , and the fourth step is to determine whether ε is greater than μ . If it is judged as yes, place it in the left child node; otherwise, place it in the right child node. The fifth step is to determine whether it is a leaf node. If it is, the process ends; otherwise, the next step is to proceed. The sixth step is to determine whether all data are equal. If they are equal, the process ends; otherwise, it enters the seventh step. The seventh step is to determine whether the preset maximum depth has been reached. If it has been reached, the process ends. Otherwise, it returns to the first step and repeats the process until it ends. The process of elevator anomaly detection method based on vibration analysis and IF algorithm is shown in Fig. 6.

From Fig. 6, the first step of the elevator anomaly detection method is to collect acceleration signals, and the second step is to estimate the elevator position. The third step is to establish a baseline, and the fourth step is envelope spectrum analysis. The fifth step is to generate data records, and the sixth step is to determine whether the sliding window is full. If it is full, the detection model is used for anomaly detection. Otherwise, the process returns to the fifth step. The seventh step is to determine if the anomaly rate is too high. If it is too high, an alarm will be triggered and the process will end. Otherwise, the data will be added to the buffer. The eighth step is to determine whether the number of records is greater than the set threshold. If it is, the detection model will be retrained based on the training data. Otherwise, it will return to the data buffer.

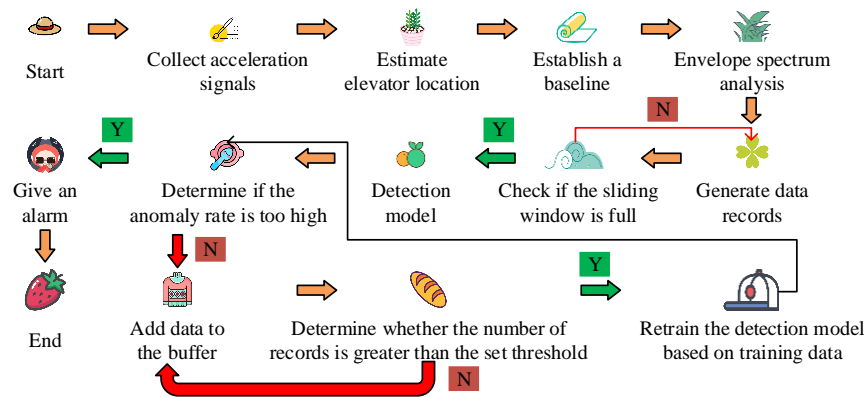


Fig. 6. The process of elevator anomaly detection method based on vibration analysis and IF algorithm.

III. RESULTS

To analyze the detection results of elevator abnormal states, the study explained the experimental data collection equipment, experimental environment, and other experimental devices, and analyzed the results of acceleration information decomposition. Afterwards, the study analyzed the results of elevator position tracking and constructed the relationship between elevator vibration energy and position mapping. Finally, the study validated the performance of the IF model in detecting elevator abnormal states.

A. Acceleration Information Decomposition and Motion Feature Estimation Results

To collect elevator data, the study used Raspberry Pi 4B and added MPU6050 acceleration sensor, BMP180 air pressure sensor, and touch switch. The experimental environment used was the elevator in the experimental building. The operating system used in the experiment was Raspbian, the built-in operating system of Raspberry Pi. The processor was Broadcom BCM2711, with a clock speed of 1.5GHz, a maximum supported memory capacity of 4GB, and a thermal design power consumption of 7.5W. The collected signal would be decomposed, and the decomposition result is shown in Fig.

7.

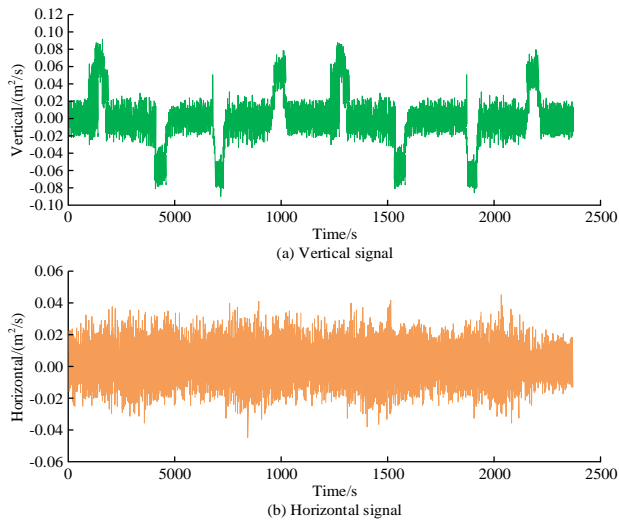


Fig. 7. The result of signal decomposition.

In information decomposition, the study first collected the residual distribution of the acceleration signal minus the calibrated gravity acceleration in three-dimensional space. Secondly, the distribution was decomposed into vertical and horizontal acceleration signals in three-dimensional space, as shown in Fig. 7 (a) and Fig. 7 (b). According to Fig. 7 (a), after decomposing the acceleration information, the maximum and minimum vertical acceleration values were $0.09375 \text{ m}^2/\text{s}$ and $-0.09063 \text{ m}^2/\text{s}$, respectively. As time increased, the vertical acceleration value exhibited a fluctuating trend of gentle, upward, and downward fluctuations. According to Fig. 7 (b), in the horizontal direction, the maximum acceleration value was $0.0450 \text{ m}^2/\text{s}$ and the minimum value was $-0.0457 \text{ m}^2/\text{s}$. As time gradually increased, the horizontal acceleration value exhibited a fluctuating upward and downward trend. In addition, the main distribution range of acceleration values in the horizontal direction was between $0.02 \text{ m}^2/\text{s}$ and $-0.02 \text{ m}^2/\text{s}$. Through signal decomposition, three-dimensional data can be transformed into one-dimensional data. This not only facilitates the installation of sensors, but also enhances the robustness of the system. The analysis of motion characteristics is shown in Fig. 8.

According to Fig. 8 (a), in terms of acceleration characteristics, the maximum and minimum values of the source data were $1.0 \text{ m}^2/\text{s}$ and $-0.75 \text{ m}^2/\text{s}$, respectively. The maximum and minimum values of the filtered data's maximum acceleration were $0.71 \text{ m}^2/\text{s}$ and $-0.72 \text{ m}^2/\text{s}$, respectively. The maximum and minimum values of the maximum deceleration were $0.76 \text{ m}^2/\text{s}$ and $-0.75 \text{ m}^2/\text{s}$, respectively. From Fig. 8 (b), at 95% of the sampled data, the maximum values of acceleration and deceleration were $0.75 \text{ m}^2/\text{s}$ and $0.68 \text{ m}^2/\text{s}$, respectively, with a difference of $0.07 \text{ m}^2/\text{s}$ between the two. In addition, the minimum acceleration value was $-0.74 \text{ m}^2/\text{s}$, which differed from the minimum deceleration value of $-0.59 \text{ m}^2/\text{s}$ by $0.15 \text{ m}^2/\text{s}$. In Fig. 8 (c), with the increase of time, the elevator speed exhibited a cyclic upward and downward trend. In addition, the maximum and minimum values of the velocity characteristics were $1.81 \text{ m}^2/\text{s}$ and $-1.89 \text{ m}^2/\text{s}$, respectively. Both the acceleration and deceleration, as well as the maximum acceleration and deceleration, were within the standard range.

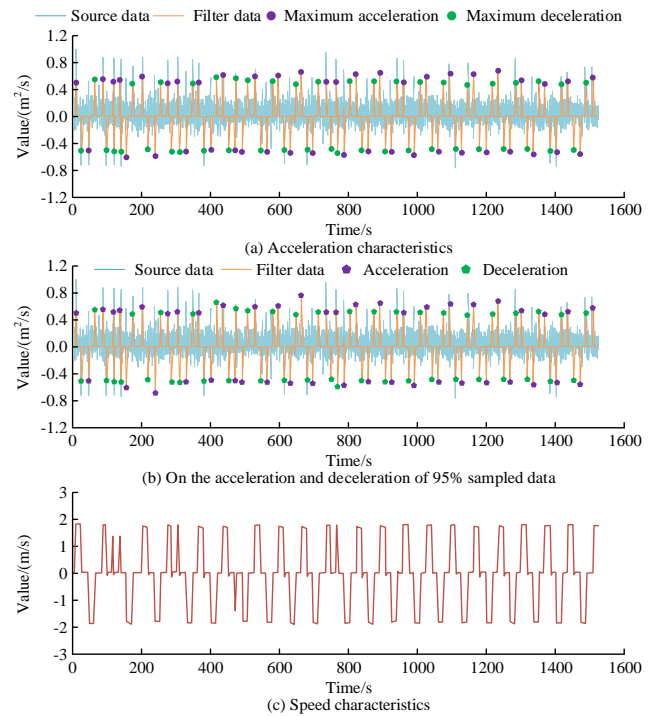


Fig. 8. Analysis of motion characteristics.

B. Analysis of Elevator Position Tracking Results

To track the position of the elevator, the study used the same Raspberry Pi and sensors to collect elevator data. The operating position of the elevator in the experimental building was a total of 8 floors, and the floor height and spacing were generally around 4 meters. The scanning range of SLAM was 0.1-30 meters, with a measurement accuracy of ± 2 centimeters, a time interval of 50Hz, and a positioning accuracy of 10 centimeters. The measurement matrix of UKF was 1, the sampling time interval was 0.5s, the process noise covariance was 1, and the measurement noise covariance matrix was 0.01. Based on the collected information from acceleration sensors and air pressure sensors, the study combined these two types of information through UKF and obtained the tracking results and estimation errors of the elevator position, as shown in Figure 9.

From Fig. 9 (a), the maximum value of elevator position using only the UKF method was 31.3m, and the minimum value was -0.8m . The maximum values of elevator position using UKF+automatic calibration method and UKF+ automatic calibration+SLAM method were 29.3m and 28.5m, respectively, with a difference of 0.8m between the two, and the minimum value of elevator position under both methods was 0m. The automatic calibration method and SLAM effectively reduced the errors observed in the UKF method. In Fig. 9 (b), the cumulative probability of errors at different positions varied under different methods. For example, when the position error was 0.5m, the cumulative probabilities of the UKF method, UKF+automatic calibration method, and UKF+automatic calibration+SLAM method were 0.327, 0.684, and 1.00, respectively. In addition, the average estimation errors of the three methods were 0.923m, 0.395m, and 0.109m, respectively, and the root mean square errors were 0.943m, 0.404m, and 0.113m, respectively. The UKF+automatic calibration+SLAM

method had the smallest estimation error, followed by the UKF+ automatic calibration method. In summary, the UKF+automatic calibration+SLAM method can effectively

solve the problem of accumulated position errors and accurately track the operating position of elevators.

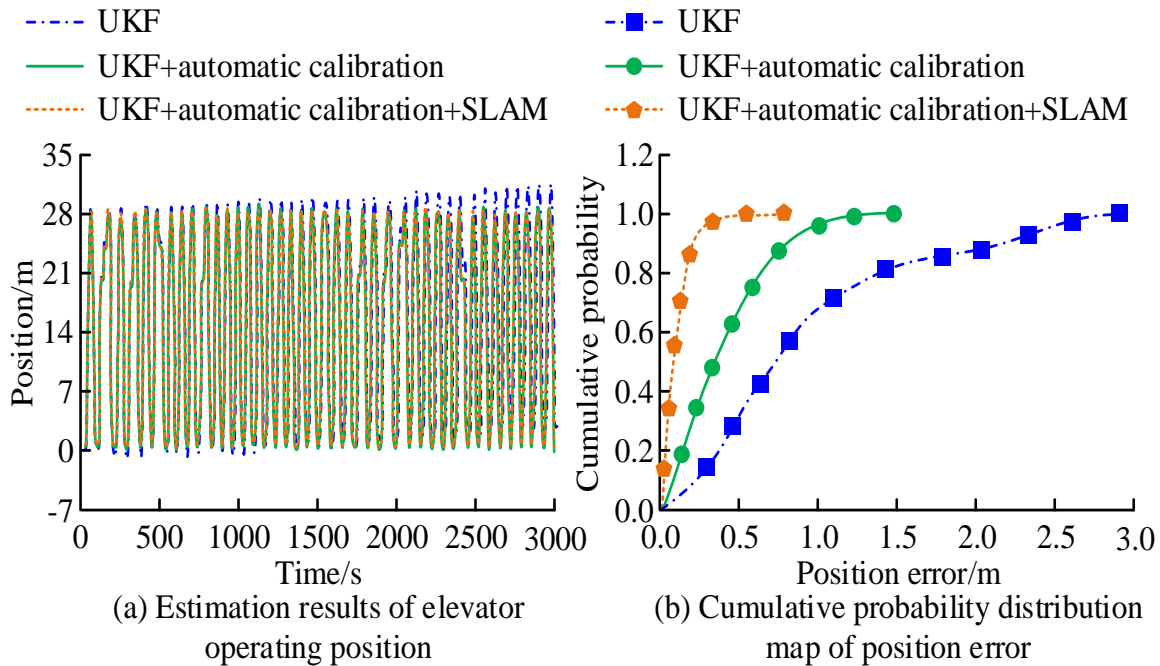


Fig. 9. Tracking results and estimation errors of elevator position.

C. Analysis of Elevator Abnormal State Detection Results

To detect abnormal states in elevators, the same experimental environment and setup were used in the study. In collecting data, the experiment also used Raspberry Pi and sensors. In the collected baseline data, the sub healthy line was 0.140m/s^2 and the fault line was 0.162m/s^2 . In addition, the study also removed useless attributes of the data, such as running time, and only retained the maximum acceleration/deceleration, maximum speed, 95th percentile of

acceleration, and 5th percentile of deceleration. The total number of isolated trees was 120, the threshold for outliers was -0.18 , the maximum sampling number was 267, and the proportion of outliers in the training data was 99.69%. The width of the sliding window was 800, and the study selected the outlier distribution of the third and ninth windows for analysis. The envelope spectrum analysis was conducted on the first product component of the vibration signal, and the amplitude before and after analysis is shown in Fig. 10.

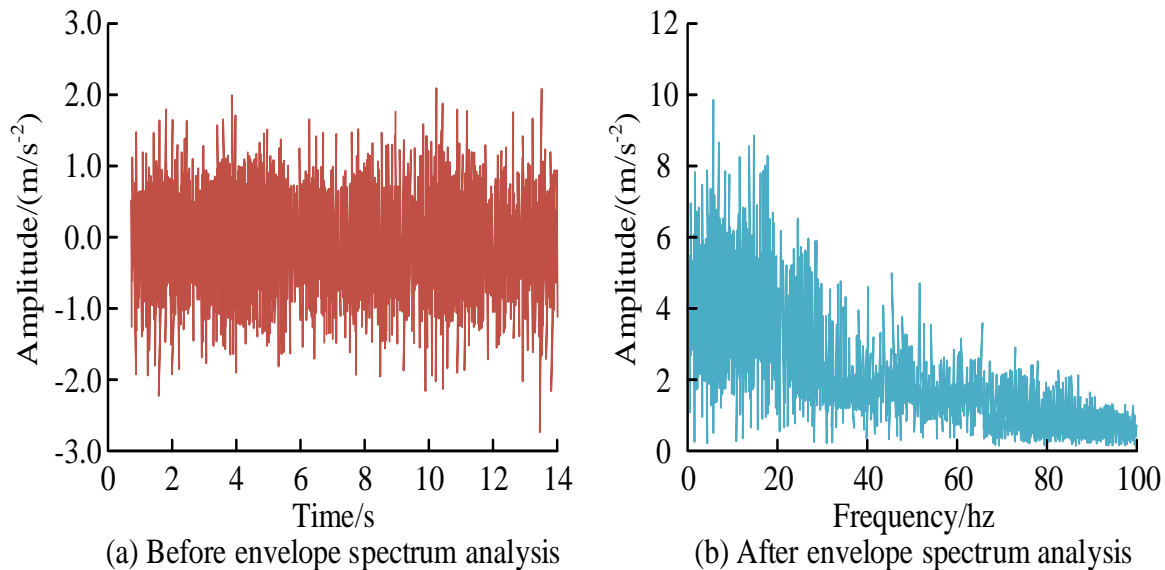


Fig. 10. The amplitude before and after the envelope spectrum analysis of the first product component.

According to Fig. 10 (a), before demodulating the product component 1, its corresponding amplitude was mainly concentrated between 0.5m/s^2 and -0.7m/s^2 , and the maximum and minimum amplitudes were 0.208m/s^2 and -0.273m/s^2 , respectively. As time went by, the vibration kept rising, falling, and fluctuating repeatedly. From Fig. 10 (b), after demodulating the product component 1, as the frequency increased, the vibration gradually decreased, and the range of up and down fluctuations also narrowed. The maximum and minimum amplitude values were 9.98m/s^2 and 0.00m/s^2 , respectively. In addition, the vibration energy of the signal was primarily concentrated in the low-frequency range, with fewer high-frequency impulsive signals. This indicated that the outlier was false and consistent with the actual situation. The relationship between elevator vibration energy and position mapping is shown in Fig. 11.

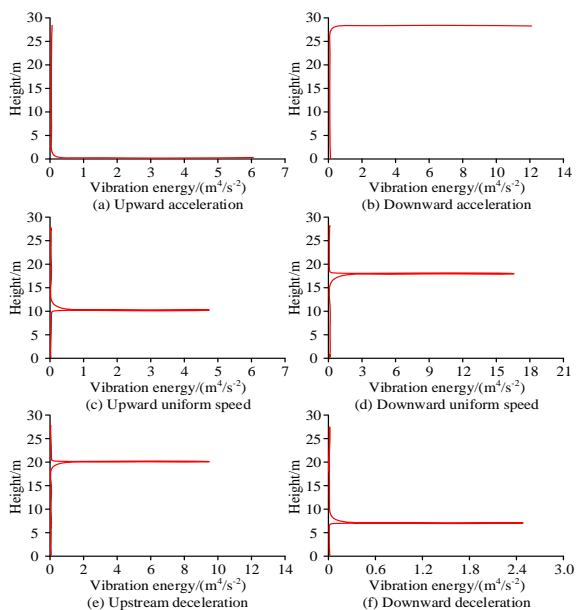


Fig. 11. The relationship between elevator vibration energy and position mapping.

In Fig. 11, the vertical axis represents the operating position of the elevator, and the horizontal axis represents the vibration energy at the corresponding position. The peaks in the spectrum are the abnormal vibrations generated by simulated faults. From Fig. 11 (a), when the elevator accelerated upwards, the abnormal vibration energy generated during the simulated fault was 6.02m/s^2 , and the corresponding height at this time was 0m . From Fig. 11 (b), 11 (c), 11 (d), 11 (e), and 11 (f), peaks appeared during the elevator's downward acceleration, upward uniform speed, downward uniform speed, upward deceleration, and downward deceleration. This indicated that abnormal vibrations occurred in the elevator in all five cases, and the vibration energy and fault height varied in different situations. For example, if the elevator was moving at a constant speed, the abnormal vibration energy corresponding to the up and down directions of the elevator was 4.78m/s^2 and 16.8m/s^2 , respectively, and the corresponding fault heights were 10m and 18m , respectively. It can be seen that the occurrence of simulated faults could be clearly detected at various stages of elevator operation, and the position of the car where the fault occurred, that is, the position where the guide rail may have malfunctioned, could be located, providing important information for the maintenance of the elevator system and the rescue of trapped personnel. The distribution of outliers for different sliding windows is shown in Fig. 12.

In Fig. 12 (a), under the third sliding window, the abnormal scores were mainly concentrated in the range of 0.10 and 0.15 . In addition, the proportion of normal values greater than the threshold of -0.18 outliers was approximately 99.91% . According to Fig. 12 (b), in the 9th sliding window, the proportion of normal values greater than the outlier threshold was about 99.57% , and the outlier scored with more than 10 data points were mainly concentrated in the range of 0.067 to 0.15 . In addition, the maximum number of outlier data points was 58, corresponding to outlier scores of 0.123 and 0.147 . Overall, the IF model could effectively detect the operational status of elevators.

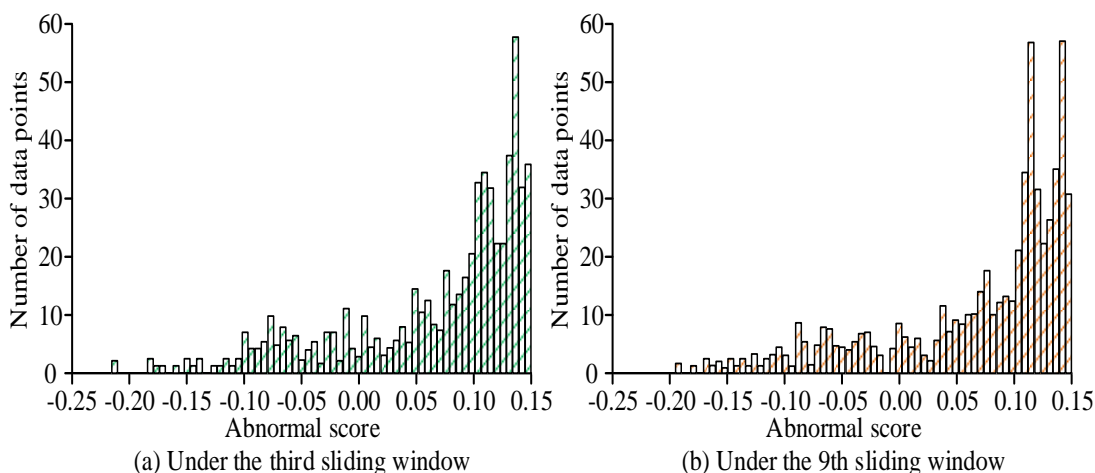


Fig. 12. Outlier distribution of different sliding windows.

IV. DISCUSSION AND CONCLUSION

Aiming at the problem of detecting abnormal states in elevators, a dynamic prediction method for elevators and an information fusion-based elevator operation position tracking method were studied and designed. An anomaly detection model based on vibration analysis and IF algorithm was also constructed. The results showed that after signal decomposition, the maximum acceleration values corresponding to the vertical and horizontal directions were $0.09375 \text{ m}^2/\text{s}$ and $0.0450 \text{ m}^2/\text{s}$, respectively, and the minimum acceleration values were $-0.09063 \text{ m}^2/\text{s}$ and $-0.0457 \text{ m}^2/\text{s}$, respectively. Signal decomposition could transform three-dimensional data into one-dimensional data, enhancing the robustness of the system. The average estimation errors of UKF method, UKF+automatic calibration method, and UKF+automatic calibration+SLAM method were 0.923m , 0.395m , and 0.109m , respectively, with root mean square errors of 0.943m , 0.404m , and 0.113m , respectively. This indicated that both automatic calibration and SLAM algorithms could to some extent solve the problem of accumulated position errors. The maximum amplitude values of product component 1 before and after demodulation were $0.208\text{m}/\text{s}^2$ and $9.98\text{m}/\text{s}^2$, respectively, and the vibration energy of the demodulated signal was mainly concentrated in the low-frequency range, with less high-frequency impulsive signals, which was in line with the actual situation. Under different operating conditions of the elevator, simulated faults had corresponding peak signals, and the vibration energy and height corresponding to the peak signals were also different under different operating conditions. In the third and fifth sliding windows, the proportion of normal values greater than the outlier threshold was 99.91% and 99.57% , respectively. The research designed anomaly detection models had good performance. This method could monitor and analyze the acceleration signals and vibration data of the equipment in real time, predict potential faults, and thus improve safety. It can be applied to elevators in construction sites, mine elevators, rail transit systems, industrial automation equipment, key components in the aerospace industry, lifting systems in marine engineering, as well as medical and emergency rescue equipment, ensuring the stability and safety of these systems during operation, reducing accident risks, and has significant practical application value.

However, there are also certain limitations to the research. Firstly, there was not much consideration given to the fault detection of elevator door systems in research. Technologies such as photosensitive sensors or image-based door anomaly detection are important components in addressing the safety of elevator door systems. Light sensors can monitor the status of elevator doors by sensing light, while image-based door anomaly detection technology requires advanced image processing algorithms and computer vision technology. Secondly, the study used unsupervised IF algorithm. However, relying solely on Unsupervised Learning is difficult to fully explore the deep information of data. Future research can combine Unsupervised Learning and Supervised Learning to further improve the detection accuracy and robustness of the model, especially in the case of labeled datasets. Thirdly, sensor data may be affected by environmental noise, which can affect the accuracy of elevator status monitoring. Future research

could consider deep learning techniques to address noise issues, improving noise processing accuracy by learning the features and patterns of noise. Fourthly, although UKF theoretically has high accuracy, it may face computational efficiency challenges in practical applications, especially in scenarios with large data volumes or high real-time requirements. Future research can explore model light weighting to reduce the time and resources required for retraining models on new tasks, as well as reduce the difficulty of model optimization and improve application efficiency through automated joint optimization. Fifthly, the deployment of this method in real environments will face complex building environments, especially large buildings with complex internal structures that may include multiple elevator shafts, different floors, and complex building layouts, leading to interference in sensor signals. Future research can optimize the layout and selection of sensors, reduce space occupation, and choose high-quality sensors. Shielding and filtering techniques can also be used to reduce interference.

REFERENCES

- [1] Nguyen T V, Jeong J H, Jo J. An efficient approach for the elevator button manipulation using the visual-based self-driving mobile manipulator. *Industrial Robot: the international journal of robotics research and application*, 2023, 50(1):84-93.
- [2] Beamurgia M, Basagoiti R, Rodríguez I, Rodríguez V. Improving waiting time and energy consumption performance of a bi-objective genetic algorithm embedded in an elevator group control system through passenger flow estimation. *Soft Computing*, 2022, 26(24):13673-13692.
- [3] Mangera M, Pedro J O, Panday A. GA-optimised nonlinear pseudo-derivative feedback control of a sustainable, high-speed, ultratall building elevator. *International Journal of Dynamics and Control*, 2022, 10(6):1903-1921.
- [4] Skog I. Nonintrusive Elevator System Fault Detection Using Learned Traffic Patterns. *IEEE Sensors Letters*, 2020, 4(11):1-4.
- [5] Oya J R G, Hidalgo-Fort E, Chavero F M, Carvajal R G. Compressive-Sensing-Based Reflectometer for Sparse-Fault Detection in Elevator Belts. *IEEE Transactions on Instrumentation and Measurement*, 2020, 69(1):947-949.
- [6] Ippili S, Russell M B, Herrin W D W. Deep learning-based mechanical fault detection and diagnosis of electric motors using directional characteristics of acoustic signals. *Noise Control Engineering Journal*, 2023, 71(5):384-389.
- [7] Mian T, Choudhary A, Fatima S. Multi-sensor fault diagnosis for misalignment and unbalance detection using machine learning. *IEEE Transactions on Industry Applications*, 2023, 59(5):5749-5759.
- [8] Neumayer S, Nimmer M, Setzer S, Steidl G. On the Robust PCA and Weiszfeld's Algorithm. *Applied Mathematics and Optimization*, 2020, 82(3):1017-1048.
- [9] Groumpos P P. A Critical Historic Overview of Artificial Intelligence: Issues, Challenges, Opportunities, and Threats. *Artificial Intelligence and Applications*. 2023, 1(4):197-213.
- [10] Feng X. Multiplant Location Involving Resource Allocation. *GEOGRAPHICAL ANALYSIS*, 2024, 56(1):97-117.
- [11] Rezova N, Kazakovtsev L, Rozhnov I, Stanimirovic P S, Shkaberina G. Hybrid Algorithms With Alternative Embedded Local Search Schemes For The p-Median Problem. *International Journal on Information Technologies & Security*, 2023, 15(4):61-72.
- [12] Anh A T H T, Duc L H. Super-capacitor energy storage system to recuperate regenerative braking energy in elevator operation of high buildings. *International journal of electrical and computer engineering*, 2022, 12(2):1358-1367.
- [13] Chen Y, Sanz-Alonso D, Willett R. Autodifferentiable ensemble Kalman filters. *SIAM Journal on Mathematics of Data Science*, 2022, 4(2):801-833.

- [14] Potokar E R, Norman K, Mangelson J G. Invariant extended kalman filtering for underwater navigation. *IEEE Robotics and Automation Letters*, 2021, 6(3):5792-5799.
- [15] Wei Y, Zhou B, Zhang J, Sun L, An D, Liu J. Review of Simultaneous Localization and Mapping Technology in the Agricultural Environment. *Journal of Beijing Institute of Technology*, 2023, 32(3):257-274.
- [16] Batuhan G, Serhat K. Multifractal detrended fluctuation analysis of insole pressure sensor data to diagnose vestibular system disorders. *Biomedical Engineering Letters*, 2023, 13(4):637-648.
- [17] Valente N F, Bilro L, Oliveira R. Hydrostatic Pressure Sensor Based on Polymer Optical Fiber Multimode Interferometer. *IEEE Sensors Journal*, 2023, 23(12):12876-12880.
- [18] Chang S, Ji B, Liu L B. Two algorithms of geodesic line length calculation considering elevation in eLoran systems. *IET radar, sonar & navigation*, 2023, 17(10):1469-1478.
- [19] Alam M I, Pasha A A, Jameel A G A, Ahmed U. High altitude airship: A review of thermal analyses and design approaches. *Archives of Computational Methods in Engineering*, 2023, 30(3):2289-2339.
- [20] Svetunkov I, Kourentzes N, Ord J K. Complex exponential smoothing. *Naval Research Logistics (NRL)*, 2022, 69(8):1108-1123.
- [21] Hamidy F, Yasin I. Implementation of Moving Average for Forecasting Inventory Data Using CodeIgniter. *Journal of Data Science and Information Systems*, 2023, 1(1):17-23.
- [22] Morales F A, Ramírez J M, Ramos E A. A mathematical assessment of the isolation random forest method for anomaly detection in big data. *Mathematical Methods in the Applied Sciences*, 2023, 46(1):1156-1177.
- [23] Xu H, Pang G, Wang W Y. Deep Isolation Forest for Anomaly Detection. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(12):12591-12604.
- [24] Runkai Z, Rong R, John Q. Reliable and fast automatic artifact rejection of Long-Term EEG recordings based on Isolation Forest. *Medical and Biological Engineering and Computing: Journal of the International Federation for Medical and Biological Engineering*, 2024, 62(2):521-535.

LFM Book Recommendation Based on Fusion of Time Information and K-Means

Dawei Ji

School of Humanities and Art, Nanchang Institute of Technology, Nanchang 330099, China

Abstract—To meet the growing demand in the field of book recommendation, the research focuses on meeting the personalized needs, behavioral patterns, and interests of readers. A book recommendation algorithm that combines K-means clustering with time information is proposed to provide more convenient and efficient book recommendation services and enhance readers' reading experience. The algorithm constructs a comprehensive user preference matrix by incorporating readers' borrowing time. Then, the K-means clustering is applied to group users with similar preferences and leverages a latent factor model to train and predict user ratings. The methodological integration of clustering and latent factor model ensures a more precise and dynamic recommendation process. The experimental results demonstrated that the proposed algorithm achieved a high average recommendation accuracy of 98.7%. Additionally, the algorithm maintained an average book popularity score of 8.2 after reaching stability, indicating its ability to suggest widely appreciated books. These outcomes validate the effectiveness of the algorithm in delivering accurate and popular book recommendations tailored to individual readers' needs. This study combines K-means clustering with time sensitive preference analysis and latent factor model to introduce an innovative method in the field of book recommendation systems. The findings provide valuable insights and practical applications for libraries seeking to enhance their personalized recommendation services, offering a significant contribution to the field of intelligent information retrieval.

Keywords—Book recommendation; K-means; time information; latent factor model; preference matrix

I. INTRODUCTION

The Book Recommendation (BR) system is an important application that can provide users with personalized book recommendations. As information technology develops, people's demand for obtaining, sharing, and purchasing books continues to increase, making recommendation systems crucial in helping users discover interesting books [1]. Traditional BR systems often use Collaborative Filtering (CF) algorithm for recommendation, which analyzes users' historical behavior data, mines similarities between users or items, and makes recommendations. However, CF algorithm still has some problems. Firstly, CF is unable to handle cold start issues well, which means that there is a lack of sufficient data to accurately recommend new users or newly listed books. Secondly, CF does not consider the specific relationship between books and users, resulting in a lack of diversity and personalization in recommendation results. To overcome these issues, a Latent Factor Model (LFM) is introduced into the BR system. LFM establishes the connection between users and books by representing them as latent feature vectors [2]. This

model not only considers the similarity between users and books, but also captures the implicit relationship between users and books, thereby improving the accuracy and personalization of recommendation results [3]. However, although LFM has achieved good results in solving some problems, the existing LFM-BR system cannot fully utilize the time information of users in the process of borrowing books. Meanwhile, it also cannot provide personalized recommendations for users, and still has certain limitations. Therefore, to solve these problems, a preference matrix is constructed by analyzing the borrowing time of readers to reflect their preferences. At the same time, the K-means algorithm and preference matrix are used for reader clustering to identify the reading needs of different preference reader groups. The implicit semantic model is trained and scored for prediction. An LFM-BR algorithm that integrates time information and K-Means clustering is designed. It is hoped to improve the recommendation performance of the BR system, provide users with more accurate personalized BR services, enhance user experience, and provide new ideas for the development of BR systems. This article consists of six sections. Section I is the background of the BR system. Related work is given in Section II. Section III reviews the research on recommendation systems both domestically and internationally. The sections designs the LFM-BR algorithm based on K-means and time information. It constructs a modified preference model based on time information. It optimizes the LFM recommendation algorithm based on K-means and modified preference models. Section IV analyzes the performance of the algorithm and its practical application effects. Finally, the entire article is summarized and its shortcomings are pointed out in Section VI.

II. RELATED WORKS

Due to the rapid development of the Internet, a large number of books can be obtained and read online. The number and variety of books continue to increase, but users often feel confused and exhausted. Therefore, to provide personalized BR services, many scholars have conducted in-depth research on recommendation systems. Guo Q et al. designed a knowledge graph recommendation system to address information explosion and enhance user experience in various online applications. The knowledge graph was used as auxiliary information to generate recommendations. These results confirmed that the system had higher recommendation accuracy [4]. Yi B et al. designed a deep matrix factorization model based on implicit feedback embedding to accurately recommend reader preferences. It directly generated potential factors of users and preferences from input information

through feature transformation functions. These results confirmed that this model had high accuracy and training efficiency [5]. Cui Z et al. designed a CF-based personalized recommendation system to provide users with accurate and fast information over time, which analyzed user behavior to provide higher quality recommendations. These results confirmed that the system could quickly and accurately make recommendations [6]. Zhou W et al. designed a graph-based personalized recommendation algorithm for sorting to improve user preference matching accuracy in recommendation systems. It matched target users with users with similar preferences through an improved resource allocation process. These results confirmed that the recommendation performance of this algorithm was good [7]. Liu Y et al. proposed a personalized library recommendation model based on small data fusion algorithm to better grasp the needs of library users and provide more accurate knowledge services. The neural network was utilized to achieve multi-dimensional small data fusion. These results confirmed that this model could effectively achieve personalized recommendations [8]. Liu Y designed a CF information recommendation algorithm based on spatiotemporal similarity to meet the academic information recommendation needs of university libraries. An academic information demand model was established through situational awareness and combined with adaptive interest models. These results confirmed that this algorithm could effectively achieve personalized recommendations [9].

Zhang S designed a personalized service method for university libraries based on data tracking technology to identify user interests and provide peer-to-peer service recommendations. The big data behavior tracking technology was utilized to analyze and track the behavioral information of user groups. These results confirmed that this method could accurately recommend and had high efficiency [10]. Frequent data scanning and excessive candidate itemset in the library lead to slow system operation. Therefore, Zhou Y proposed an information recommendation book management system based on improved Apriori. The method integrated C/S and B/S architectures to open book information to staff and borrowers. These results confirmed that the CPU usage of this system was relatively low [11]. Fu M proposed a personalized library resource recommendation system to address the low accuracy and user satisfaction of traditional library recommendation systems. It corrected the bias values and weights of visible and hidden layers in deep belief networks through contrastive divergence method. These results confirmed that this system had high accuracy, recall, and user satisfaction [12]. Chendhur K M K et al. designed an improved CF based on user preferences to improve the execution time and accuracy of prediction problems in BR. A small batch gradient descent algorithm was introduced to make predictions based on user preferences. These results confirmed that this algorithm had high prediction efficiency [13]. Anwar T et al. designed a cross domain BR for sequential pattern mining and rule mining to meet user needs in a shorter amount of time. The semantic similarity was utilized to expand domain recommendations and recommend books that users preferred through rule mining algorithms. These results confirmed that the system had a high-performance score [14]. Saraswat M et

al. designed a BR model based on neural recursive network classification to consider the combination of book types and reviews in BR. It categorized book plots and comments into various categories and recommends books to users based on these categories. These results confirmed that the accuracy and F1 value of this model were relatively high [15].

In summary, scholars have proposed various innovative methods aimed at providing personalized and accurate BR services. However, most of these methods rely on the user's historical behavioral data for recommendations, ignoring their real-time or immediate needs. Therefore, the study first constructs a comprehensive preference model for reader borrowing duration. Then, the K-means algorithm is used to cluster the readers. The clustering results are trained using the LFM model. A LFM-BR algorithm that integrates time information and K-means clustering is proposed. Compared with existing research, this method emphasizes the importance of temporal information and identifies reader groups with different preferences through clustering methods. It can better handle data with obvious group and time series characteristics, thus better grasping changes in readers' interests and needs, and making timely recommendations.

III. LFM BOOK RECOMMENDATION MODEL BASED ON TIME INFORMATION AND K-MEANS

This chapter mainly studies the improvement method of LFM-BR based on K-means and time information. Firstly, a preference model based on time information is constructed. Next is to improve the function design of BR.

A. Construction of a Modified Preference Model Based on Time Information

In the library, readers' borrowing preferences are a constantly changing dynamic process. Over time, readers' interests and needs will change, leading to new interests in different types of books. Faced with this situation, traditional CF often cannot provide satisfactory recommendation results [16]. To address the dynamic changes in reader borrowing preferences over time, this study analyzes the borrowing duration to deeply explore the potential preferences of readers and constructs a comprehensive preference degree model. The set of readers is $A = \{a_1, a_2, \dots, a_m\}$ and the set of books is $D = \{d_1, d_2, \dots, d_n\}$. Based on sets of books and readers, the personal reading preference in Eq. (1) can be obtained.

$$L_p(a_i, d_j) = 1 - \exp\left(-\frac{ct_{ij}}{t_i}\right) \quad (1)$$

In Eq. (1), L_p refers to individual reading preferences. $\exp()$ represents an exponential function. t_{ij} means the duration of time for reader a_i to borrow book d_j . \bar{t}_i is the average borrowing time of books borrowed by reader a_i . c represents a parameter that can be adjusted. However, the borrowing situation of each book varies due to factors such as content, number of pages, or category. Some books may attract more readers to borrow and pay attention to them due to their in-depth and engaging content, or their popular themes.

However, other books may have relatively fewer borrowed volumes due to their large number of pages or special categories. Therefore, the next step is to calculate the preference for borrowing books, represented by Eq. (2).

$$L_b(a_i, d_j) = \frac{t_{ij} - t_j(\min)}{t_j(\max) - t_j(\min) + b} \quad (2)$$

In Eq. (2), L_b refers to the degree of preference for book borrowing. $t_j(\min)$ represents the minimum borrowing time of book d_j . $t_j(\max)$ means the maximum borrowing time of book d_j . b is a bias term. Personal reading preferences can reflect the user preference for different themes or types of books, while book borrowing preferences reflect the user's tendency to choose a certain book in actual borrowing behavior. Therefore, considering the two preferences comprehensively, a comprehensive preference model is established by weighted sum to better understand user preferences and improve the accuracy of recommendations, represented by Eq. (3).

$$L_s(a_i, d_j) = \mu \cdot L_p(a_i, d_j) + (1 - \mu) \cdot L_b(a_i, d_j) \quad (3)$$

In Eq. (3), L_s represents the comprehensive preference of readers for borrowing books. μ is the weighting coefficient. The matrix in Table I shows the comprehensive preference.

TABLE I. COMPREHENSIVE PREFERENCE MATRIX

Reader /Book	d1	d2	...	dj	...	dn
a1	L (a1, d1)	L (a1, d2)	...	L (a1, dj)	...	L (a1, dn)
a2	L (a2, d1)	L (a2, d2)	...	L (a2, dj)	...	L (a2, dn)
...
ai	L (ai, d1)	L (ai, d2)	...	L (ai, dj)	...	L (ai, dn)
...
am	L (am, d1)	L (am, d2)	...	L (am, dj)	...	L (am, dn)

The interests and preferences of readers will also change over time. Therefore, by constructing preference transfer functions, different weights are assigned to preferences in different time periods. Based on the preference transfer function, the comprehensive preference model is optimized and a comprehensive preference correction model incorporating time information is designed. Considering the borrowing history of readers, recently returned books better reflect their current interests and preferences. Therefore, they should be given higher weight. Books that have been returned for a long time may reflect outdated interests and preferences, so certain punishments should be imposed on them. This is consistent with the law of human forgetting, which means that people are more likely to remember recent events, while their memory of distant events gradually becomes blurred. The Ebbinghaus forgetting curve can describe the forgetting pattern of the human brain. Fig. 1 shows the curve of human memory over time.

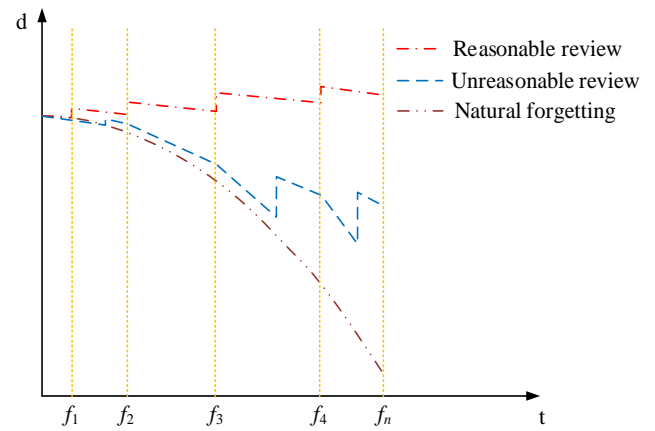


Fig. 1. Time dependent curve of human memory level.

According to Fig. 1, a function can be used to fit the Ebbinghaus forgetting curve and quantify the degree of forgetting. The study adopts Newton's cooling law, represented by Eq. (4).

$$-\frac{dT(t)}{dt} = \varphi [T(t) - T_C] \quad (4)$$

In Eq. (4), $-\frac{dT(t)}{dt}$ represents the rate at which the temperature $T(t)$ of the object decreases over time t . T_C represents the temperature of the surrounding environment at time t . φ represents the cooling coefficient, which is a proportional constant. The next step is to calculate the relationship between the object temperature at time φ and the initial time through mathematical operations, represented by Eq. (5).

$$T(t) = T_C + [T(t_0) - T_C] \cdot \exp[-\kappa(t - t_0)] \quad (5)$$

In Eq. (5), t_0 represents the initial time. The last time the reader returns the book before the current moment is used as the evaluation criterion. Eq. (5) is adjusted to obtain the preference transfer function, which is represented by Eq. (6).

$$\omega(a_i, d_j) = \varepsilon + (1 - \varepsilon) \cdot \exp[-\zeta(t_c - t_{last})], \zeta \in (0, 1) \quad (6)$$

In Eq. (6), ω is the preference weight. ε represents a constant. t_{last} is the last time the book is returned. ζ means a time decay coefficient. Finally, by combining the preference transfer function with the comprehensive preference model, a preference correction model based on time information can be obtained, represented by Eq. (7).

$$L'(a_i, d_j) = \omega(a_i, d_j) \cdot [\mu \cdot L_p(a_i, d_j) + (1 - \mu) \cdot L_b(a_i, d_j)] \quad (7)$$

In Eq. (7), L' is the modified preference based on time information. Fig. 2 shows the modified preference matrix.

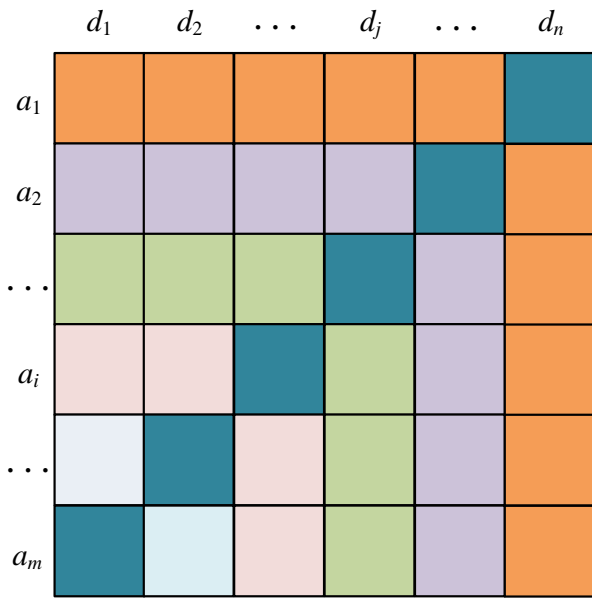


Fig. 2. Modified preference matrix diagram.

B. LFM Recommendation Algorithm Based on K-Means and Modified Preference Degree

The relationship between readers and book categories was not taken into account in the design and modification of preferences. Therefore, it is impossible to fully explore the potential preference information of users. In some libraries, there are more books. The borrowing records of readers are relatively rare [17]. Therefore, the first step is to regard the set of book categories as $G = \{g_1, g_2, \dots, g_z\}$ and combine the set of books with $G = \{g_1, g_2, \dots, g_z\}$ to form a category matrix. The elements in the j row and k column of the category matrix are set to Boolean values, either 0 or 1. When $bg_{jk} = 1$, it indicates that the book b_j belongs to the g_k class. When $bg_{jk} = 0$, it indicates that b_j does not belong to g_k . The category preference of books in Eq. (8) can be obtained.

$$Q_{ik} = \frac{f_{ik}}{\sum_{z=1}^z f_{iz}} \quad (8)$$

In Eq. (8), Q_{ik} means the reader's preference for borrowing books. f_{ik} represents the frequency of borrowing g_k books. f_{iz} is the frequency of borrowing g_z books.

The next step is to cluster readers using K-means based on the category preference matrix between readers and books, as displayed in Fig. 3.

When clustering, cluster centers are selected based on the similarity between reader preferences for different book categories. The next step is to combine reader clustering with preference correction matrix to establish a preference matrix for the same cluster of readers. The cosine similarity is introduced, and the category preference matrix is inputted to calculate the similarity, which is represented by Eq. (9).

$$similarity(a_x, a_y) = \frac{\sum_{k=1}^z (Q_{xk} \cdot Q_{yk})}{\sqrt{\sum_{k=1}^z (Q_{xk})^2} \cdot \sqrt{\sum_{k=1}^z (Q_{yk})^2}} \quad (9)$$

In Eq. (9), $similarity(a_x, a_y)$ is the similarity in book category preferences among different readers [18]. The next step is to train using the LFM recommendation algorithm, which decomposes the matrix into two low dimensional matrices. One matrix represents the relationship between users and potential features, while the other matrix represents the relationship between items and potential features. These potential features can capture the implicit relationship between users and projects, namely implicit classification. By learning these potential features, users can predict their ratings for projects they have never interacted with before. The user rating of the project is represented by Eq. (10).

$$\begin{cases} S = U \times P^T \\ r_{zi} = \sum_{h=1}^H p_{zh} q_{ih} \end{cases} \quad (10)$$

In Eq. (10), S is the rating matrix. U represents the relationship matrix between decomposed users and potential features. P means the relationship matrix between the decomposed project and potential features. r_{zi} refers to the predicted score. H is the number of hidden classifications. p_{zh} represents interest level. q_{ih} means the association between projects and implicit classification. The next step is to obtain the values of parameters p_{zh} and q_{ih} through the objective function. Meanwhile, to prevent overfitting, a regularization term is added to the objective function, represented by Eq. (11).

$$\Phi = \sum_{(z,i) \in R} (r_{zi} - \sum_{h=1}^H p_{zh} q_{ih})^2 + \lambda (\|p_z\|^2 + \|q_i\|^2) \quad (11)$$

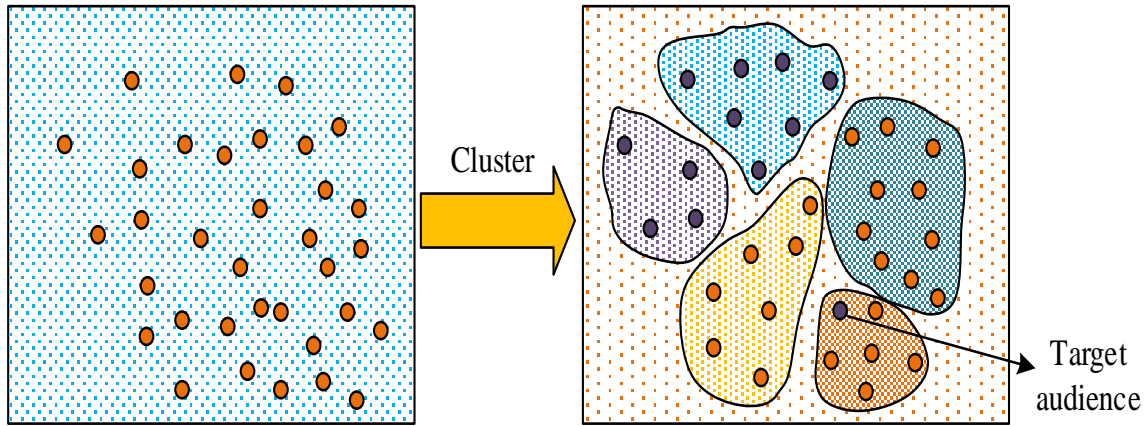


Fig. 3. K-means reader clustering diagram.

In Eq. (11), Φ represents the objective function. R refers to the scoring set. λ is the regularization coefficient. The solution of the objective function usually uses gradient descent method, which minimizes the objective function by taking its derivative and gradually reducing the parameters values. In the gradient descent method, the parameters values are first initialized. Then, the partial derivatives of each parameter in the objective function are calculated to obtain the gradient of the parameters. Next, based on the learning rate setting, the update amount of the parameter is obtained by multiplying it by the gradient value, and it is added to the current parameter value. This process continues until the specified stopping criterion is reached, that is, the objective function has converged or reached a certain number of iterations. The solution of parameters p_{zh} and q_{ih} is represented by Eq. (12).

$$\begin{cases} p_{zh} = s_z q_{ih} (q_{ih}^T q_{ih} + \lambda I)^{-1} \\ q_{ih} = s_z p_{zh} (p_{zh}^T p_{zh} + \lambda I)^{-1} \end{cases} \quad (12)$$

In Eq. (12), s_z represents the rating of user z . I is the identity matrix. The training process has been completed. Fig. 4 shows the training process of the LFM recommendation algorithm.

The next step is to introduce Sum of Squared Errors (SSE) to evaluate the clustering performance of K-means-based reader clustering, represented by Eq. (13).

$$SSE = \sum_{i=1}^k \sum_{x \in A_i} |sim(C_i, x)|^2 \quad (13)$$

In Eq. (13), A_i represents the set of reader clusters. C_i represents the clustering center of the reader cluster. The next step is to introduce Mean Absolute Error (MAE) and Root

Mean Squared Error (RMSE) to evaluate the rating accuracy of the designed recommendation system. MAE is the average absolute difference between the predicted score and the actual score. A smaller value indicates that the predicted score is more accurate. RMSE is calculated based on squared error, taking into account the error between each predicted score and the actual score, and averaging the error. A small RMSE indicates that the predicted score is close to the actual score. These two indicators are represented by Eq. (14).

$$\begin{cases} MAE = \frac{\sum_{z,i \in N} |r'_{zi} - r_{zi}|}{N} \\ RSME = \sqrt{\frac{\sum_{z,i \in N} (r'_{zi} - r_{zi})^2}{N}} \end{cases} \quad (14)$$

In Eq. (14), N represents the total amount of data. r'_{zi} stands for the reader's true rating of the book. Finally, the accuracy, recall, and F1 score are used to predict the accuracy of the recommendation list, represented by Eq. (14).

$$\begin{cases} Accuracy = \frac{\sum_{o \in O} |S(o) \cap R(o)|}{\sum_{o \in O} |S(o)|} \\ Recall = \frac{\sum_{o \in O} |S(o) \cap R(o)|}{\sum_{o \in O} |R(o)|} \\ F1 = \frac{2 \cdot Accuracy \cdot Recall}{Accuracy + Recall} \end{cases} \quad (15)$$

In Eq. (15), O represents the set of users. $S(o)$ means the recommended list. $R(o)$ is a set of user preferences. Fig. 5 shows the designed LFM recommendation algorithm based on K-means and modified preference.

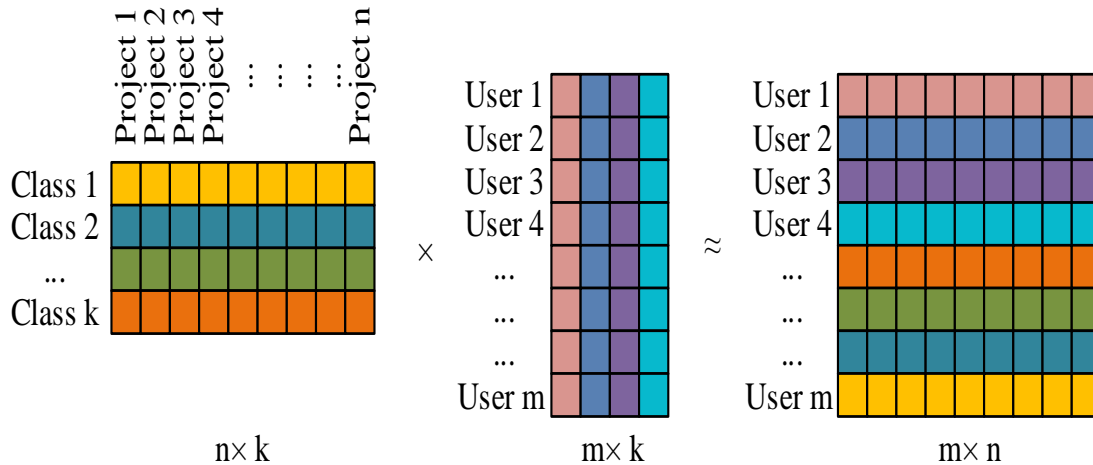


Fig. 4. The training process of the LFM recommendation algorithm.

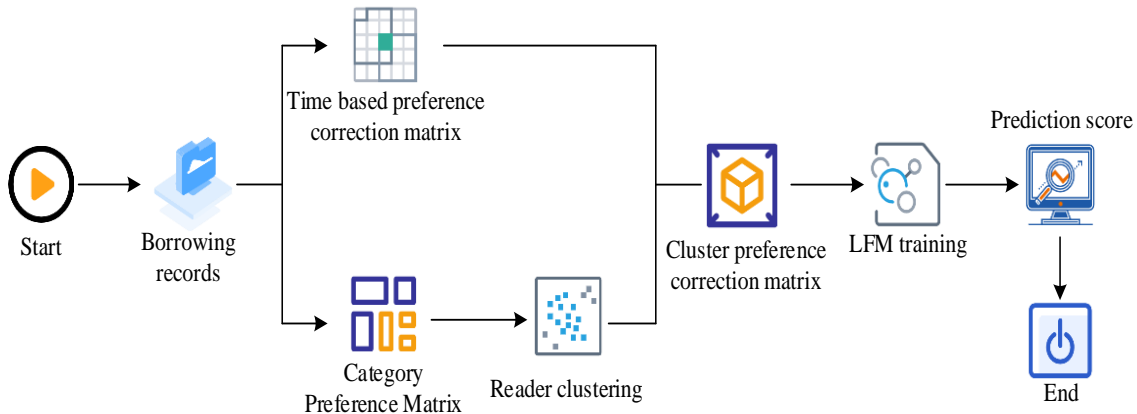


Fig. 5. LFM recommendation algorithm based on K-means and modified preference.

IV. RESULTS OF LFM RECOMMENDATION ALGORITHM BASED ON K-MEANS AND TIME INFORMATION

This section mainly analyzes the experimental results of the designed LFM-BR. The first step is to analyze the performance of the designed algorithm. The second step is to design simulation experiments for its practical application.

A. Performance of LFM Recommendation Algorithm based on K-Means and Time Information

To verify the performance of the designed LFM recommendation algorithm, the study first selected different numbers of clusters and calculated SSE to obtain the optimal number of clusters, as displayed in Table II.

TABLE II. SSE VALUES FOR DIFFERENT NUMBER OF CLUSTERS

Number of clusters	SSE value	Number of clusters	SSE value	Number of clusters	SSE value
2	2.279	8	0.898	14	0.608
3	1.875	9	0.784	15	0.473
4	1.663	10	0.715	16	0.498
5	1.512	11	0.692	17	0.457
6	1.318	12	0.618	18	0.473
7	1.301	13	0.635	19	0.463

From Table II, as the clusters increased, the SSE of the designed BR gradually decreased. When the cluster was less than 7, the decrease rate of SSE was faster. When the cluster was greater than 7, the decrease rate of SSE became slower. Therefore, the optimal clusters were 7. The above results indicate that selecting the appropriate number of clusters has a significant impact on clustering effectiveness. In practical applications, it is necessary to choose the appropriate number of clusters based on different business needs and data characteristics to achieve the best recommendation effect. The next step is to calculate the MAE and RMSE of the designed BR, and compare them with CF, LFM, and LFM based on time information. Fig. 6 shows the results.

From Fig. 6 (a), as the hidden classification increased, the MAE of CF remained unchanged, while the MAE of other three algorithms showed a decreasing trend and gradually flattened out. Among them, the MAE of CF was always 0.26. The maximum MAE of LFM was 0.32 and the minimum value was 0.24. The maximum MAE of LFM based on time information was 0.29 and the minimum value was 0.23. The maximum MAE of the designed LFM recommendation algorithm based on K-means and time information was 0.29, and the minimum value was 0.21. From Fig. 6 (b), the RMSE of CF still did not vary with the hidden classifications, and its

value remained at 0.33. The maximum RMAE of LFM was 0.39 and the minimum value was 0.30. The maximum RMAE of LFM based on time information was 0.34 and the minimum value was 0.28. The maximum RMAE of the designed BR was 0.31 and the minimum value was 0.23. The above results indicate that the designed recommendation algorithm has good performance on prediction accuracy. Finally, to further validate the performance of the designed BR, the accuracy, recall, and F1 score of the four recommendation algorithms were calculated, as displayed in Fig. 7.

From Fig. 7 (a), as the iteration increased, the accuracy of all four algorithms showed an upward trend but gradually stabilized. When CF reached a flat state, the accuracy was 87.8%, and the maximum accuracy of LFM was 93.5%. The

LFM based on time information achieved a stable accuracy of 95.7%. The maximum accuracy of the designed BR was 97.3%. From Fig. 7 (b) and 7 (c), the recall and F1 score trends of these four algorithms were consistent with the accuracy trends. When CF reached stability, the recall rate and F1 score were 92.1% and 0.91, respectively. When LFM reached stability, the recall rate and F1 score were 92.6% and 0.93, respectively. When LFM based on time information tended to flatten, the recall rate and F1 score were 95.1% and 0.953, respectively. When the designed algorithm tended to flatten, the recall rate and F1 score were 98.2% and 0.965, respectively. The three indicators of this designed algorithm are significantly higher than the other three algorithms, proving its good overall performance.

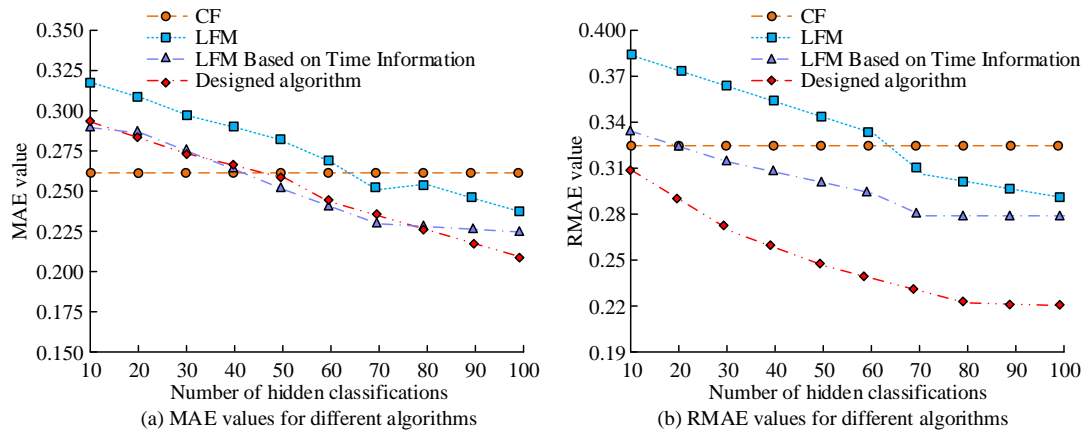


Fig. 6. MAE and RMSE values of four book recommendation algorithms.

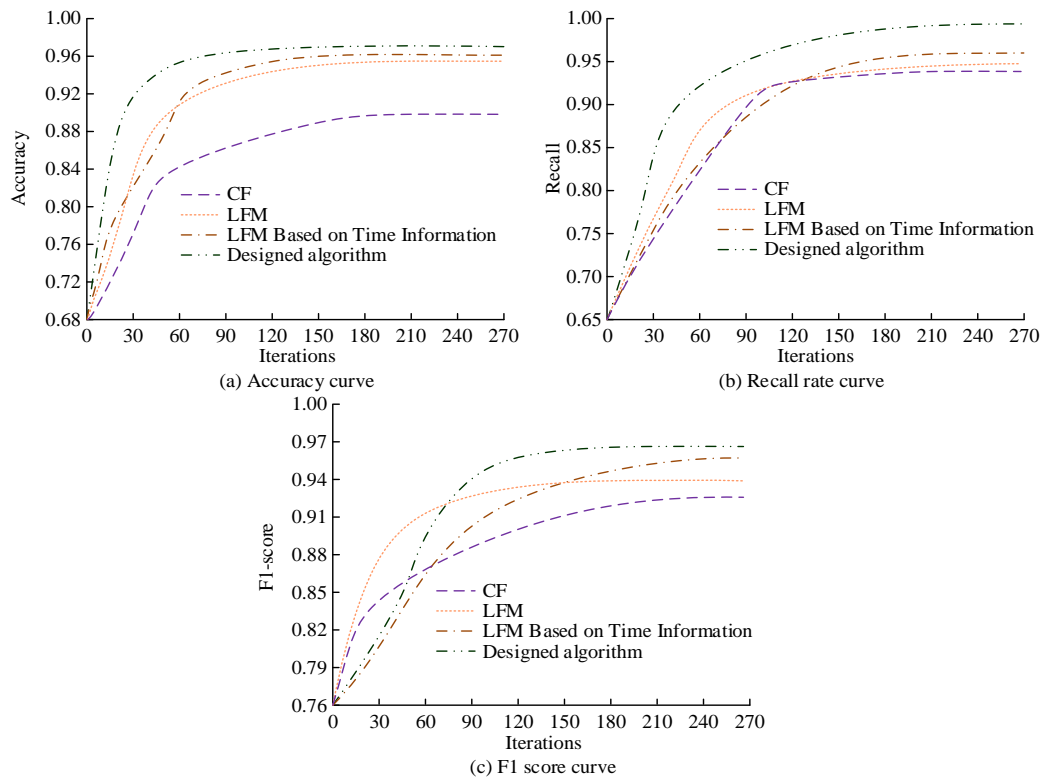


Fig. 7. Results of different indicators.

B. Application Effectiveness of LFM Recommendation Algorithm based on K-Means and Time Information

To verify the effectiveness of the designed BR in practical applications, a simulation experiment was conducted using Python 3.9 in a hardware environment with a 64 bit Windows 10 operating system, Intel (R) Core (TM) i5-12500K processor, 8GB of RAM, and 1TB of hard disk capacity. Firstly, the sparsity and computation time of the four algorithms are calculated. Fig. 8 shows the comparison results.

From Fig. 8 (a), as the iteration increased, the sparsity of different algorithms was effectively reduced. The sparsity

reduction effect of the designed BR was significantly better than other algorithms, with a sparsity of 66.5% at 180 iterations. From Fig. 8 (b), the computation time for using all four algorithms for recommendation showed a decreasing trend and tended to stabilize after reaching a certain iteration. The calculation time when the designed algorithm reached stability was 10.2s. The above results demonstrate that the designed algorithm can better solve the data sparsity in BR, and also prove its high computational efficiency. The next step is to calculate the average accuracy and coverage of different algorithms separately, as displayed in Fig. 9.

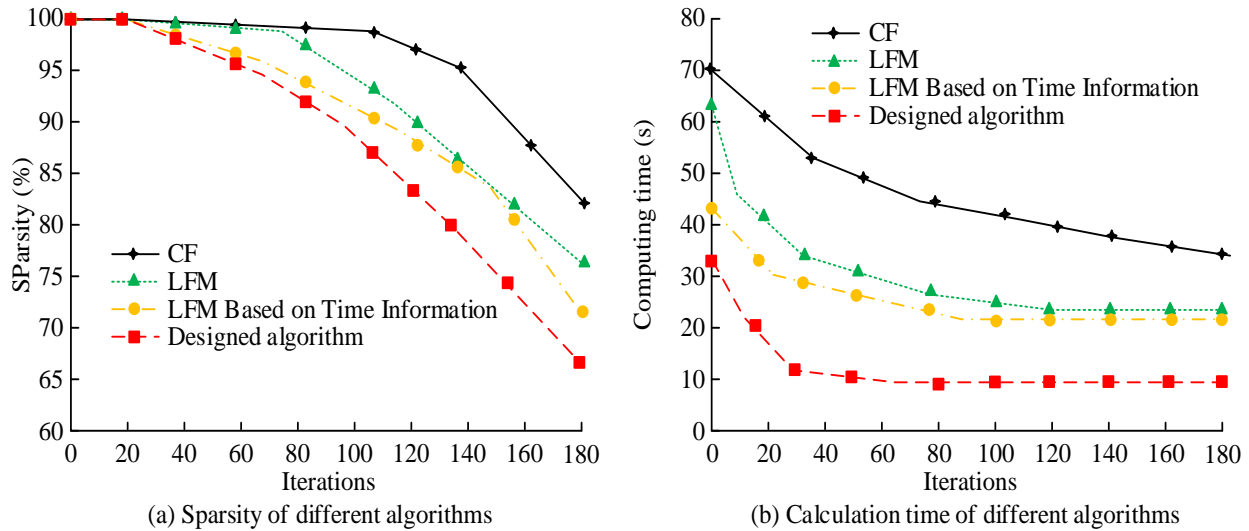


Fig. 8. Sparsity and computational time of different algorithms.

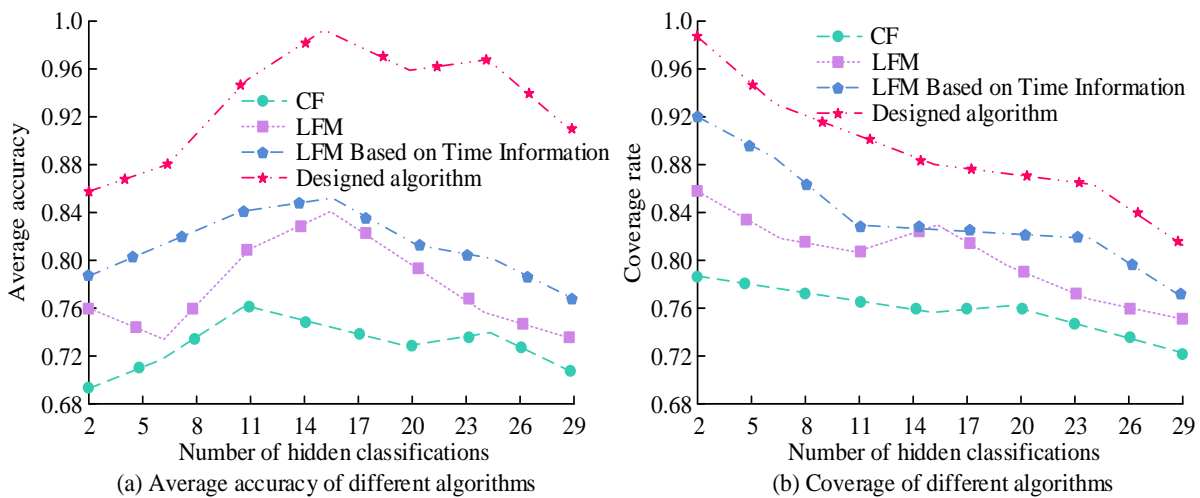


Fig. 9. Average accuracy and coverage of different algorithms.

From Fig. 9 (a), the average accuracy of the designed algorithm was significantly higher than other algorithms, reaching a maximum of 98.7%, further proving its recommendation accuracy. From Fig. 9 (b), as the number of hidden classifications increased, the coverage trends of these four algorithms were consistent. However, the coverage curve

of the designed algorithm had always been above that of other algorithms, indicating that the designed algorithm could cover more items in the recommendation process, proving its comprehensiveness and diversity. Finally, the average popularity of different algorithms was calculated to evaluate the effectiveness of the designed recommendation algorithm, as displayed in Fig. 10.

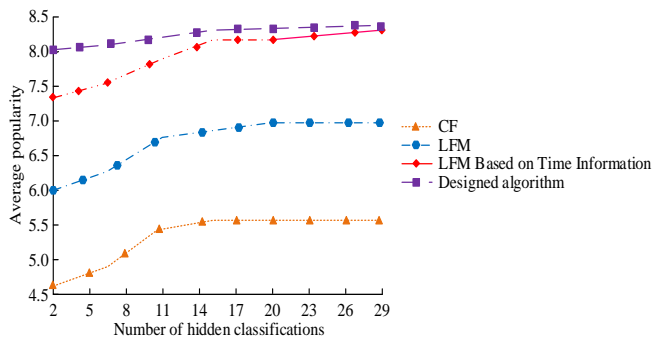


Fig. 10. Average popularity of different algorithms.

From Fig. 10, the average popularity of these four algorithms gradually increased and tended to stabilize. The average popularity of CF reaching a flat state was 5.5. The average popularity of LFM reaching a flat state was 6.8. The average popularity of LFM based on time information after stabilizing was 7.9. The average popularity of the designed recommendation algorithm after reaching stability was 8.2. The above results indicate that the book recommended by the designed algorithm has a high popularity, proving its good recommendation effect.

V. DISCUSSION

With the progress of the times, the number of books has increased sharply, and the readership has become more diverse. Traditional BR methods have been unable to meet the diverse needs. The research aims to provide more accurate and personalized BR services, and proposes an LFM-BR algorithm that integrates time information and K-means clustering. The results showed that the MAE and RMSE values of the proposed algorithm were 0.21 and 0.23, respectively, which were higher than other algorithms, proving its high accuracy. Similar conclusions were drawn by Yi B et al. [5], who proposed an implicit feedback embedding deep matrix factorization model that can capture potential features of users, but it cannot effectively handle cold start problems. The proposed method quantifies the impact of time factors on user preferences by constructing a preference transfer function, effectively alleviating the cold start problem. The sparsity of the proposed algorithm was 66.5% at 180 iterations, indicating that it could better solve the data sparsity problem in BR. This conclusion is similar to the conclusion drawn by Fu M et al. [12], but the proposed algorithm is significantly better. This is because the proposed algorithm introduces the K-means algorithm to cluster readers with similar preferences and incorporates time information to enhance the system robustness, thus better addressing the data sparsity. The proposed algorithm took 10.2 seconds to reach a steady state, significantly lower than other algorithms, demonstrating its high computational efficiency. This is similar to the conclusion drawn by Zhou Y [11], and the proposed algorithm is superior. This is because the proposed algorithm optimizes the objective function and uses gradient descent to update parameters, ensuring efficient computation on large-scale datasets. In summary, the proposed method effectively improves the effectiveness of BR by integrating time information, optimizing clustering strategies, and enhancing LFM models, which is of great significance for promoting

personalized library services.

VI. CONCLUSION

In the era of information explosion, BR has become an important tool to help users quickly find suitable reading interests. To improve the accuracy and personalization of BR, an LFM-BR based on K-means and time information was designed. Firstly, a comprehensive preference model based on time was constructed through borrowed books, followed by reader clustering using K-means, and finally training using LFM. These results confirmed that the maximum MAE of the designed algorithm was 0.29 and its minimum value was 0.21. The maximum RMSE value was 0.31 and its minimum value was 0.23, all higher than other algorithms, proving its high prediction accuracy. In terms of accuracy, recall, and F1 score calculation, the designed algorithm was 97.3%, 98.2%, and 0.965, respectively, proving its good overall performance. In terms of sparsity and computation time, the designed recommendation algorithm had a sparsity of 66.5% at 180 iterations, and a computation time of 10.2s to reach a stable state. These prove that it can effectively solve the data sparsity in BR and has high computational efficiency. The above results confirm that the designed algorithm has high accuracy and personalization in the BR problem, and can accurately reflect the reading interests and needs of users. However, the study does not consider other behavioral data of readers, which may have a certain impact on the results. Future research will incorporate more user behavior data to construct more comprehensive user preference models, and introduce new natural language processing techniques to further enhance the performance of BR systems.

REFERENCES

- [1] Wu S, Sun F, Zhang W, Xie X, Cui B. Graph neural networks in recommender systems: a survey. *ACM Computing Surveys*, 2022, 55(5): 1-37.
- [2] Wu D, Shang M, Luo X, Wang Z. An L 1-and-L 2-norm-oriented latent factor model for recommender systems. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 33(10): 5775-5788.
- [3] Nengem S M. Symmetric Kernel-Based Approach for Elliptic Partial Differential Equation. *Journal of Data Science and Intelligent Systems*, 2023, 1(2): 99-104.
- [4] Guo Q, Zhuang F, Qin C, Zhu H, Xie X, Xiong H, He Q. A survey on knowledge graph-based recommender systems. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 34(8): 3549-3568.
- [5] Yi B, Shen X, Liu H, Zhang Z, Zhang W, Liu S, Xiong N. Deep matrix factorization with implicit feedback embedding for recommendation system. *IEEE Transactions on Industrial Informatics*, 2019, 15(8): 4591-4601.
- [6] Cui Z, Xu X, Fei X U E, Cai X, Cao Y, Zhang W, Chen J. Personalized recommendation system based on collaborative filtering for IoT scenarios. *IEEE Transactions on Services Computing*, 2020, 13(4): 685-695.
- [7] Zhou W, Han W. Personalized recommendation via user preference matching. *Information Processing & Management*, 2019, 56(3): 955-968.
- [8] Liu Y, Xu T W, Xiao M. Small Data Fusion Algorithm for Personalized Library Recommendations. *International Journal of Information and Communication Technology Education (IJICTE)*, 2023, 19(1): 1-14.
- [9] Liu Y. Survey of Intelligent Recommendation of Academic Information in University Libraries Based on Situational Perception Method. *Journal of Education and Learning*, 2020, 9(2): 197-202.
- [10] Zhang S. Personalized Service for University Library Users Based on Data Tracking. *Voice of the Publisher*, 2022, 8(2): 41-49.

- [11] Zhou Y. Design and implementation of book recommendation management system based on improved Apriori algorithm. *Intelligent Information Management*, 2020, 12(3): 75-87.
- [12] Fu M. The design of library resource personalised recommendation system based on deep belief network. *International Journal of Applied Systemic Studies*, 2023, 10(3): 205-219.
- [13] Chendhur K M K, Priya V, Priya R M, Lakshmi S L. Book recommender system using improved collaborative filtering. *International Journal of Research in Engineering, Science and Management*, 2021, 4(4): 51-56.
- [14] Anwar T, Uma V. CD-SPM: Cross-domain book recommendation using sequential pattern mining and rule mining. *Journal of King Saud University-Computer and Information Sciences*, 2022, 34(3): 793-800.
- [15] Saraswat M, Srishti. Leveraging genre classification with RNN for Book recommendation. *International Journal of Information Technology*, 2022, 14(7): 3751-3756.
- [16] Luo X, Yuan Y, Chen S, Zeng N, Wang Z. Position-transitional particle swarm optimization-incorporated latent factor analysis. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 34(8): 3958-3970.
- [17] Yi B, Shen X, Liu H, Zhang Z, Zhang W, Liu S, Xiong N. Deep matrix factorization with implicit feedback embedding for recommendation system. *IEEE Transactions on Industrial Informatics*, 2019, 15(8): 4591-4601.
- [18] Choudhuri S, Adeniye S, Sen A. Distribution Alignment Using Complement Entropy Objective and Adaptive Consensus-Based Label Refinement For Partial Domain Adaptation//*Artificial Intelligence and Applications*. 2023, 1(1): 43-51.

A Proposed Approach for Agile IoT Smart Cities Transformation— Intelligent, Fast and Flexible

Othman Asiry¹, Ayman E. Khedr², Amira M. Idrees³

University of Jeddah, College of Computing and Information Technology at Khulais-Department of Information Technology,
Jeddah, Saudi Arabia¹

University of Jeddah, College of Computing and Information Technology at Khulais-Department of Information Systems,
Jeddah, Saudi Arabia²

Faculty of Computers and Information Technology, Future University in Egypt, Egypt³

Abstract—Smart city architectures have been varying from one community to another. Each community leader develops their own perspective of smart cities. Some of these communities focus on data management, while others focus on provided services and infrastructure. In this research, an attempt to propose a clear, complete, and efficient perspective of smart cities is accomplished. The proposed generic architecture clarifies the full capabilities, requirements, and layers' contribution to a successful smart city development. The proposed architecture utilizes Internet of Things tools as well as agile standards in the description of each layer. The research aims to discuss each layer in detail, the relationships among layers, the applied technology, and every aspect that leads to the success of using the recommended architecture. Although smart cities, IoT, and agile research have previously tackled the relation of each one of them with the other, up to the researchers' knowledge, the three paradigms have not been previously considered as a unified collaborative approach. In order to reach the research target, the proposed architecture intelligently utilizes these paradigms and presents a robust architecture with high-quality standards.

Keywords—Smart city; agility; Internet of Things; cloud computing; intelligent systems

I. INTRODUCTION

Smart cities can drive intelligence to every aspect of the lives of the population. It affects management, decision-making, culture, and regularities. Smart cities consider re-inventing the way of living while generating value for the public. The main objective for developing smart cities is to enhance the quality of the population's way of living and provide them with an easy, sustainable, and satisfying life either in social or professional directions. In order to reach this goal, it is vital to identify all the requirements, obstacles, and challenges to be able to build smart cities. The term "smart city" has been tackled in different research studies; however, other aspects related to information technology are usually associated with it. The continuous population growth has become a primary reason for services' quality and hinders the development. Almost all fields have been affected by this exponential population growth, such as education, health, energy, and many others. Smart city development has been an effective solution supporting the city's sustainability. Quality is an associated aspect of smart cities. Embarking digitalization is vital for smart cities' transformation. Utilizing recent

technology, initiating new regulations, motivating the economy, and other factors are related to the concept of smart cities [1].

The idea of smart cities emerged in the nineties. The first wave focused on the contribution of the information technology concept to the main cities' infrastructure. Then, a focus on the community and its players highlighted how these players can have an easier life in smart cities. Later, smart cities term has been closely related to digitalization and intelligence; several factors effect has been discussed, such as people's resistance, the communities' culture, and others. Along the journey of smart cities' emergence, a broader perspective was highlighted, including the service's intelligence such as transportation systems, electricity system [2], construction field, banking systems, and others. These perspectives strongly emphasized the sustainability value of the cities' resources.

The concept of a smart city is exponentially increasing either for small or large cities. This global emergence drove the world to initiate new rules and regulations for the new era of cities. Many investments in smart cities have been proposed with the objective of innovation, growth, and enhancing life. All existing architectures depend mainly on the enabled technology for the smart features of the community. It has become feasible now to develop smart cities with the current acceptance and interoperability of IoT and advanced data analysis. Another perspective of smart cities is the ability to smart intelligent decisions by utilizing intelligent techniques for predictive analysis, better planning, and management [3]

This research traces all the proposed architectures for smart cities, traces the findings for embedding the technologies in the architectures infrastructures, and identifies the regularities and critical aspects for the transformation goal. This research proposes a generic architecture for smart cities, which is applicable for implementation with all its aspects, however, flexible. The proposed architecture discusses six layers and contributors to each one with giving the implementers the opportunity to implement the valid component without affecting the remaining architecture. Several factors are related to the concept of smart cities development including culture, quality, acceptance, resistance, and sustainability. These factors will be discussed later in this research. Moreover, other emerged factors such as IoT, agility, and motivating population are also considered as pillars in the proposed architecture.

II. RELATED WORK

IoT devices strongly contribute to smart cities as one of the main pillars. The devices are considered as sources that provide numerous amounts of data. Therefore, unification to a single data model is recommended, which should be able to deal with such data. Several researchers have proposed different models to deal with the data heterogeneity, temporal, and complexity. The research [4] included a model that received data from IoT devices and determined the correlation between the growth of both industries and urban. The research concluded that managing knowledge directly influences the development of the industry. The study recommended that the enterprises' technological communication enhances knowledge management and promising technologies investments which leads to alliance flexibility. Qian et al. (2019) [5] discussed the positive impact of IoT technology on empowering the infrastructure and, consequently economic growth and sustainability. Chen et al. (2020) [6] focused on the contribution of IoT technology and deep learning techniques in construction field. The research concluded that these contributing fields have leveraged the process of construction by exploring the shortest time consumed in the evacuation process. Watson et al. (2020) [7] also focused on IoT's contribution to the construction of smart cities and concluded that the field still needs further research for further development requirements of IoT technologies to leverage the construction of smart cities.

Focusing on cloud computing contribution in smart cities. Lv et al. (2018) [8] proposed a platform for smart cities in which cloud computing and geographical information systems successfully contributed. The research highlighted that both fields could positively affect the success level of monitoring environment phenomena and predicting expected disasters as well as exploring transportation system performance. Hossain et al. (2018) [9] highlighted the need for cloud computing platforms for higher computing technology when contributing to IoT devices access in smart cities such as edge computing for more effective computations and latency avoidance. Javadzadeh et al. (2020) [10] focused on the same latency issue and proposed the same fog computation solution. Giannakoulis et al. (2019) [11] focused on the security issues of cloud environments and their effect on smart cities platforms as well as [12]. The research in [13] highlighted that IoT and cloud computing contribution in smart cities enhances the services performance with the ability to reduce the servers' load as well as reducing their number which greatly contribute the construction cost.

The previous research highlighted the positive contribution of both IoT and cloud computing with focusing on a determined subject, however, up to the researchers knowledge, there was no research that provided a complete perspective of smart cities which is one of the vital issues to present the whole smart cities pictures, the interaction between all contributing aspects and the need for homogenous cooperation among all contributing players. Moreover, the research [14] discussed the issues of smart cities and confirmed that there is no complete structure

for smart cities while all current projects are pilot individual projects that included trade-offs of some aspects over the others.

III. RESEARCH CONTRIBUTION

To be able to present a complete flexible generic platform, twelve platforms have been extensively studied. These platforms have been developed by international organizations following the required standards. Moreover, different projects and studies have been also studied. The aim of this extensive study is to explore the characteristics of a smart city architecture including core, extendable, and supportive characteristics. Ensuring the collecting of all aspects reflects that the proposed platform proposes the best practice when compared with any other platform. This comparison will be performed in the evaluation section. While there were differences between the collected platforms, however, it was also clear that common segments could be highlighted. Extensive study has been performed to explore the strength and weakness between these platforms which lead to the initiative for a generic platform that provide solutions to hinder the possible obstacles and fully consider the reality of cities. The proposed platform has taken into consideration the international standards, system requirements, services, data, and security management criteria for smart cities [16].

IV. A PROPOSED AGILE IOT ARCHITECTURE

Different smart cities platforms recommendations have been studied aiming to explore the possibility to propose a successful platform with maximum interoperability with the common infrastructure of cities [15]. This target ensures accepting the proposed platform and encourages its implementation. The authors also aimed at proposing a generic platform which utilizes the commonly used technologies to ensure its implementation adequacy in different environments. The proposed generic platform that could be implemented using different technologies with various levels of maturity. The proposed platform could also be easily extended or reduced according to the capabilities; the platform flexibility will be discussed in detail in the following sections.

The proposed architecture (Fig. 1) consists of five main layers in addition to a continuing monitoring and governance layer. Each layer components will be discussed in detail in the next subsections. Moreover, Fig. 1 illustrates the proposed architecture.

The following key aspects identify the smart city platform that fits with the available capabilities.

- Identify the needed capabilities and the specifications for each capability.
- Ensure the delivery of the required capabilities with sufficient flexibility.
- Ensure all compliant solutions interoperability.

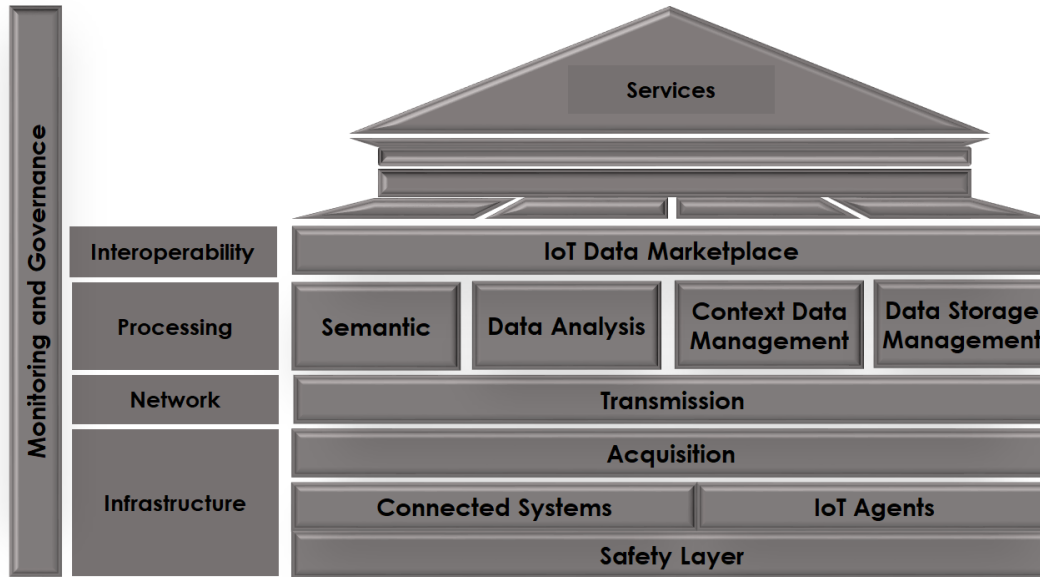


Fig. 1. Proposed generic architecture.

A. Infrastructure Layer

The infrastructure layer describes all the required utilities, devices, sensors, facilities, as well as the sources and its indicators. The infrastructure supports all the services, transportation, and environmental, as well as governmental aspects. Moreover, networking infrastructure is one of the compulsory components which configuration should be compatible with the predicted participating items. The network configuration could be considered one of the most critical aspects in the construction of smart cities.

1) *City sources components and indicator:* Many vital components which could be successful key aspects in smart cities are discussed earlier [17] (Fig. 2). These components are demonstrated in this section with discussing their effective contribution. The following points discusses each component, its standards, perspective, and contribution to the smart cities' paradigm. It is worth highlighting that although this section discusses each component individually, however; practically speaking; these components are interrelated. Therefore, a component could be discussed in various places in this section.

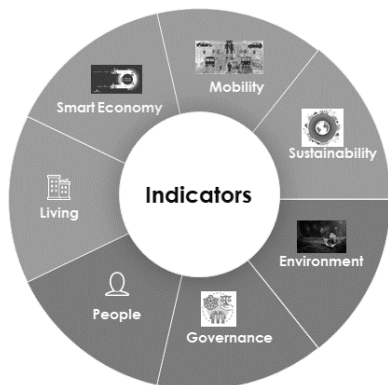


Fig. 2. Smart city indicators.

a) *Smart Economy:* Smart economy is achieved by incorporating a set of features into the economy such as innovation, digital transformation, sustainability, and high productivity. Smart economy focuses on the level of quality and standards of life which is evaluated by a new set of variables including basic needs, community participation, technology advancement, and scientific aspects. The concept of smart cities is multidimensional dynamic as well as adaptable. It includes economic, social, and motivating individuals' aspects. There is an immense need to balance between the current and next generations. Therefore, the focus of developing the current generation could result in identifying the effective methods and polices for development while maintaining the required quality with continuous quality upgrading. Smart economy concept leads to a clear identification to the relationship between the value of the economy and the satisfaction level such as the employees' satisfaction, customers' satisfaction, etc. it is also related to the level of availability either products or services with the satisfying quality. Therefore, the bond between smart economy and scientific research is vital which could be the key success to achieve such targets. A society that adopts the concept of developing its knowledge succeeds in raising the intelligence level. With efficient management, the goal becomes a fact.

b) *Mobility:* Most researchers focus on mobility with the aspect of traffic. Although this is one of the most crucial factors, however, it should not be the only mobility focus. Mobility is a crucial factor for emerging smart cities. The main discrimination between mobility and moving to smart mobility resides in general to the ability for easy access to the required information in a real time manner. The information continuous availability and easy access genuinely leads to improving services, saving time and effort, as well as raise the satisfaction level. Main factors that initially affect mobility could be the

success in developing sustainable plans, identifying measures to reduce conjunctions (for example, reduce number of simultaneously moving cars in the city), a successful management plan, and deploying alternatives for the services. Focusing on the mentioned example, one of the vital factors for reducing the number of cars is the online information availability without the need to move to the information source. This could be highlighted as a clear example for the effectiveness of the real time management system.

c) Sustainability: Sustainability is a crucial factor for smart cities which is in focus for many researchers. Most of the researchers' objectives are to reduce energy consumption and CO₂ emissions. The balance and economic usage of energy sources are also effective for sustainability. Moreover, seeking optimal sources' usage, developing a green environment, recycling, intelligent grid management, and developing supportive policies to reach these goals are also different pillars for efficient sustainability. For example, water sources management sustainability is challenging which could be achieved through intelligent technologies such as IoT technology.

d) Environments: Transforming to smart environments is a significant factor for smart cities. Green environment, water margins, controlling gas emissions, controlling wastes, and efficient energy management are some of the smart environment-affecting factors. Different intelligent solutions could be proposed. Gasses emissions prediction could contribute to gasses control which also leads to maintaining a green environment. Gasses emissions prediction is related to monitoring different activities such as traffic. As mentioned earlier, these factors are interrelated, therefore, in this example, it is clear that smart traffic systems could lead to minimizing the gasses emissions.

e) People: As smart cities main aspect is increasing the quality of life, therefore, People are considered the core factor of smart cities. They are the main players who have the benefits of the offered services and devices. Living in smart cities requires people to accept and be able to utilize this environment. On the other hand, securing people privacy as well as facilitating the usage of offered services are two of many factors that lead to the required level of acceptance. Educating people about the smart cities aspects is another factor. Their growing knowledge of how these services facilitates their lifestyle reduces their risk level and raise the level of trust. One of the emerging educating sources for smart cities is the social networks. Since connection is the key aspect for the concept of smart cities, therefore, considering all communicating channels should have given attention. Social communication among people should be utilized to exchange experiences and share opinions. Moreover, sharing resources, data, and benefits could also be intelligently performed with maintaining people privacy. Smart people as a main pillar of smart cities should not only share services, but they should also share data and knowledge targeting to ensure long-term benefits. For example, sharing data and knowledge about the city roads may lead to more enhancement in the roads infrastructure which leads to enhancing the traffic services. Smart people should have skills

such as being easily adaptable, accepting creativity, flexible, ability to participate in the community, able to identify their goals, as well as having knowledge about the rules.

f) Living: Smart living is comprised of the living in smart surroundings including smart buildings such as smart educational institutions, smart healthcare system, and smart tourism services and places. An example of the facilities that facilitates smart living in houses could be the contribution of IoT in home appliances, voice commands, and others. IVR is proved to be one of the useful solutions for smart living, especially being easy to learn and access. Different focuses have been performed to many other contributing systems to smart living such as healthcare system which is a non-negotiable one of the most vital systems in the community. Real time monitoring systems could contribute to various aspects such as healthcare, transportation, educational, and many other systems. Such innovation in the different business models reflects the required levels of management, leads to a high-quality level, provides a high value for the cooperative environment and contribute with a high impact for smart living concept.

g) Governance: The government is the core vital barrier that is able to promote smart cities concept to the people. The success of smart cities relies on the success of the government in providing the smart services, maintaining the relating channels, issue the suitable laws and regulations for maintaining privacy, fair facilities usage, and other aspects. E-government services should be one of the investment targets targeting to contribute to prosperity, satisfaction and enabling organizations. As the impact of social networks is already mentioned, governments could utilize such active sources as an effective strategy to motivate people in contributing to the smart cities concept. Achieving such a goal also relies on the success of maintaining security and privacy. Although technological aspects are key factor for e-government, however, managing policies are also vital to enable people to use the offered services but with regularities. Different metrics could raise the level of trust between the government and the people including the services effectiveness, citizen engagement, the process transparency, and the clear collaboration. Sustainable governance could be supported by effective environments such as cloud computing systems. This environment could heavily support the success of such framework due to the continuous engagement, communication, and collaboration.

2) Safety Component: Smart cities platforms need to maintain security requirements for both services as well as data. However, supporting security could be performed in a flexible environment to accommodate the needs' variations. Users' integrity, accounts' authentication, confidentiality, and trustiness are required for maintaining the platform protection as needed [18] [19].

a) Protecting Data Privacy: The privacy protection for all levels of data should be addressed starting for the lower level represented in the infrastructure of the proposed platform to the higher level representing the offered applications for the

people. One of the main methods is encryption. Encryption supports securing data while migrating as well as while residing in the physical data repositories. It is also vital to support the system against breaching by unauthorized users. Therefore, security procedures should be clearly defined and maintained on a regular basis. This task is a non-trivial task since data providers and consumers may engage with third parties to access data sources. Therefore, policies should be professionally managed for the success of controlling the access of data and its sources. All the contributing parties should comply with policies that maintain data privacy.

b) Devices Privacy: The variety of the services contributing to the smart cities leads to an impossible situation of identifying a unified security method for all devices as each device has its nature and capabilities and requires its own security procedure. On the other hand, balancing the equation of securing the contributing devices on one hand, and reducing resources consumption on the other hand makes this task more critical. Smart cities system should provide end-to-end system for security; for example, starting with the API level for IoT devices; moving forward to apply authentication and integrity methods. The boundaries of such critical systems should be governed by a set of policies and the boundaries should be clearly identified.

3) Sensing Component: Sensors have a key role in smart cities. They gather data, then transfer this data to its corresponding objects. This data also moves through the network layer to the cloud in an agile manner targeting to a fast response and immediate message passing for reply. With the expected network load, several paradigms could participate in the framework to avoid congestion. Load balancing techniques could be utilized to avoid network servers' overload. Data handling and resource allocation algorithms could be utilized with a cloud monitoring system for homogenous resource allocation. Resources allocation and task scheduling techniques support the assignment tasks in the application level which ensures a prominent level of service quality. Moreover, the dynamic computation complexity at the sensors level should be considered to ensure resource provisioning.

4) Connected Systems (IoT Management Agent): The main idea of the IoT paradigm is enhancing the working processes through the ability to receive instant services that are transformed from physical objects in a real-time manner [20]. IoT is basically a communication between objects either these objects are machines, utilities, devices, or human. RFID, sensors, and embedded systems are the main players in the IoT system. It is a fact that IoT paradigm is one of the most challenging evolutions in digitalization phenomena. This evolution has moved the field of digital business into a new level of competence, efficiency, and effectiveness. IoT offers an improved business models with the ability for cost reduction.

The current architecture for IoT systems suffers from the lack of agility for services. This situation highlighted the idea of embedding the agile concepts and practical aspects into IoT systems architecture [21]. Agile strategies support business

processes with maintaining optimization in a real-time management manner which consequently supports many opportunities to successful business solutions [22]. The current framework proposes a flexible agile-based IoT architecture which helps in the ability to accommodate with the dynamic environment of the smart cities. The proposed architecture considers the database security over the cloud environment with the availability for end-to-end access under the agility umbrella for more efficient computation and ensure high scalability to ensure leveraging the resources allocation. IoT platforms perspectives either focus on the objects or on the internet, the proposed architecture adopts a modular perspective targeting the services quality with maintaining agility for higher satisfaction rate.

It is a fact that technologies emerge quickly which requires continuous adaptation in a fast and flexible manner. The proposed framework has loosely distributed components to provide a flexible approach that permits components replacement with nearly non-impact to other components. Following the proposed approach minimizes the risks associated with the traditional IoT architecture deployment and ensures the multi-operating machines with maintaining the working scale. One of the fundamental aspects for the proposed framework is the concept of encapsulating the development issues and its capability of integration flexibility of trusted services with establishing the sufficient data repositories that have common basics for the smart cities. The proposed architecture ensures having a reliable scalable system with future adaptation capabilities. The recommended platform should have vital specifications to ensure the digital connection between all systems as well as ensure the high system scale with maintaining privacy and security. The factor of the platform interoperability and trustiness ensures the system stability with addressing the economy benefits from different perspectives.

The proposed architecture (in Fig. 1) illustrates the main components in a layered approach with demonstrating the corresponding standards. The following principles have been taken into consideration.

- The proposed solution is suitable for large, medium, and small cities based on its simplicity and adaptation availability.
- The verities of deployment solutions could be applied by the proposed architecture.
- The main international standards have been considered.
- Common features have been presented to ensure the generality perspective.
- Modular approach to ensure the architecture technology adaptation.
- Enabling interoperability by employing open API with semantic data interpretation.

5) Acquisition Layer: This layer considers gathering the required data from various sources. Data streams are produced from edge devices, it is then processed on the go during migration through the network to avoid delays. The process involves the physical devise which is the closest to the

determined data source and is responsible for the basic aspects of data processing and primary analyzing data. The analysis results are then migrated through the shortest channel with the real-time conditions to boost the cloud center for the data and initiate the decision-making alternatives.

B. Transmission (Network) Layer

The network layer provides the required support for both data and applications. The associated network servers and the sensors manipulate the data starting from gathering to processing over the cloud environment. The network layer task is to maintain this process quality including the delivery time and ensuring the satisfying of the business requirements with the required performance. This situation raises the need of embedding the agility concept in the network architecture especially for the IoT architecture. Network infrastructure sensors' task is continuous monitoring and recording the parameters. Detection stations keep collecting data as active nodes for all conditions such as weather, power, chemical, and other data.

To achieve the required goal, several requirements should be accomplished. The network should be able to provide the available services to multiple tenants. The network scalability should be of acceptable level with respect to the number of players and devices, traffic scale, bandwidth and other parameters. Most of the entities require either L2 or L3 VPN with security approaches for IoT devices such as isolation. Services should be continuously available to all players and contributors; therefore, the network should adopt the redundancy approach at all the network layers in order to avoid possible failure at any time and ensure instant re-convergence. The contributing devices should have the ability to tolerate under the extremes of the environmental parameters. To accomplish successful operations and other previous requirements, it is critical to adopt simplicity in the network operations.

IoT allows disseminating information for the purpose of enhancing business processes. A continuous communication between the participating machines in the network is performed through the IoT devices. Related technologies such as RFID, sensors, and embedded systems need to be critically selected for higher scalability. Engaging IoT for smart cities ensures productivity, safety, and quality. However, the traditional systems are lacking agility in supplying the system services. Therefore, introducing agile manifesto into the IoT system provides a real-time adaptation to the devices' management process, ability for optimal execution, and ensure optimizing services. The interconnectivity between IoT and agility provides the smart city architecture with massive successful opportunities. Introducing agility with its main concepts of simplicity, flexibility, adaptation ability, incremental approach, and refactoring provides an opportunity for continuous adaptation for the business services with ensuring high competing level. Leveraging the agile environment for IoT with the cloud platform is a triple based approach for smart cities which can provide high performance services for people. In the proposed approach, the internet based IoT approach is followed as the main focus is to provide scalable services for users.

In order to highlight the advancement of agile based IoT in smart cities over traditional IoT systems, the following points could be highlighted. Usually, the devices are connected through the sensors with transferring the data and readings. By embedding agile-based IoT cloud platform, these records are instantly updated through the cloud platform and this update is considered in a timely manner. The traditional system usually suffers from performance issues which hinder the ability to promote the services that has higher demands. The proposed architecture ensures a high performance. The traditional architecture does not adopt defined standards, rather, each pilot project has its own individual standards. The proposed architecture ensures the ability to adopt the same standards and ensures the concept of universal standardization. The traditional architecture suffers from excessive cost as a high number of supportive devices such as sensors require to be embedded in many places to ensure good coverage and continuous communication. On the other hand, the proposed system extensively reduces the cost according to the domination of cloud environment which can be accessed irrespectively of time or place with the lowest communication cost. There are no quality standards in the traditional system while in the proposed system, as agile manifesto is introduced, then following quality standards could be considered as one of the main parameters on focus.

C. Processing Layer

1) *Semantic Component*: Integrating semantic concepts into smart cities platforms is one of the emerging approaches [23] [24]. IoT services that are based on semantics are introduced in some research [24]. The research in [25] proposed a method for processing with semantic indexing while in [26], the concept of semantic IoT is introduced but was limited to the IoT applications only. The semantic component is based on multi-level analysis of data and explored knowledge. First level is for infrastructure data, then the data gathered from the IoT contributed devices is considered the second layer, the third layer includes the associations between players through the communication channels, business, social, and other available channels. Embedding the semantic component provide more meaning to the distributed data through the network. Moreover, applying machine learning techniques could contribute to enhancing the data quality through avoiding the data sparsity, outliers, and other data challenges. Machine learning techniques are well known for their ability to successfully deal with such challenges [27]. Semantic component deals with data in a collaborative environment. Data relationships are explored by stamping these data with the semantic annotations representing the relationship nature, time stamps, the defining factors explanation and others. Feature connectivity and weighting also contribute to the semantic explanation process [28]. Semantic components are responsible for providing meaningful explanation. For example, if a request for traffic in a certain territory is initiated, then the data is collected, aggregation is applied from various sources layers, and a virtual

entity is created. The features vector is explored, and the answer of the user query is provided through the predicted result.

2) *Data Analysis Component*: The data analysis component utilizes a set of models for processing data. A model repository should be included with a supportive architecture for identifying the suitable model for each received data. The model set could be divided to various categories such as statistical, learning [29], optimizing, forward and backward sensing, and forecasting. The availability of a variety of models highlights the opportunity for the analysis of different data nature with various levels of complexity [30]. The selection of the suitable model is based on the user requirements, the available data, and the purpose. The main goals of the data analysis component are to identify the exploration process that enables a selective set of data models to support both systems and applications interoperability between the interrelated communities. Moreover, classifying the available data models according to sectors and interoperability ability is on focus. Replicable models are accepted, even recommended for different sectors. Additionally, the higher capability of data volume manipulation that data models could consider provides a higher level of trust. This supports the system applications to reach their requirements and ensures higher efficiency and effectiveness. The different data models can play a communication role among different communities. Therefore, one of the main success factors is the clear and well discrimination of different data models manipulation methodology as it provides a level of trust for data migration among communicate. Moreover, as different data formats and structures are usually required by different application wither structured, semi structured, or unstructured data, therefore, a variety of different analysis and storage methods for different data sources formats as well as introducing analysis methods for data formats transformations could provide successful support.

3) *Context Data Management Component*: This component provides a set of management plans for data. These plans are initiated according to the integrated data and the different comprehensive understanding perspectives to the data. Understanding data is accomplished through gathering data from various sources, integrating these data, applying methods for semantic data explanation which context is received from these sources. The meta-data supports the comprehension task according to the associated entities and their functionalities. These entities gather events and migrate these events' data for the required explanation. The required pillars to perform this process efficiently is the data availability, accessibility, usability, and sharing ability. This component is an enabler for the applications to be able to explore their relevant data and be able to apply enquires over this data. Although the heterogeneity and diversity of data could be an obstacle, however, the structured architecture of the integrated data lakes enables the data migration among different applications.

Context validation should be also applied to confirm the operability and data validity.

4) *Data Management Component (Agile-based component)*: Management of data is accomplished through an identified standards API. This direction provides the ability for the re-using facility as there are common solutions standards. The ability for re-use also ensures blocking the lock-in issue which consequently sets the availability for continuous monitoring and improving ability. The re-use ability could also be supported by presenting a common taxonomy which is pre-identified for the data and the services which directly leads to ensure standardization. Tracking the updates is accomplished through the identified APIs which supports the access simplicity, resources accessibility while avoids issues of inconsistencies.

5) *Data Storage Management Component*: The aim of this component is to provide continuous and easiness in the data access ability through the services that are supported by AI technologies [31], IoT devices, and communication channels. This architecture provides an easy low-risk environment for data access through contributed cities. As digitalization is already established universally. One of the keys for successful data storage and migration is the ability to ensure the continuous accessibility to the data sources with nearly zero-risk through the infrastructure's owners. Then, re-using and sharing data provides a healthy environment for the different applications for common solutions, fair competition, minimum risk, and consequently supports sustainability. One of the key issues is the data privacy and security which is crucial aspects. Data usage agreements should be established. Consequently, this enables the required communication among providers and consumers and facilitates ensuring the prevention of data misuse and protects data flows in the system. However, the contributing entities should be able to control their data accessibility and valid process.

On the users' level, managing data is accomplished through their own controlling attributes. Users set their own rules in sharing their data, services, or applications. Trustiness is one of the key aspects for personal- management. Users have their security credentials which enables them to protect their data from others including infrastructure providers. Users have their right to set the access level and scope for others, manipulate their own data and use the agreement when needed. Users' authentication is a critical aspect to consider. Each user must have his own credentials. Cloud environment is known to be a flexible easily accessible system in general, however, data could reside in an on-premises cloud based system to ensure the required data protection. On the other hand, thin/thick client-based cloud architecture could also provide the balance between flexibility, accessibility and privacy.

D. Interoperability and Layout Layer (IoT Data Market Place)

Expanding the system scaling should be available as the more users intervening the system or the more devices

contributing to the system resulting the raising of data streaming. To successfully manage the increase in data scaling, additional network nodes should be added according to the need. On the other hand, additional storage repository and additional memory should be on focus to satisfy the increase in the computation according to the new requirements. One of the key aspects is the ability for the community contributors to have access to the system resources and be able to perform their needed functionalities, therefore, agility paradigm could be one of the main successful key aspects to satisfy this requirement. Moreover, as the contributors in the IoT system is the devices, sensors, and the gateway, the communication between these parties can be accomplished through many paths with different standards and protocols. Flexibility is required for updating the platform contributors, communication patterns, and possible changes. Communication patterns could be one-way from the devices to the gateway, information requests from devices, transferring messages through the system, or accomplish a determined task by one of the devices.

In order to ensure the architecture success, there is an immense need to ensure some factors such as interoperability, easiness, usability, and availability. As proposing a good design is one vital factor, however, this design should be efficient and effective for successful operation. Entities have their own data and naturally use different data models and different methods to process this data. Therefore, the need for unified standards has become critical. Interoperability is supported by applying public network standards and open protocols which organize the flow of data and information through gateways and APIs. The data and information migration between the network components should be according to these protocols and set agreements. Accordingly, new contributors could easily detect these gateways and APIs, and the integration is performed with the required agile approach. Accordingly, practicing such environment should be easily performed.

Pivotal points identify the Setting the system pivotal points should be intelligently established to connect between the constructed smart city system and the external environment as well as the system components themselves including devices, sensors, and the different smart management systems such as traffic, e-market, and others. It also controls the data migration through the system. Successful identification of the pivotal points supports the concept of reusability wither within the system on focus or other similar systems. Concrete identification of the level of coupling in the system is critical, from one hand, tight coupling is secured but hard to change and higher expectations for failure while from the other hand, loosely coupling systems is more flexible but high securing risk. Pivotal points could contribute enable integration of the different architecture components.

E. Applications (Services) layer

This layer provides the set of services and applications to the individuals. As discussed by much previous research, various applications have emerged such as smart grid, smart transportation, smart environment, smart living, smart health, smart energy, and others. Significant obstacles have emerged in the offered smart city applications. For example, the service provider can illegally access the data of the people living in smart cities such as medical or financial data. Moreover, smart

mobility can use different techniques to detect the users' patterns which could expose users to risks. This situation arises the vital impact of safety component and highlights the immense need for providing the protection layer over both data and application layers. The applications quality by logic depend on the quality of the previous layers, however, the enhancements of the services relies on a set of pillars including the ability of integrating services, how users can depend on these services, their quality level, ability for expansion and updating, and the services' standardization level. The evolution of embedding agility to the provided services can raise the guaranteed level of these services as it ensures higher performance and consequently higher satisfaction.

F. Monitoring and Governance Layer

Continuous monitoring should be performed to confirm the applicability and continuous operating to the smart cities' architecture. Identification of all contributors either people or devices should be clearly accomplished. Moreover, on the user level, a clear monitoring of the user activities and even expectations of his future activities should be performed. As this is a key success factor, however, it is not an easy task to perform. Several intelligent machine learning, data science, and artificial intelligence techniques contribute to this task in order to be successfully performed [32]. The concept of governance is applied to both data and applications. It is not considered on simple management tasks, rather, it is considered about providing a procedural action of continuous protection and tracking [33]. The following factors should be continuously monitored.

Resilience, failure could occur to devices, applications, or network components. Therefore, a self-healing system is recommended to avoid failure complications and ensure resilience. Self-healing includes many procedures such as adopting redundant links, continuous monitoring to IoT devices and the interaction between different components and users.

Performance, the system performance should be maintained through guaranteeing the instant response to the real time users' interaction. Applications should be continuously available and efficiently responding. Continuous automation to testing scenarios should be applied. Moreover, upgrading plan should also be on focus. Licenses should be available and complete authorization is expected.

One of the main factors is the feedback monitoring and recommendations to various aspects in the architecture. This has become a well-known principle for enhancements and avoid acceptance failure.

G. Enablers of smart cities

Several factors could affect the progress of smart cities implementation, some of these factors are discussed in this section (Fig. 3). Governmental support is one of the most effective enablers for smart cities construction. The government should be able to remove any arising barriers that hinder this change. Sustainability is another enabler which support the continuous development with the ability to gain public trust and economic stability. Approaches recommendations availability to reduce change resistance is another enabler. Fund availability is one of the critical enablers to be able to provide the

requirements of constructing smart cities. Agility adaption in the digital environment of management, marketing, and other services with courage in adopting the change could provide more innovations and the required changes could be accomplished with minimal risk.

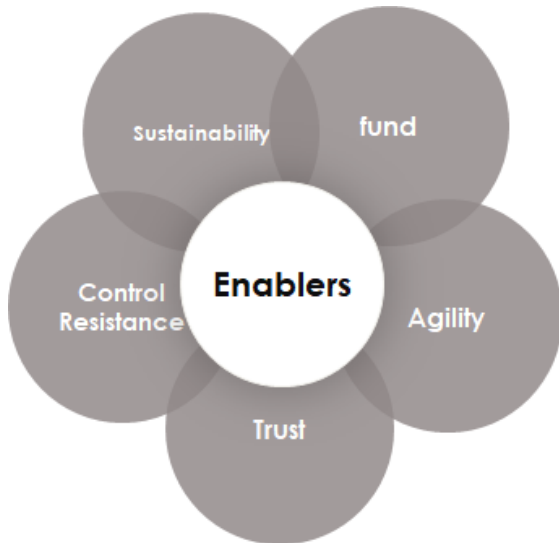


Fig. 3. Smart city enablers.

H. Challenges on focus

There are different challenges directions that should be highlighted. First, the data sources and formats heterogeneity as it is collected from different data sources. Moreover, data characteristics such as velocity, volatility, variability, veracity, and value should be determined and evaluated. The method to exchange the required knowledge by the players should be intelligently performed. Storage repositories access could be one of the major challenges in the platform especially for the governmental assets. Data standardization is another challenging aspect with the storage diversity. The construction cost debate with the return on investment is challenging as most of the return aspects are quality perspective. Ensuring the new paradigm acceptance by the users in the cities who are represented in the citizens should have a robust plan and complete governance support.

V. APPLICATIONS SAMPLE MODEL

A vast set of domains has seen the development of intelligent applications. These applications are not yet widely accessible; however, initial research suggests the potential of IoT to enhance the quality of life in our society. IoT applications are utilized in home automation, fitness tracking, health monitoring, environmental protection, smart cities, and industrial environments. Focusing on residential automation as a sample, smart houses are gaining popularity nowadays for two reasons. The sensor and actuation technologies, together with wireless sensor networks, have substantially advanced. Secondly, contemporary individuals exhibit trust technology to mitigate their concerns regarding quality of life and home security (Fig. 4).

Smart homes utilize an array of sensors that deliver intelligent and automated services to users. They facilitate the automation of daily routines and assist in establishing a routine for persons prone to forgetfulness. They facilitate energy conservation by automatically deactivating lighting and electronic devices. Motion sensors are generally employed for this purpose. Motion sensors can also be utilized for security purposes. An intelligent agent is offered that employs diverse predictive algorithms to do automated chores in reaction to user-initiated events and adjusts to the residents' routines. Prediction algorithms are employed to forecast the sequence of events in a household. A technique for sequence matching preserves event sequences in a queue while simultaneously recording their frequency. A prediction is subsequently generated utilizing the match length and frequency.

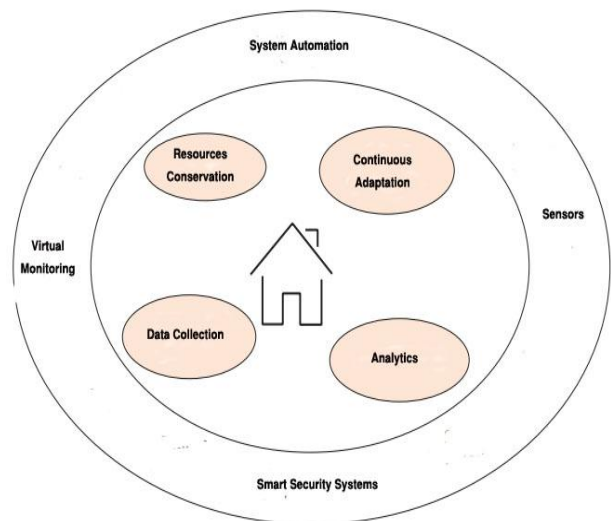


Fig. 4. A proposed block diagram of adapting smart home system.

Other algorithms employed by analogous applications utilize compression-based prediction and Markov models. Energy conservation in smart homes is often accomplished by sensors and contextual awareness. The sensors gather data from the environment, including light, temperature, humidity, gas, and fire incidents. The data from diverse sensors is transmitted to a context aggregator, which relays the gathered information to the context-aware service engine. This engine chooses services according to the circumstance. An application can autonomously activate the air conditioning when humidity levels increase. Alternatively, in the event of a gas leak, it may extinguish all the lights. Smart home applications are highly advantageous for the elderly and individuals with disabilities. Health is checked, and families are promptly notified in crises.

The floors are fitted with pressure sensors that monitor an individual's mobility within the smart home and assist in identifying if a person has fallen. CCTV cameras in smart homes can record significant occurrences. These can subsequently be utilized for feature extraction to ascertain the underlying phenomena. Fall detection applications in smart environments are effective for identifying instances where elderly individuals have fallen. Fall dynamics is identified by evaluating motion patterns and also detects idleness, comparing

it with previous activity levels. Neural networks are utilized, and sample data is sent to the system for various categories of falls. Numerous smartphone applications are available that detect falls using data from accelerometers and gyroscopes.

Numerous obstacles and issues pertain to smart home applications. Security and privacy are paramount, as all data on occurrences occurring within the home is being documented. If the system's security and reliability are not assured, an attacker may compromise the system and induce harmful behavior. Smart home systems are designed to alert owners upon detecting anomalies. This is achievable with AI and machine learning algorithms, and academics have commenced efforts in this area. Comparing the sample model with the literature [3, 20, 29, 34], most of the researchers focused on one enhancement direction; including cost, processing, and security; with no ability for a complete vision.

VI. CONCLUSION

The research proposed a complete generic architecture of smart cities. Smart cities consider allowing both people and governance to benefit from technology. The concept of embedding intelligence into the cities' aspects requires ensuring that the entire process could be altered. One of the main pillars is the data availability for the required level of quality and sustainability. Therefore, the proposed architecture presented all required aspects for a complete transformation to the digital smart city including infrastructure, processes, data management, players, roles, enablers, and services. The research highlighted the enablers for the architecture as well as the SWOT analysis for implementing the proposed architecture. Finally, challenges for the transformation process are discussed. It is expected that the proposed smart city architecture is encouraging and could move the field from the pilot individual projects to the standardization model which consequently provide a universal perspective to the smart cities concept rather than having different understanding to the same concept in the pilot projects. A main future consideration is to apply the proposed architecture and evaluate the transformation process in a real-life example. Another future direction is to provide more details and recommendations for practical development to each component.

ACKNOWLEDGMENT

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant No. (UJ-23-DR-15). Therefore, the authors thank the University of Jeddah for its technical and financial support.

REFERENCES

- [1] A. Kirimtata, O. Krejcar, A. Kertesz and M. F. TASGETIREN, "Future Trends and Current State of Smart City Concepts: A Survey," *IEEE Access*, vol. 8, 2020.
- [2] S. Shaker, M. Tamer, A. E. Khedr and S. Kholeif, "A Proposed Framework for Reducing Electricity Consumption in Smart Homes using Big Data Analytics," *Journal of Computer Science*, vol. 15, no. 4, 2019.
- [3] H. Samih, "Smart cities and internet of things," *Journal of Information Technology Case and Application Research*, vol. 21, no. 1, 2019.
- [4] S. Bresciani, A. Ferraris and M. Del Giudice, "The management of organizational ambidexterity through alliances in a new context of analysis: Internet of Things (IoT) smart city projects," *Technol Forecast Soc Chang*, vol. 136, p. 331–338, 2018.
- [5] Y. Qian, D. Wu and W. Bao, "The internet of things for smart cities: Technologies and applications," *IEEE Network*, vol. 33, no. 2, pp. 4-5, 2019.
- [6] Y. Chen, S. Hu and H. Mao, "Application of the best evacuation model of deep learning in the design of public structures," *Image Vis Comput*, vol. 102, no. 103975, 2020.
- [7] A. Watson, Z. Musova, V. Machova and Z. Rowland, "Internet of things-enabled smart cities: big data-driven decision-making processes in the knowledge-based urban economy," *Geopolitics History and International Relations*, vol. 12, no. 1, p. 94100, 2020.
- [8] Z. Lv, X. Li, W. Wang, B. Zhang, J. Hu and S. Feng, "Government affairs service platform for smart city," *Future Generation Computer Systems*, vol. 81, pp. 443-451, 2018.
- [9] S. A. Hossain, M. A. Rahman and M. A. Hossain, "Edge computing framework for enabling situation awareness in IoT based smart city," *Journal of Parallel and Distributed Computing*, vol. 122, pp. 226-237, 2018.
- [10] G. Javadzadeh and A. M. Rahmani, "Fog computing applications in smart cities: a systematic survey," *Wireless Networks*, vol. 26, p. 1433–1457, 2020.
- [11] A. Giannakoulis, "Cloud computing security: protecting cloud-based smart city applications," *Journal of Smart Cities*, vol. 2, no. 1, 2017.
- [12] S. Naiem, M. Marie, A. M. Idrees and A. E. Khedr, "DDoS Attacks Defense Approaches And Mechanism In Cloud Environment," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 13, 2022.
- [13] P. Su, Y. Chen and M. Lu, "Smart city information processing under internet of things and cloud computing," *The Journal of Supercomputing*, vol. 78, 2022.
- [14] S. Choenni, M. S. Bargh, T. Busker and N. Netten, "Data governance in smart cities: Challenges and solution directions," *Journal of Smart Cities and Society*, vol. 1, 2022.
- [15] O. Embarak, "Smart City Transition Pillars With Layered Applications Architecture," *Procedia Computer Science*, The 18th International Conference on Mobile Systems and Pervasive Computing (MobiSPC), vol. 191, 2021.
- [16] H. M. Elmasry, A. E. Khedr and H. M. Abdelkader, "Enhancing the Intrusion Detection Efficiency using a Partitioning-based Recursive Feature Elimination in Big Cloud Environment," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 1, 2023.
- [17] S. A. Almaqashi, S. S. Lomte, S. Almansob, A. Al-Rumaim and A. A. A. Jalil, "The Impact of ICTS in the Development of Smart City: Opportunities and Challenges," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, 2019.
- [18] S. Naiem, M. Marie, A. E. Khedr and A. M. Idrees, "Distributed Denial Of Services Attacks And Their Prevention In Cloud Services," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 4, 2022.
- [19] S. Naiem, A. M. Idrees, A. E. Khedr and M. Marie, "Iterative Feature Selection-Based DDoS attack Prevention Approach in Cloud," *International Journal of Electrical and Computer Engineering Systems*, vol. 14, no. 2, 2023.
- [20] A. M. Idrees and E. Shaaban, "Reforming home energy consumption behavior based on mining techniques a collaborative home appliances approach," *Kuwait Journal of Science*, vol. 47, no. 4, 2020.
- [21] P. Upadhyay, G. Matharu and N. Garg, "Modeling Agility in Internet of Things (IoT) Architecture," *Information Systems Design and Intelligent Applications - Advances in Intelligent Systems and Computing*, vol. 340, 2015.
- [22] N. A. Eldanasory, E. Yehia and A. M. Idrees, "A Literature Review on Agile Methodologies Quality, eXtreme Programming and SCRUM," *Future Computing and Informatics Journal*, vol. 7, no. 2, 2023.
- [23] A. A. Qaffas, A. M. Idrees, A. E. Khedr and A. S. Kholeif, "A Smart Testing Model Based on Mining Semantic Relations," *IEEE Access*, vol. 11, 2023.
- [24] N. Zhang, H. Chen, X. Chen and J. Chen, "Semantic Framework of Internet of Things for Smart Cities: Case Studies," *Sensors*, vol. 16, 2016.
- [25] Z. Li, C. Chu, W. Yao and A. Richard, "Ontology-Driven Event Detection and Indexing in Smart Spaces," *Proceedings of the 2010 IEEE Fourth International Conference on Semantic Computing (ICSC)*, 2010.

AUTHORS' PROFILE

- [26] X. Chen, H. Chen, N. Zhang, J. Huang and W. Zhang, "Large-scale real-time semantic processing framework for Internet of Things," International Journal of Distributed Sensor Networks, vol. 2, 2015.
- [27] S. Zaki, N. Ghali, A. Abo Elfetoh and A. M. Idrees, "Comparison of Four ML Predictive Models Predictive Analysis of Big Data," Journal of Theoretical and Applied Information Technology, vol. 101, no. 1, 2023.
- [28] A. M. Idrees, A. E. Khedr and A. A. Almazroi, "Utilizing Data Mining Techniques for Attributes' Intra-Relationship Detection in a Higher Collaborative Environment," International Journal of Human-Computer Interaction, 2022.
- [29] A. H. Z. Hassan, A. M. Idrees and A. I. B. Elseddawy, "Neural Network-Based Prediction Model for Sites' Overhead in Commercial Projects," International Journal of e-Collaboration, vol. 19, no. 1, 2023.
- [30] B. N. Silva, M. Khan and K. Han, "Big Data Analytics Embedded Smart City Architecture for Performance Enhancement through Real-Time Data Processing and Decision-Making," Wireless Communications and Mobile Computing, vol. 9429676, 2017.
- [31] N. Y. Hegazy, M. H. Khafagy and A. E. Khedr, "Big Scholarly Data Techniques, Issues, and Challenges Survey," Journal of Theoretical and Applied Information Technology, vol. 100, no. 5, 2022.
- [32] S. Zaki, N. Ghali, A. Abo-Elfetoh and A. M. Idrees, "Exploratory Big Data Statistical Analysis The Impact Of People Life's Characteristics On Their Educational Level," Journal of Theoretical and Applied Information Technology, vol. 100, no. 5, 2022.
- [33] S. Zaki, N. Ghali, A. Abo Elfetoh and A. M. Idrees, "Predictive Analysis of Big data in Egypt Census 2017 Comparison of Four ML Predictive Models," Journal of Theoretical and Applied Information Technology, vol. 101, no. 1, 2022.
- [34] M. Attia; A. H. Abed, "A Comprehensive Investigation for Quantifying and Assessing the Advantages of Blockchain Adoption in Banking Industry", 2024 6th International Conference on Computing and Informatics (ICCI), Egypt, IEEE, 2024.



Ayman E. Khedr
Professor

aeelsayed@uj.edu.sa

I'm currently a professor at the University of Jeddah. I have been the vice dean of post-graduation and research and the head of the Information Systems Department in the Faculty of Computers and Information Technology, at Future University in Egypt. I am a professor in the Faculty of Computers and Information, at Helwan University in Egypt. I have previously worked as the general manager of the Helwan E-Learning Center. My research is focused on the themes (scientific) data and model management, Data Science, Big Data, IoT, E-learning, Data Mining, Bioinformatics, and Cloud Computing.

Othman Asiry
Assistant Professor
asiry@uj.edu.sa

Dr. Othman Asiry is currently working as an Assistant Professor in the Department of Information Technology at Khulis College. His research interests encompass IoT, Medical Image Processing, Deep Learning, Machine Learning, and Speech Recognition.



Amira M. Idrees

Professor

amira.mohamed@fue.edu.eg

I'm a professor in information systems. I have been the head of scientific departments and the vice dean of community services and environmental development, at the Faculty of Computers and Information, at Fayoum University. A professor in the Faculty of Computers and Information Technology at Future University, the head of IS department, and the head of the University Requirements Unit. And a professor in King Khalid University. My research interests include Knowledge Discovery, Text Mining, Opinion Mining, Cloud Computing, E-Learning, Software Engineering, Data Science, and Data warehousing.

A Novel Optimization Strategy for CNN Models in Palembang Songket Motif Recognition

Yohannes, Muhammad Ezar Al Rivan, Siska Devella, Tinaliah

Informatics, Faculty of Computer Science and Engineering, Universitas Multi Data Palembang,
Palembang, Indonesia

Abstract—Palembang Songket is an essential part of Indonesian cultural heritage, and its introduction and preservation present challenges, particularly in recognizing various motifs. This research introduces a novel strategy to optimize the performance of Convolutional Neural Networks (CNNs) by presenting a hierarchical integration of Ghost Module operations and Max Pooling, referred as Ghost Feature Maps. While the Ghost Module is effective in reducing parameters and enhancing feature extraction, it has limitations in filtering irrelevant information. To address this shortcoming, we propose a hierarchy in which Max Pooling works in conjunction with the Ghost Module, strengthening its performance by not only extracting dominant features but also eliminating excess, non-essential information. This hierarchical design enables more efficient feature extraction, thus enhancing the model's recognition accuracy. By combining Ghost Modules and Max Pooling in a structured manner, this approach advances established methodologies and offers a new perspective on feature representation within CNN architectures. Utilizing a dataset of 10 augmented classes of Palembang Songket motifs totaling 1000 images, we conducted experiments using varying ratios of Ghost Feature Maps. The results indicate that a ratio of 2 achieves an impressive accuracy of 0.98 with minimal parameter reduction. Additionally, a ratio of 3 results in a 34% decrease in parameters while maintaining a competitive accuracy of 0.95. Ratios of 4 and 5 continue to demonstrate robust performance, achieving accuracy levels of 0.93 while delivering over 60% reductions in model size and parameters. This research not only contributes to the optimization of CNN architectures but also supports the preservation of cultural heritage by improving the recognition capabilities of Palembang Songket motifs.

Keywords—Convolutional neural network; ghost module; Palembang songket motif; recognition

I. INTRODUCTION

One of the artistically significant pieces of Indonesian cultural heritage is Songket. The term "songkit" which describes the process of embroidering gold and silver threads, is the source of the word "songket" [1]. Ten connected steps are involved in the production of Songket fabric, including thread dyeing, klose processing, lungsin coating, thread type selection, and weaving designs with lidi. Songket fabric has different characteristics and philosophies depending on its region of origin [2]. One of the Songket fabrics registered as an Indonesian Intangible Cultural Heritage is Palembang Songket [3]. Palembang Songket has various types of motifs. The motifs of Palembang Songket reflect its beauty, uniqueness, and traditional values. Palembang Songket faces challenges in its preservation because the motif recognition process still relies

on manual or semi-automatic approaches that are vulnerable to human error and limitations. Successfully recognizing and understanding Palembang Songket motifs is essential for preserving local art and culture and has significant economic impacts through promoting Songket products in the global market.

The problem of identifying Songket motifs has been addressed in the past by a number of feature extraction techniques, including Felzenszwalb segmentation [4], and Gray Level Co-occurrence Matrix (GLCM) [5]. These techniques are then combined with a variety of classifier algorithms, including Naive Bayes [6], Decision Tree [6], and Support Vector Machine (SVM) [5]. Even while these techniques have produced acceptable outcomes in certain situations, there are still a number of important drawbacks. The primary drawback of these methods is their dependence on manual feature extraction, which frequently results in challenges in capturing the crucial elements of the complex Songket motifs.

Enhancement of motif identification performance is a promising area in deep learning. Convolutional Neural Network (CNN) models have demonstrated their capacity to recognize patterns in a variety of domains, including the identification of images. CNN has a lot of potential to improve Songket motif recognition. Applying CNN to motif recognition has the benefit of minimizing reliance on manual feature extraction by automatically extracting pertinent characteristics from data. Although there has been some prior research, not much has been accomplished in terms of training CNN to identify Songket motifs [7], [8]. The primary constraint is to the model's capability to manage intricate motif modifications, such as rotation, scaling, and distortion. The intricacy of Palembang Songket patterns is too great for a conventional CNN training model to handle.

Due to their numerous convolution layers, CNNs have substantial computational costs. In each convolution layer, mathematical operations are conducted to each input in order to extract important properties from the input data. The number of convolution layers that are conducted therefore increases the amount of computing operations needed, which could result in significant computational expenses in terms of processing time and energy consumption [9], [10], [11]. Thus, the convolution operations of CNN can be assumed by the Ghost Module. By using a method called the Ghost Module, which involves joining convolution filters with smaller "ghost filters", the CNN model is able to retain a significant amount of representation information while requiring fewer calculations and parameters [12]. In order to improve the performance of Palembang

Songket motif recognition, this research suggests a novel method that incorporates the Ghost Module into the CNN architecture.

The main contribution in this research is the introduction of Ghost Feature Maps through the integration of Ghost Module as a substitute for conventional convolutional layers, with Max Pooling applied hierarchically afterward. Ghost Feature Maps function to facilitate feature learning in CNN models, with the aim of increasing the efficiency of Palembang Songket motif recognition. In this hierarchical approach, Ghost Module is applied first to improve feature extraction efficiency by reducing computational complexity. Afterward, Max Pooling is used to further reduce spatial dimensions, enhancing computational efficiency and focusing on dominant features while suppressing irrelevant information. This combination reduces the number of parameters and model size while increasing accuracy performance. In addition, this study offers a solution to the challenge of recognizing complex Songket motif variations, such as rotation, scale, and deformation. This approach provides an innovative solution in addressing problems that have not been fully resolved in Palembang Songket motif recognition.

The remaining content of the paper is organized as follows: Section II discusses the related works in the field of Palembang Songket Motif, CNN and Ghost Module. Section III presents a detailed explanation of the proposed method. Section IV presents results and discussion, and performance. The last Section V summarizes the overall conclusion of the paper.

II. LITERATURE REVIEW

This section discusses the literature review of the Indonesian traditional fabrics, CNN, and Ghost Module research.

A. Image-based Recognition of Indonesian Traditional Fabrics

Sriani, Hasibuan, and Ananda [5] used SVM to classify Batu Bara Songket motifs, which are characterized by distinctive patterns such as Bunga Tanjung, Pucuk Betikam, Pucuk Cempaka, Pucuk Pandan, Tampuk Manggis, and Tolab Berantai. Gray-level texture features were extracted using the Co-Occurrence Matrix method, considering parameters like Contrast, Correlation, Energy, and Homogeneity. These extracted features were then processed as input for classification using the SVM. Despite the challenges, the study achieved a classification accuracy of 57% with 60 training data and 30 test data.

Aprianti et al. [6] classified Lombok songket fabric motifs, which are characterized by geometric patterns, varying density, color, and motif positioning. The algorithms used included Naive Bayes and Decision Tree, tested with different pixel sizes to compare accuracy levels. The results showed that the Naive Bayes algorithm achieved the highest accuracy of 90% at a 100×100 pixel size, while the Decision Tree algorithm was optimal at 400×400 pixels with the same accuracy. This approach demonstrated that combining algorithms with pixel size adjustments could significantly enhance motif recognition accuracy.

Ariessaputra et al. [7] classified Lombok songket motifs using a Convolutional Neural Network (CNN) algorithm, demonstrating the potential of image processing for traditional fabric pattern recognition. The dataset consisted of 20 Lombok Songket images with identical motifs and colors, 14 with the same but different colors, and 10 with various motifs and colors. In the preprocessing phase, each image underwent resizing, followed by CNN layers for convolution, pooling, and fully connected operations. Data augmentation through 150-degree rotations was applied to enhance model robustness. The results indicated that motif classification with consistent colors achieved an accuracy of 84%, highlighting the effectiveness of CNNs for identifying and distinguishing Lombok Songket motifs across varying visual parameters.

Hambali, Mahayadi, and Imran [8] applied CNN to classify Lombok Songket motifs, focusing on Songket from two prominent Lombok regions, Sade and Pringgasela. The study utilized a dataset of 64 images, comprising 40 samples from Sade and 24 from Pringgasela. The model's testing results showed an accuracy of 86%, with 87% precision and 86% recall. These results demonstrated CNN's effectiveness in differentiating textures within traditional Songket fabrics, offering valuable insights for preserving and recognizing regional textile characteristics.

Andrian et al. [13] used CNN architectures, including AlexNet, EfficientNet, LeNet, and MobileNet, to classify Lampung Batik motifs. The study utilized a dataset of 1000 images representing ten distinct motifs, enhanced through preprocessing techniques like rotation, shifting, brightness adjustment, and zooming. The results showed that LeNet achieved the highest accuracy of 99.33%, highlighting its suitability for small datasets, while other architectures also demonstrated strong performance despite occasional classification errors due to motif similarities.

Elvitaria et al. [14] proposed an ensemble deep learning method for batik image classification that combines texture feature extraction using Gray Level Co-occurrence Matrix (GLCM) with the Residual Neural Network (ResNet) classification model. By extracting texture features such as contrast, dissimilarity, and entropy using GLCM and combining them with ResNet, the proposed ensemble method achieved high performance, with accuracy, precision, recall, and F1-score all above 90%. The study demonstrated that the ensemble deep learning approach, particularly with the standard deviation feature, improved classification accuracy and can be applied to preserve batik culture digitally.

Muliono, Iranita, and Syah [15] proposed a deep learning model for classifying traditional Batak Ulos fabrics, utilizing CNN to recognize and classify different Ulos motifs. The study employed the Modular Neural Network (MNN) to simplify complex computations, achieving an accuracy of 97.83% with a loss value of 0.0793 during training. The validation results showed a validation loss of 2.1885 and a validation accuracy of 74.29%, demonstrating the model's strong performance while indicating areas for potential improvement in generalization.

Overall, the studies reviewed highlight significant advancements in image-based recognition techniques for recognizing Indonesian traditional fabrics, such as Songket and

Batik. Various machine learning models, including SVM, Naive Bayes, Decision Trees, CNN, and ensemble deep learning methods, have been applied to address the challenges of recognizing intricate fabric patterns characterized by texture, color, and motif positioning. These findings underscore the potential of image-based recognition systems in advancing the digital preservation and recognition of traditional Indonesian fabrics, offering valuable contributions to the cultural heritage field.

B. Ghost Module and CNN

Han et al. [12] proposed the Ghost Module in their research to produce feature maps with low computational costs, thereby simplifying the architecture of conventional CNN. GhostNet, built using the Ghost Module, demonstrated better recognition performance than MobileNetV3, with an accuracy of 75.7% on the ImageNet ILSVRC-2012 dataset. According to these results, the Ghost Module has the potential to be an efficient solution that can be added to current CNN networks.

Wang and Li [16] discussed the creation of a CNN model to enhance recognition systems' efficiency. This study used the GhostNet and Convolutional Block Attention Module (CBAM) methods. GhostNet and Ghost Bottleneck improved the model's capacity to extract significant image characteristics. Furthermore, employing GhostNet resulted in fewer parameters while retaining good accuracy.

Zhao and Cheng [17] proposed a more efficient approach that minimizes computation compared to traditional CNN by merging the Yolov5 model and GhostNet. This method increases processing speed and accuracy. The proposed approach's performance test showed that it achieved high detection accuracy while needing less memory and compute. Overall, this study provides a solution for image identification for human security screening purposes. This framework is also done for vehicle detection [18].

Huangfu, Li, and Yan [19] proposes the Ghost-YOLO v8 algorithm to improve the detection of surface floating litter in artificial lakes. This efficient and lightweight algorithm includes an SE mechanism for better feature extraction, a small-target detection layer to reduce semantic loss, and a GhostConv module to decrease computational demands.

Fang, Chen, and He [20] suggested an efficient CNN solution for facial expression recognition called Ghost-based Convolutional Neural Network (GCNN). This approach seeks to overcome CNN-related overfitting concerns. The Ghost Module architecture is less computationally expensive since it may minimize the number of parameters while producing more feature maps than CNN approaches. Based on this research, GCNN can efficiently extract and classify face expression features.

Alansari et al. [21] introduced Lightweight Face Recognition using GhostFaceNets, which requires less processing than normal CNN models. GhostFaceNets is a very accurate and efficient face recognition system. The proposed method was tested using a variety of datasets, including LFW, AgeDB-30, IJB-B, IJB-C, and MegaFace. GhostFaceNets uses the Ghost Module method to execute linear modifications on

feature maps, resulting in better and more thorough feature extraction.

Luan, Mu, and Yuan [22] addresses challenges in Online Signature Verification (OSV), by proposing the one-dimensional Ghost-ACmix Residual Network (1D-ACGRNet). The network is designed to combine convolution with a self-attention mechanism to effectively capture both global and local signature features. Simplification of operations is achieved through the Ghost-based Convolution and Self-Attention (ACG) block, which reduces computational load. Significant accuracy improvements are shown in experiments on the MCYT-100 and SVC-2004 Task2 datasets, with equal error rates reaching as low as 0.91% for genuine and forged signatures.

Paoletti et al. [23] conducted Hyperspectral Image Classification (HSI) which is one of the remote sensing techniques used in Earth observation for health, robotic vision, and quality control. The challenge in HSI is that each HSI image has hundreds of spectral bands that produce large amounts of data, requiring high computation. This study introduces an approach by combining ghost-module architecture and CNN. Test results show that using Ghost module can reduce HSI's cost and computation time.

Tang et al. [24] proposed an efficient mechanism called DFC attention is proposed, using GhostNetV2. GhostNetV2 can overcome limitations in conventional CNN methods. In testing, GhostNetV2 showed better performance than CNN architecture, achieving an accuracy of 75.3% on the ImageNet dataset, with efficient FLOPs. Therefore, GhostNetV2 is the right choice for mobile applications requiring efficiency and high performance.

Liu et al. [25] investigated effective training strategies for compact neural networks to address performance gaps and proposed GhostNetV3. The strategy focused on essential methods, including re-parameterization to improve efficiency, knowledge distillation to enhance smaller model performance through learning from larger models, and optimized learning schedules and data augmentation to increase training data diversity. As a result, GhostNetV3 achieves an optimal balance between accuracy and inference costs.

He et al. [26] proposed the Ghost module-based convolution network approach for superresolution (SR) in satellite video in this research. This approach is called Ghost module-based video SR (GVSR) and consists of two main modules: the preliminary image generation module and the SR results' reconstruction module. Experimental results on Jilin-1 and OVS-1 videos show that this method is superior in quality and quantity to other Deep Learning methods.

Liu et al. [27] introduced a new approach for hyperspectral image classification called Ghost module extended morphological profile (GhostEMP), which employs Ghost Module and extended morphological profile (EMP) features. This method can reduce model complexity and the amount of calculations, hence increasing operational efficiency. The experimental results suggest that this strategy effectively preserves model performance by maximizing hyperspectral data features. Not only does it apply to hyperspectral data, but

it also utilizes the Ghost Features Network (GFN) for super-resolution by cascading residual-in-residual ghost blocks [28].

The studies emphasize the effectiveness of the Ghost Module in enhancing the efficiency of deep learning models for diverse applications, including image recognition and hyperspectral image processing. GhostNet and its variants, such as GhostNetV2 and GhostNetV3, have significantly improved accuracy while reducing computational complexity. These advancements make Ghost Module in GhostNet architecture a valuable solution for high-performing applications with minimal resource usage.

III. METHODOLOGY

A. Dataset

The dataset used in this research consists of 10 classes of Palembang Songket motifs, namely Bintang Berantai, Bunga Cina, Bunga Jatuh, Cantik Manis, Jando Beraes, Kenanga Makan Ulat, Naga Besaung, Nampan Perak, Pacar Cina, and Pulir. The motif images were captured using a Canon 7D DSLR camera equipped with a Canon EF 70-200mm F2.8 lens, Canon Speedlite 600 Mark II, and tripod, with a consistent portrait distance of 45 cm from the object and a front-facing angle of 0 degrees. A total of 50 Songket fabrics were photographed and thoroughly validated by a Songket motif expert. The expert ensured that each image was following the traditions and authenticity of Palembang Songket culture.

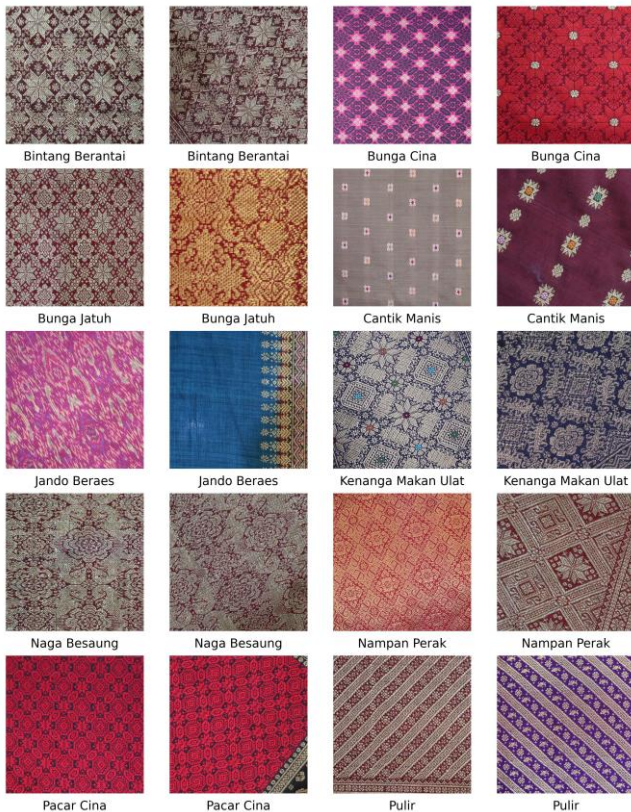


Fig. 1. Dataset of Palembang songket motif.

The Palembang Songket motif photos were cropped to 2048 x 2048 pixel size with 300 dpi resolution, and then augmented.

Augmentation techniques applied include rotation, scaling, and horizontal and vertical flipping, as these techniques are more suitable for the complex patterns of Songket and do not alter the basic motif design. This augmentation technique produces a total of 1000 images, with each motif containing 100 images that correspond to the Palembang Songket motif collection with geometric patterns, as shown in Fig. 1. Afterwards, the images were resized to 256×256 for use as input in the CNN architecture with Ghost Feature Maps. As part of the preprocessing stage, each Songket motif image was resized to 256×256 pixels to ensure uniform image sizes.

B. Proposed Method

The proposed model architecture can be divided into two major components, namely feature learning and classification, each responsible for different aspects of its functionality. Fig. 2 illustrates the architecture of the proposed model.

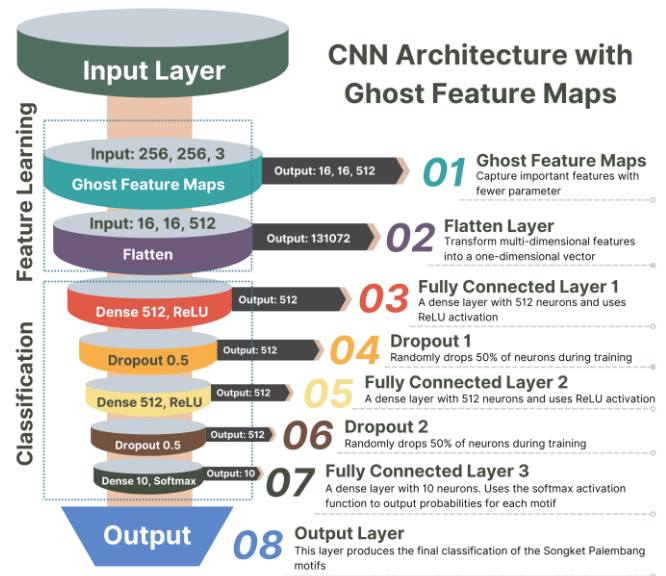


Fig. 2. CNN architecture with ghost feature maps.

1) *Feature Learning*: The process starts with applying Ghost Feature Maps. This layer is designed to extract feature maps efficiently by generating primary feature maps and applying simple transformations to produce additional maps. The objective here is to capture essential patterns in the data while minimizing computational resources and parameters, making the approach both efficient and effective. Following this, the extracted feature maps are processed by a Flatten layer. The Flatten layer transforms these high-dimensional feature maps into a one-dimensional vector, which is necessary for subsequent classification layers. The Flatten layer ensures that all the spatial information captured during the feature learning process is retained but represented in a format compatible with fully connected (dense) layers.

2) *Classification*: The classification phase includes three Fully Connected (Dense) layers. The first two Dense layers each consist of 512 neurons and use the ReLU activation function to introduce non-linearity, enabling the model to learn

more complex relationships in the data. These layers take the flattened feature vector from the feature learning phase and process it further to identify the patterns necessary for classification. To address the risk of overfitting, the model incorporates two Dropout layers after the first and second Dense layers. Dropout randomly turns off 50% of the neurons during each training iteration, encouraging the model to learn more robust and generalized features by not overly relying on specific neurons. The final step involves a third Dense layer, the output layer. This layer contains ten neurons, corresponding to the 10 Palembang Songket motifs being classified. It uses the softmax activation function to generate a probability distribution across the ten classes, with the class having the highest probability selected as the predicted motif.

C. Ghost Module

The GhostNet architecture, which features a layer known as the Ghost bottleneck, was the primary inspiration for this work. The Ghost bottleneck combines the Ghost module, a batch normalization, two or three Ghost Modules in a row, and then interleaved depthwise convolution, batch normalization, and ReLU activation [12]. However, in this research, only the Ghost module is employed without the full implementation of the Ghost bottleneck.

The dataset is a Palembang Songket pattern and it is unique and thus very rare, so it is hard to get a lot of data. Batch normalization was excluded from the Ghost Module. Instead, the Ghost Module was used alone, assuming that its simplicity would adequately convey the distinctive qualities of the Palembang Songket motifs without the added complication of the Ghost bottleneck.

The Ghost module consists of multiple phases, including primary convolution, cheap convolution, feature concatenation, and channel trimming. Each of these levels contributes significantly to the module's efficiency by reducing the amount of parameters and computational complexity while maintaining overall network performance. The Ghost module is a way to create more feature maps while doing fewer computations than usual, which is very helpful in deep learning models. Fig. 3 shows how the Ghost module works.

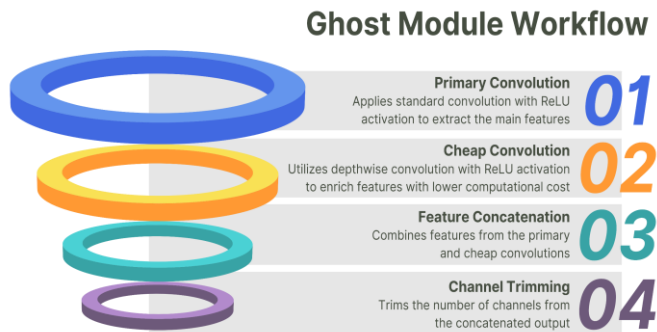


Fig. 3. Ghost module workflow.

1) *Primary convolution:* In the initial stage, the Ghost module conducts feature extraction through a primary convolution operation. This convolutional process decreases

the number of output channels by a defined reduction ratio, denoted as r , which indicates the degree of parameter reduction in comparison to a typical convolution operation. If the desired number of output channels is C_{out} , the output channels from the primary convolution, C_{cheap} are calculated in Eq. (1).

$$C_{cheap} = \frac{C_{out}}{r} \quad (1)$$

This convolution employs a kernel size of $k \times k$, where k represents the kernel dimension. It is applied with the same padding to maintain the input's spatial dimensions. A ReLU activation function is applied after the convolution to introduce non-linearity, enabling the model to capture more complex patterns. The mathematical expression for this step is calculated in Eq. (2).

$$P = \sigma(X * W_1 + b) \quad (2)$$

Where, X is the input, W_1 is the convolution filter of size k , b is the bias term, and σ represents the ReLU activation function. The result of this process, P , contains features with C_{cheap} channels, representing a portion of the total desired channels, C_{out} .

2) *Cheap convolution:* In the next stage, the features produced by the primary convolution undergo a second convolution process using a depthwise convolution. This operation processes each feature channel independently, meaning each channel is convolved with a separate filter, which reduces computational cost while generating additional feature details. The depthwise convolution is applied with a kernel size of $k \times k$, where k is the size of the depthwise kernel. The output of this process can be calculated in Eq. (3).

$$C = \sigma(P * W_2) \quad (3)$$

Where W_2 is the depthwise convolution filter applied to each feature channel in P , and C represents the output of this operation. A ReLU activation function (σ) is also applied to maintain non-linearity in the feature representation. The Ghost module enriches the features with minimal computational overhead compared to full convolutions by using depthwise convolution, contributing to a more efficient model.

3) *Feature Concatenation:* Following the two convolution processes, the outputs from the primary and depthwise convolutions are aggregated. Feature aggregation combines these outputs to create a richer feature representation, merging primary and additional features into a single output set. If the output of the primary convolution is denoted as P and the output of the depthwise convolution as C , the aggregation process can be expressed mathematically in Eq. (4).

$$O = [P, C] \quad (4)$$

Where O represents the aggregation output, and the notation $[.,.]$ signifies concatenation along the channel dimension. This aggregation ensures that the final feature set captures both the main and additional details from the input, leading to a more informative feature representation without significantly increasing computational complexity.

4) *Channel Trimming*: The final stage is channel trimming to ensure that the number of output channels corresponds to the target, C_{out} . After feature aggregation, the number of channels may surpass the target number. Channel trimming is done to keep only the necessary C_{out} channels. If the aggregated feature map O has C_{concat} channels, where $C_{concat} > C_{out}$, only retain the first C_{out} channels that can be represented as in Eq. (5).

$$Y(i, j) = O(i, j, 1: C_{out}) \quad (5)$$

Where, $O(i, j, 1: C_{out})$ denotes selecting the first C_{out} channels from the aggregated feature map O , where i and j are the spatial indices of the feature map. The operation trims the excess channels, ensuring the correct output dimensionality, thus providing an output feature map Y with dimensions $H \times W \times N$, where H and W are the height and width of the spatial dimensions, and N is the number of required channels. This trimming guarantees that the final output size is consistent with the architecture's design without unnecessarily increasing computational costs.

D. Ghost Feature Maps

The proposed Ghost feature maps, illustrated in Fig. 4, integrate Ghost modules with Max Pooling. In CNNs, multiple filters are applied within each convolutional layer to extract various features [29]. These convolutional and pooling layers are arranged sequentially, forming a hierarchical structure that progressively captures and reduces feature dimensions [30]. In this research, traditional convolutional layers are replaced by Ghost modules, which provide a more efficient alternative while maintaining the standard CNN architecture. Utilizing a ratio (r) parameter can generate more features from input with significantly fewer parameters. This research uses a kernel size k of 3 for both the primary [12] and depthwise convolution in the Ghost module, with the bias b set to 0. For instance, in the first layer, Ghost Module (32, $r = 2$) produces 32 features using only half the parameters conventional convolution requires. This not only accelerates training but also reduces the risk of overfitting. Additionally, as the number of features increases in subsequent Ghost Modules (32, 64, 128, 512), the model captures more complex variations in the input data. Increasing the feature channels as the network deepens enriches representation and allows the model to learn more abstract features.

Max pooling then is strategically applied after each Ghost module to reduce the dimensionality of the features significantly extracted. It also greatly reduces computational complexity in later layers, allowing the model to focus on the most dominant features, significantly improving its generalization ability. Max pooling's function in promoting invariance makes the model much more robust to small rotational or translational changes in the input data, which is very important in pattern recognition problems, where such variations are the norm. This makes the audience feel more confident that the model is not only efficient but strong as well.

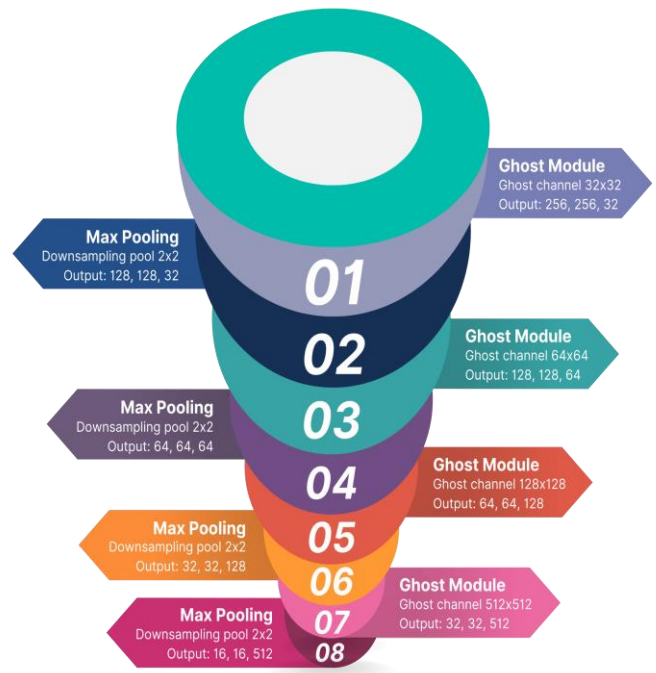


Fig. 4. Ghost feature maps workflow.

Doing that four times, Ghost module and max pooling, gives even more advantages. Every pair builds a pyramid of feature representation. As the layers get deeper, the learned feature hierarchies become more and more complex, starting from simple features in the lower layers to more abstract representations in the higher layers. The architecture actually exploits this by using a Ghost module and then max pooling right after to ensure maximum feature extraction, but without losing any computational efficiency.

E. Parameter Distribution of Proposed Model Architecture

The distribution of parameters in the proposed CNN model's layer structure is presented in Fig. 5. Fig. 5 illustrates a comparative analysis of the total number of parameters between standard 2D convolutional (Conv2D) and Ghost module layers with varying expansion ratios (2 to 5). The Conv2D layer has the highest parameter count at 683,584, indicating its substantial computational complexity. In contrast, the Ghost module layers demonstrate a progressive reduction in the number of parameters as the expansion ratio increases, with 344,736 parameters at a ratio of 2; 150,633 at a ratio of 3; 87,120 at a ratio of 4, and 54,603 at a ratio of 5. This trend highlights the efficiency of Ghost module in significantly reducing parameter count while maintaining performance. The results suggest that higher Ghost module ratios can drastically minimize the model's computational load, offering a more resource-efficient alternative to traditional Conv2D layers. This parameter reduction is advantageous for applications with limited computational resources without compromising the model's representative capacity.

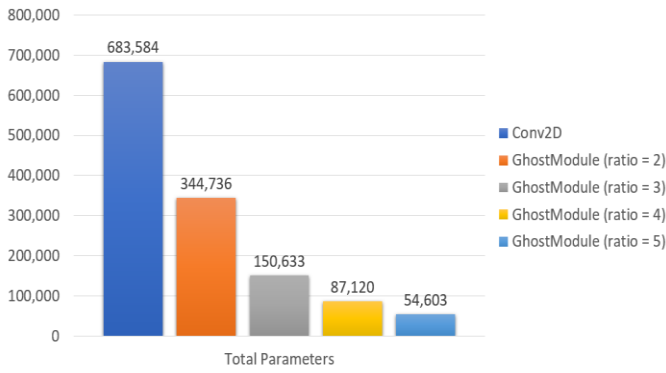


Fig. 5. Comparison of total parameters by layer type.

F. Experimental Setup

The experimental setup is to test the Ghost Feature Maps with ratios of 2, 3, 4, and 5 [12] versus Conv2D layers for image classification. The research will utilize a dataset of Palembang Songket motif images organized into ten distinct classes, with the data split structured as 80% for training, 10% for validation, and 10% for testing purposes. A single learning rate of 0.001 and a batch size of 32 will be employed, with the Adam optimizer selected for model training over 50 epochs.

The model architectures will encompass Ghost Modules for each specified ratio, followed by a Flatten layer, three Dense layers, and a Dropout for regularization. In parallel, a standard 2D convolutional model will be constructed with a similar architectural framework to facilitate direct comparison.

The evaluation will follow some standards: accuracy, precision, recall, and F1-score. The experimental procedure will also preprocess the data so the dataset is normalized and a consistent input is entered into the model. After training, however, all models will be tested thoroughly on the test set, and a complete analysis will be done comparing Ghost Feature Maps to normal convolutional layers and examining how variations in hyperparameters affect the model's overall performance.

G. Classification Performance

Many different performance measures are used to assess the classification models effectiveness. Some more common ones are accuracy, precision, recall, F1-score, and overall accuracy. They all express the model's capability to classify the Palembang Songket patterns differently.

1) *Accuracy*: simply the ratio of correct predictions to the total number of instances [31]. The accuracy value can be calculated using Eq. (6).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

2) *Precision*: a measure used to determine how accurate the model is in its predictions. It is calculated by dividing the number of true positives by the number of predicted positives. Precision is the number of true positives divided by the total number of positives the model predicted [31]. The precision value can be calculated using Eq. (7).

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

3) *Recall*: is the model's ability to find all the true positives. It is the percentage of positive examples that the model correctly labels [31]. The recall value can be calculated using Eq. (8).

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

4) *F1-Score*: is the harmonic mean of precision and recall, used as a single value to balance precision and recall [31]. It can be calculated using Eq. (9).

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9)$$

5) *Overall accuracy*: is simply the number of correct predictions divided by the total number of samples over all of the classes. It is especially useful when performing multiclass classification [31]. The overall accuracy value can be calculated using Eq. (10).

$$Overall Accuracy = \frac{\sum TP + \sum TN}{\sum TP + \sum TN + \sum FP + \sum FN} \quad (10)$$

IV. RESULTS AND DISCUSSION

A. Results

A comparison of two Convolutional Neural Network models using Conv2D layers optimized with Ghost Module produces significant differences. This difference can be seen from the pattern of accuracy or loss that is difficult to stabilize and fluctuations increase towards several epochs when considering training and validation data. The Conv2D model (Fig. 6) in the accuracy and loss graph has four sharp fluctuations in several epochs. The increasing loss and accuracy fluctuations indicate that the model chooses unstable predictions under certain conditions. This pattern indicates the difficulty of the model in identifying patterns in data that are consistent for each epoch in training and validation data. This condition affects the process to remain stable so that the model cannot generalize to new data.

When given the Ghost Module (Fig. 7), some fluctuations in the learning process appear better. The number of sharp fluctuations decreases from four to two, and all remaining conditions are flatter. This effect indicates that the Ghost Module can make the model stable for each epoch so that the trend of decreasing loss becomes more stable and accuracy increases continuously. The remaining training and validation data provide smoother and less extreme patterns like the Conv2D model so that it can detect better patterns each epoch. In addition, the validation data also looks better by decreasing the trend in several epochs. Accuracy increases more stably, and the loss decreases, making the learning model more effective on data without overfitting training and validation data. Overall, Ghost Module can provide a stable model and reduce sharp fluctuations as evidenced by more stable accuracy and decreased loss in training and validation datasets.

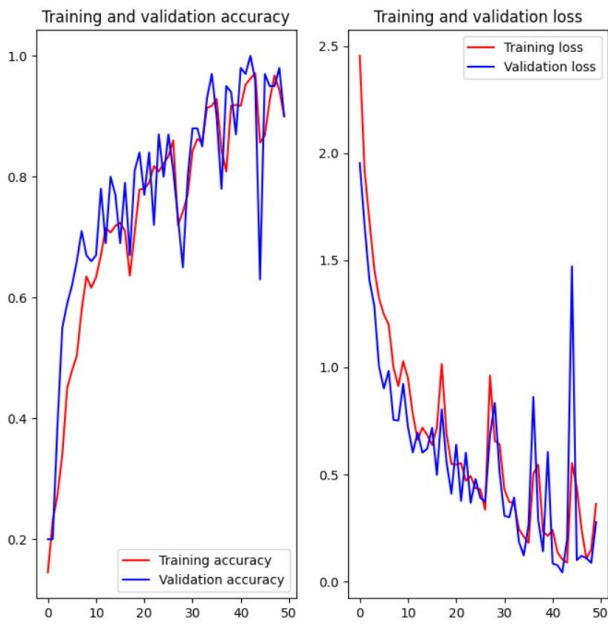


Fig. 6. Training and validation performance, Conv2D.

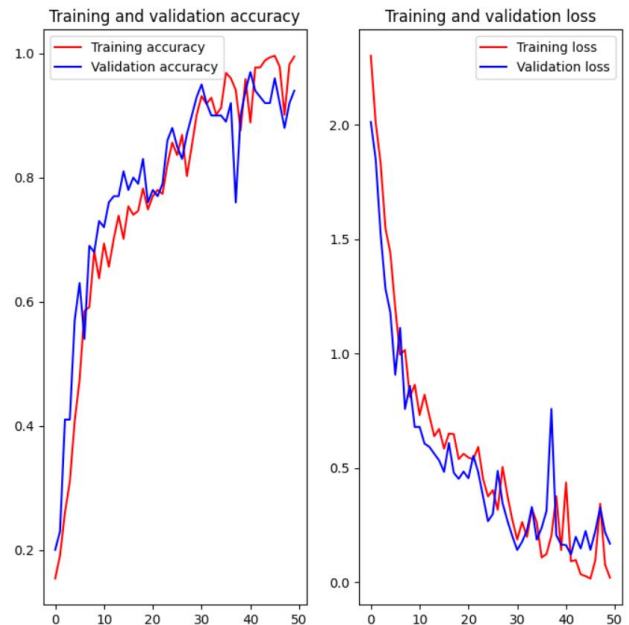


Fig. 8. Training and validation performance, Ghost Feature Maps ($r = 3$).

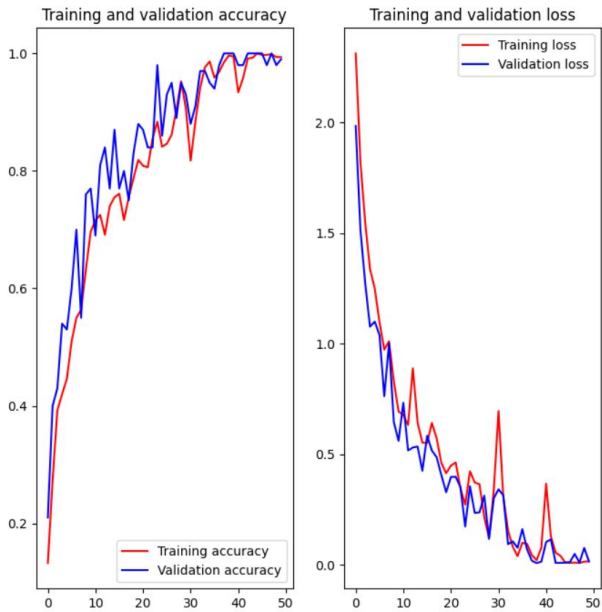


Fig. 7. Training and validation performance, Ghost Feature Maps ($r = 2$).

Furthermore, based on Fig. 8, the accuracy and loss results for the training and validation process of the Ghost Module model with a ratio of 3 are similar to the Ghost Module model with a ratio of 2. However, at ratio 3, the graph shows slightly inconsistent accuracy and loss fluctuation in training and validation. This occurs at epochs 40 to 50, where a gap begins to widen slightly in the accuracy and loss values for the training and validation processes. From the training model graph results, it can be seen that this model can still learn the Palembang Songket motif data well, where there are still similarities in the fluctuations in accuracy and loss values with the Ghost module ratio 2.

For the performance results of the Ghost module with a ratio of 4 in Fig. 9, the fluctuations in loss values for the training and validation processes begin to increase. At least starting from epochs 10 to 20, there have been quite large fluctuations, but for the following epochs up to 50, the loss value begins to decrease. Compared to ratio 3, the Ghost module for ratio 4 tends to be less stable at epochs 30 to 50. In this epoch range, a gap begins to move away from the accuracy and loss values. The results of ratio four show that the training model began to experience a decrease in performance for the classification of Palembang Songket motifs compared to ratio 3.

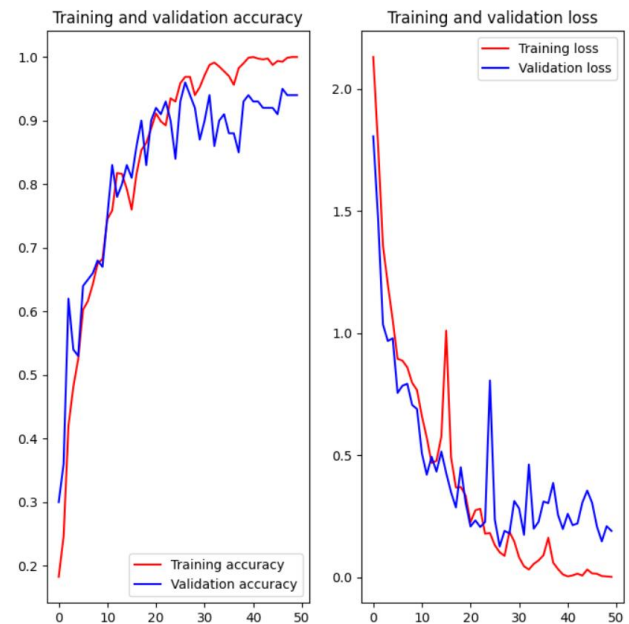


Fig. 9. Training and validation performance, Ghost Feature Maps ($r = 4$).

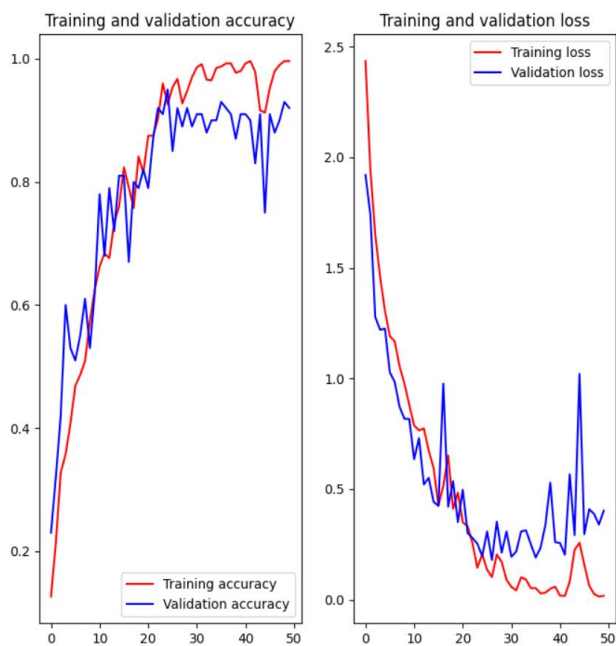


Fig. 10. Training and validation performance, Ghost Feature Maps ($r = 5$).

At ratio five shown in Fig. 10, the performance of the Ghost Module shows a wider gap for accuracy and loss values. This is almost the same as ratio four, where the condition of the training model experienced quite large fluctuations from epoch 30 to 50. However, more significant fluctuations occurred in epoch 40 to 50 compared to ratios 2, 3, and 4. This is due to a significant reduction in the number of parameters in the Ghost module, so it cannot capture the pattern of the Palembang Songket motif features.

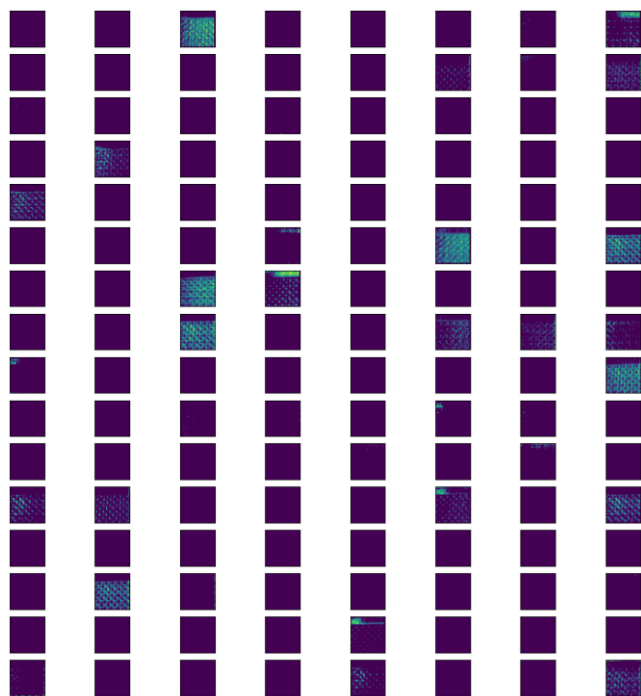


Fig. 11. Visualization of Conv2D feature maps.

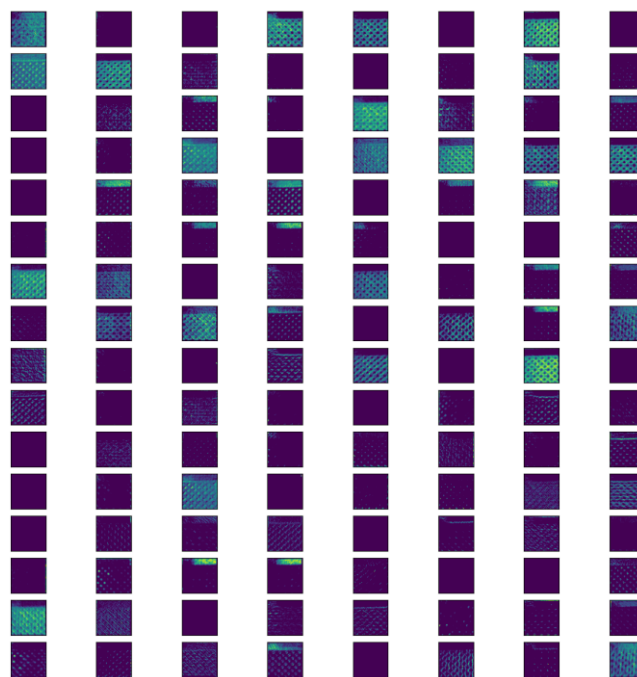


Fig. 12. Visualization of Ghost Feature Maps.

Another way to test the model is to visualize the feature maps and see what kind of features the model has learned from the Palembang Songket motif image. The visualization results are then compared with the CNN model with standard Conv2D and Ghost Feature Maps. As shown in Fig. 11, the visualization results of the feature maps that the CNN model with Conv2D produced have a lot of dark spots and only a few obvious motif feature patterns using this trained model. It turns out that even with the trained model it is still hard to get the Palembang Songket motif pattern.

As seen in Fig. 12, unlike the regular CNN model with Conv2D, the CNN model with Ghost Feature Maps can actually yield clearer feature maps than the regular CNN model. As evidence by the feature map of the Ghost feature map, it performs better than the normal Conv2D. The CNN model with ghost really brings out the Songket motif pattern in many places. The shape and edge features are more distinct than in the regular CNN model. This is good because it will allow the model to learn to recognize the intricate Palembang Songket pattern much more accurately than the old Conv2D model.

Table I shows the results of the comparison of the classification performance of the CNN model with Ghost Feature for 10 classes of Palembang Songket motifs. Compared to CNN with Conv2D, CNN with Ghost feature (ratio 2) is able to provide better classification performance. This is proven by the increase in accuracy, precision, recall, and f1-score for all classes of Palembang Songket motifs. For the Songket Bintang Berantai motif, there was an increase in accuracy from 0.94 to 0.98, precision from 0.63 to 0.83, recall remained at 1.00, and f1-score from 0.77 to 0.91. Furthermore, the Bunga Jatuh motif increased with accuracy from 0.99 to 1.00, precision remained at 1.00, recall from 0.90 to 1.00, and f1-score from 0.95 to 1.00. Then, the Kenanga Makan Ulat motif increased with accuracy from 0.98 to 1.00; precision remained at 1.00, recall from 0.80

to 1.00, and f1-score from 0.89 to 1.00. After that, for the Naga Besaung motif, accuracy increased from 0.95 to 0.98, precision from 0.86 to 1.00, recall from 0.60 to 0.80, and f1-score from 0.71 to 0.89. Then, for the Nampan Perak motif, there was an increase in accuracy from 0.98 to 1.00, precision from 0.90 to 1.00, recall from 0.90 to 1.00, and f1-score from 0.90 to 1.00. The rest, the Bunga Cina, Cantik Manis, Jando Beraes, Pacar Cina, and Pulir motifs, did not increase because they were already very well recognized.

Table II shows the results of the overall comparison of the performance of the CNN model with Ghost Feature. The

comparison consists of total parameters, overall accuracy, and model size. The total parameters referred to are the overall parameters used from the input layer to the output layer, namely ghost feature maps, flatten, up to the fully connected layer. Based on the results in Table II, it was found that the CNN model with Ghost Feature provided an overall accuracy of 0.98 with fewer total parameters compared to the CNN model with Conv2D which had an overall accuracy of 0.92. In addition, the CNN model with Ghost Feature ratios of 4 and 5 was able to provide the same overall accuracy of 0.93 with much more reduced parameters with a smaller model size.

TABLE I. COMPARISON OF MODEL PERFORMANCE BASED ON PALEMBANG SONGKET MOTIF CLASS

Songket Motif	Accuracy		Precision		Recall		F1 – Score	
	Conv2D	Ghost Feature	Conv2D	Ghost Feature	Conv2D	Ghost Feature	Conv2D	Ghost Feature
Bintang Berantai	0.94	0.98	0.63	0.83	1.00	1.00	0.77	0.91
Bunga Cina	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Bunga Jatuh	0.99	1.00	1.00	1.00	0.90	1.00	0.95	1.00
Cantik Manis	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Jando Beraes	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Kenanga Makan Ulat	0.98	1.00	1.00	1.00	0.80	1.00	0.89	1.00
Naga Besaung	0.95	0.98	0.86	1.00	0.60	0.80	0.71	0.89
Nampan Perak	0.98	1.00	0.90	1.00	0.90	1.00	0.90	1.00
Pacar Cina	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Pulir	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

TABLE II. COMPARATIVE RESULTS OF THE PROPOSED METHOD

Model	Total Parameters	Overall Accuracy	Model Size (MB)
CNN with Conv2D	68,060,746	0.92	259.63
CNN with Ghost Feature ($r = 2$)	67,721,898	0.98	258.34
CNN with Ghost Feature ($r = 3$)	44,983,411	0.95	171.60
CNN with Ghost Feature ($r = 4$)	33,909,850	0.93	129.36
CNN with Ghost Feature ($r = 5$)	27,061,049	0.93	103.23

B. Discussion

A comparison was also conducted with two previous studies on Songket motif classification using CNN-based methods, as summarized in Table III. Ariessaputra et al. [7] utilized a CNN architecture comprising Conv2D, Max Pooling, and Fully Connected layers, while Hambali et al. [8] implemented a CNN model with additional Dropout layers to enhance generalization. Both approaches leveraged CNN's feature extraction and classification capability for traditional fabric patterns. These studies highlight the relevance of CNN architectures for motif recognition, providing a contextual foundation for evaluating the proposed method's design and performance.

TABLE III. METHOD COMPARISON

Authors	Methods	Dataset	Overall Accuracy
Ariessaputra et al. [7]	CNN (Conv2D, MaxPooling, Fully Connected)	Lombok Songket Motifs	0.84
Hambali et al. [8]	CNN (Conv2D, MaxPooling, Dropout, Fully Connected)	Lombok Songket Motifs	0.86
Ours	CNN with Ghost Feature ($r = 2$)	Songket Palembang Motifs	0.98

Each method has its strengths and limitations. However, the proposed approach, which retains the basic CNN architecture with Dropout but replaces Conv2D and Max Pooling with Ghost Feature maps involving the Ghost Module and Max Pooling, shows improved performance over the previous studies. The hierarchical combination of Ghost Module and Max Pooling in the proposed method leads to better classification results compared to the state-of-the-art methods from Ariessaputra et al. [7] and Hambali et al. [8]. This improvement highlights the effectiveness of the proposed method in enhancing motif recognition accuracy.

In addition to key parameters such as the learning rate and batch size, the experiment also involved setting a Dropout rate

of 0.5 in the dense layer. This configuration aimed to improve the model's generalization in recognizing complex and diverse Songket motifs. During training, Dropout randomly deactivates units, helping the model learn feature representations that are more adaptive to the distinctive patterns of Songket motifs, such as geometric curves and overlapping color variations.

The proposed model addresses the limitations of previous approaches by integrating the Ghost Module and a hierarchical combination of Max Pooling, which significantly enhances the efficiency of dominant feature extraction without losing the primary characteristics of motif patterns. Unlike conventional convolutional layers that rely on a large number of parameters, this model leverages the Ghost Module for a more lightweight feature generation mechanism, while the hierarchical integration of Max Pooling reduces redundancy and ensures a more focused feature extraction process. This approach not only improves efficiency in filtering less relevant features but also strengthens the model's ability to capture intricate patterns, especially in motifs with significant visual similarities and minor variations. With this structure, the model demonstrates enhanced efficiency and effectiveness in recognizing Songket motifs.

V. CONCLUSION

The use of Ghost feature maps, which involve the Ghost module, in the CNN model leads to a significant reduction in the number of parameters and the model size compared to traditional CNNs utilizing Conv2D. This efficiency is highlighted by the model achieving an impressive accuracy of 0.98 at a ratio of 2, with only a minor parameter reduction of about 0.5% and a slightly smaller model size. However, as the ratio of Ghost feature maps increases, a further decrease in parameters and model size occurs, accompanied by a decline in accuracy. Specifically, a ratio of 3 results in a 34% reduction in parameters but lowers accuracy to 0.95. Ratios 4 and 5 stabilize accuracy at 0.93 while achieving over 60% reductions in model size and parameters compared to the Conv2D model. Thus, a trade-off between accuracy and model size becomes evident, particularly at a ratio of 3, where significant size reductions are achieved with only a slight impact on accuracy.

The proposed Ghost Feature maps in this model are constructed hierarchically through a combination of Ghost Modules and Max Pooling, applied four times. Each pair forms a pyramid-like feature representation, allowing the model to learn increasingly complex feature hierarchies as the depth of the layers increases. However, the optimal number of feature repetition levels required for achieving the best performance remains to be explored. Future research should investigate whether adding deeper hierarchical layers could reduce performance or significantly improve recognition accuracy. Therefore, further development of deeper architectures and evaluation at various layer depths is necessary to determine whether this approach can significantly improve Songket motif recognition.

ACKNOWLEDGMENT

This research was supported by Ministry of Education, Culture, Research, and Technology of Indonesia through the

research grant for the fundamental research scheme in 2024 with contract number 104/E5/PG.02.00.PL/2024.

REFERENCES

- [1] A. Suzianti, R. D. Amaradhanny, and S. N. Fathia, "Fashion heritage future: Factors influencing Indonesian millenials and generation Z's interest in using traditional fabrics," *J. Open Innov. Technol. Mark. Complex.*, vol. 9, no. 4, p. 100141, Dec. 2023, doi: 10.1016/j.joitmc.2023.100141.
- [2] K. Sedyastuti, E. Suwarni, D. R. Rahadi, and M. A. Handayani, "Human Resources Competency at Micro, Small and Medium Enterprises in Palembang Songket Industry," in *Proceedings of the 2nd Annual Conference on Social Science and Humanities (ANCOSH 2020)*, 2021, pp. 248–251. doi: 10.2991/assehr.k.210413.057.
- [3] "Songket Palembang," *Warisan Budaya Takbenda Indonesia*, 2013. <https://budaya-data.kemdikbud.go.id/wbtb/objek/AA000222> (accessed Oct. 10, 2024).
- [4] Y. Yullyana, D. Irmayani, and M. N. S. Hasibuan, "Content-Based Image Retrieval for Songket Motifs using Graph Matching," *Sinkron*, vol. 7, no. 2, pp. 714–719, May 2022, doi: 10.33395/sinkron.v7i2.11411.
- [5] S. Sriani, M. S. Hasibuan, and R. Ananda, "Classification of Batu Bara Songket Using Gray-Level Co-Occurrence Matrix and Support Vector Machine," *J. Ris. Inform.*, vol. 5, no. 1, pp. 481–490, Dec. 2022, doi: 10.34288/jri.v5i1.469.
- [6] R. Aprianti, K. Evandari, R. A. Pramunendar, and M. Soeleman, "Comparison Of Classification Method On Lombok Songket Woven Fabric Based On Histogram Feature," in *2021 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Sep. 2021, pp. 196–200. doi: 10.1109/iSemantic52711.2021.9573223.
- [7] S. Ariessaputra, V. H. Vidiyari, S. M. Al Sasongko, B. Darmawan, and S. Nababan, "Classification of Lombok Songket and Sasambo Batik Motifs Using the Convolution Neural Network (CNN) Algorithm," *JOIV Int. J. Informatics Vis.*, vol. 8, no. 1, pp. 38–44, Mar. 2024, doi: 10.62527/joiv.8.1.1386.
- [8] H. Hambali, M. Mahayadi, and B. Imran, "Classification of Lombok Songket Cloth Image Using Convolution Neural Network Method (CNN)," *J. Pilar Nusa Mandiri*, vol. 17, no. 2, pp. 149–156, 2021, doi: <https://doi.org/10.33480/pilar.v17i2.2705>.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016. doi: 10.1109/CVPR.2016.90.
- [10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *Advances in Neural Information Processing Systems*, 2012, vol. 25. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf
- [11] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *3rd Int. Conf. Learn. Represent. ICLR* 2015, pp. 1–14, 2015.
- [12] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More Features From Cheap Operations," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2020, pp. 1577–1586. doi: 10.1109/CVPR42600.2020.00165.
- [13] R. Andrian, R. Taufik, D. Kurniawan, A. S. Nahri, and H. C. Herwanto, "Lampung Batik Classification Using AlexNet, EfficientNet, LeNet and MobileNet Architecture," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 11, 2024, doi: 10.14569/IJACSA.2024.0151191.
- [14] L. Elvitaria, E. F. A. Shaubari, N. A. Samsudin, S. K. A. Khalid, S. -, and Z. Indra, "A Proposed Batik Automatic Classification System Based on Ensemble Deep Learning and GLCM Feature Extraction Method," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 10, 2024, doi: 10.14569/IJACSA.2024.0151058.
- [15] R. Muliono, M. S. Iranita, and R. B. Syah, "An Effectivity Deep Learning Optimization Model to Traditional Batak Culture Ulos Classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 4, pp. 634–638, 2023, doi: 10.14569/IJACSA.2023.0140469.

- [16] Z. Wang and T. Li, "A Lightweight CNN Model Based on GhostNet," *Comput. Intell. Neurosci.*, vol. 2022, pp. 1–12, Jul. 2022, doi: 10.1155/2022/8396550.
- [17] Q. Zhao and H. Cheng, "An Efficient Approach to Human Security Screening Image Recognition Through a Lightweight CNN Utilizing Yolov5s and GhostNet," *Trait. du Signal*, vol. 40, no. 4, pp. 1653–1660, Aug. 2023, doi: 10.18280/ts.400433.
- [18] W. Chen, Y. Zhang, X. Chen, and W. Li, "Research On Vehicle Detection based on Ghost Net and Se Attention Mechanism," in *2023 11th International Conference on Information Technology: IoT and Smart City (ITIoTSC)*, Aug. 2023, pp. 268–271. doi: 10.1109/ITIoTSC60379.2023.00055.
- [19] Z. Huangfu, S. Li, and L. Yan, "Ghost-YOLO v8: An Attention-Guided Enhanced Small Target Detection Algorithm for Floating Litter on Water Surfaces," *Comput. Mater. Contin.*, vol. 80, no. 3, pp. 3713–3731, 2024, doi: 10.32604/cmc.2024.054188.
- [20] B. Fang, G. Chen, and J. He, "Ghost-based Convolutional Neural Network for Effective Facial Expression Recognition," in *2022 International Conference on Machine Learning and Knowledge Engineering (MLKE)*, Feb. 2022, pp. 121–124. doi: 10.1109/MLKE55170.2022.00029.
- [21] M. Alansari, O. A. Hay, S. Javed, A. Shoufan, Y. Zweiri, and N. Werghi, "GhostFaceNets: Lightweight Face Recognition Model From Cheap Operations," *IEEE Access*, vol. 11, pp. 35429–35446, 2023, doi: 10.1109/ACCESS.2023.3266068.
- [22] F. Luan, X. Mu, and S. Yuan, "Ghost Module Based Residual Mixture of Self-Attention and Convolution for Online Signature Verification," *Comput. Mater. Contin.*, vol. 79, no. 1, pp. 695–712, 2024, doi: 10.32604/cmc.2024.048502.
- [23] M. E. Paoletti, J. M. Haut, N. S. Pereira, J. Plaza, and A. Plaza, "Ghostnet for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10378–10393, Dec. 2021, doi: 10.1109/TGRS.2021.3050257.
- [24] Y. Tang, K. Han, J. Guo, C. Xu, C. Xu, and Y. Wang, "GhostNetV2: Enhance Cheap Operation with Long-Range Attention," in *Advances in Neural Information Processing Systems*, Nov. 2022, pp. 9969–9982. [Online]. Available: <http://arxiv.org/abs/2211.12905>
- [25] Z. Liu, Z. Hao, K. Han, Y. Tang, and Y. Wang, "GhostNetV3: Exploring the Training Strategies for Compact Models," 2024, [Online]. Available: <http://arxiv.org/abs/2404.11202>
- [26] Z. He, D. He, X. Li, and R. Qu, "Blind Superresolution of Satellite Videos by Ghost Module-Based Convolutional Networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–19, 2023, doi: 10.1109/TGRS.2022.3233099.
- [27] S. Liu, B. Ding, J. Bai, and Z. Xiao, "Hyperspectral Image Classification Based on Extended Morphological Profile Features and Ghost Module," in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Jul. 2021, pp. 3617–3620. doi: 10.1109/IGARSS47720.2021.9554092.
- [28] Y. Huang, Y. Zhou, J. Lan, Y. Deng, Q. Gao, and T. Tong, "Ghost Feature Network for Super-Resolution," in *2020 Cross Strait Radio Science & Wireless Technology Conference (CSRSWTC)*, Dec. 2020, pp. 1–3. doi: 10.1109/CSRSWTC50769.2020.9372549.
- [29] A. Ghosh, A. Sufian, F. Sultana, A. Chakrabarti, and D. De, "Fundamental Concepts of Convolutional Neural Network," in *Recent Trends and Advances in Artificial Intelligence and Internet of Things*, V. E. Balas, R. Kumar, and R. Srivastava, Eds. Cham: Springer International Publishing, 2020, pp. 519–567. doi: 10.1007/978-3-030-32644-9_36.
- [30] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, p. 53, 2021, doi: 10.1186/s40537-021-00444-8.
- [31] P. Włodarczak, *Machine Learning and its Applications*. University of Southern Queensland, Toowoomba, Queensland, Australia: CRC Press, 2020.

A Novel Hybrid Algorithm Based on Butterfly and Flower Pollination Algorithms for Scheduling Independent Tasks on Cloud Computing

Huiying SHAO

Hebei Vocational University of Technology and Engineering, Xingtai 054000, China

Abstract—Cloud computing is an Internet-based computing paradigm where virtual servers or workstations are offered as platforms, software, infrastructure, and resources. Task scheduling is considered one of the major NP-hard problems in cloud environments, posing several challenges to efficient resource allocation. Many metaheuristic algorithms have been extensively employed to address these task-scheduling problems as discrete optimization problems and have given rise to some proposals. However, these algorithms have inherent limitations due to local optima and convergence to poor results. This paper suggests a hybrid strategy for organizing independent tasks in heterogeneous cloud resources by incorporating the Butterfly Optimization Algorithm (BOA) and Flower Pollination Algorithm (FPA). Although BOA suffers from local optima and loss of diversity, which may cause an early convergence of the swarm, our hybrid approach outperforms such weaknesses by exploiting a mutualism-based mechanism. Indeed, the proposed hybrid algorithm outperforms existing methods while considering different task quantities with better scalability. Experiments are conducted within the CloudSim simulation framework with many task instances. Statistical analysis is performed to test the significance of the obtained results, which confirms that the suggested algorithm is effective at solving cloud-based task scheduling issues. The study findings indicate that the hybrid metaheuristic algorithm could be a promising approach to improving resource utilization and optimizing cloud task scheduling.

Keywords—Task scheduling; cloud computing; butterfly optimization algorithm; flower pollination algorithm; mutualism

I. INTRODUCTION

Cloud computing is an Internet-based approach that enables elastic, easy-to-scale access to a broad set of resources, including storage, computing, and networking applications delivered via the Internet [1]. In contrast to traditional systems, dependent on locally stored resources, cloud computing allows flexible and scalable access to resources from any location. There are three main service configurations: Platform as a Service (PaaS), Infrastructure as a Service (IaaS), and Software as a Service (SaaS) [2]. IaaS involves virtualized assets accessible through the internet, offering elementary amenities like servers, storage facilities, and networking, empowering companies to expand slanted utopias without trusting physical devices [3].

While IaaS provides virtualized computing resources, PaaS is an extension that offers developers a platform complete with

tools and frameworks [4]. This allows the developer to create, evaluate, and launch applications without explicitly managing the underlying infrastructure. SaaS directly delivers ready-to-consumer applications to end-users, including email, CRM, and collaborative software, accessed via web browsers [5]. Together, these models catalyze innovation and cost-efficiency in sectors by allowing companies to lessen their IT overhead, hasten product deployment, and respond dynamically to market demands.

Scheduling tasks in cloud computing belongs to fundamental NP-hard problems that need to be solved to ensure better efficiency in resource allocation within virtual environments [6]. This problem falls under the combinatorial optimization class wherein multiple heterogeneous tasks must be assigned to available resources for maximum efficiency [7]. As finding an optimal resource allocation problem in scheduling tasks with different requirements is often combinatorial, more advanced strategies provide an alternative to conventional approaches. Common objectives in task scheduling include reducing execution time (or makespan) to ensure tasks are completed as quickly as possible, which enhances user satisfaction and system performance [8].

The other objective is to perform load balancing, in which tasks should be allocated to resources to avoid bottleneck situations and overutilization of particular servers, providing better system resiliency [9]. Last but not least, efficient usage of resources will prevent the idleness of resources and minimize operational costs by utilizing the maximum availability of infrastructure [10]. Therefore, efficient scheduling strategies are crucial for cloud environments, where dynamic scaling of resources relies on accurate and adaptive scheduling to accommodate the diversified requirements of end-users and applications.

Simultaneously, advancements in mobile robotics, particularly in navigation and mapping, provide valuable insights into addressing dynamic resource allocation challenges in cloud environments. Techniques such as reinforcement learning have demonstrated the potential to enhance decision-making and adaptability in complex scenarios [11]. These insights could inspire novel approaches to optimizing task scheduling in cloud computing, where dynamic and unpredictable demands necessitate intelligent and resilient solutions.

Metaheuristic algorithms, including the Flower Pollination Algorithm (FPA) and Butterfly Optimization Algorithm (BOA), are employed in cloud task scheduling because of their flexibility in handling complicated optimization challenges [12]. BOA is inspired by butterflies' sensory communication through fragrance. The fragrance guides each solution or member chemically to an optimal solution, mirroring the prey's natural process. This mechanism will help explore potential solutions within the search space and zoom into promising, high-quality areas [13]. On the other hand, FPA draws inspiration from flowers' pollination behavior, combining local and global pollination processes to examine the solution domain effectively. The global pollination phase facilitates the exploration of diverse solutions, while local pollination fine-tunes promising areas [14].

While BOA and FPA have emerged as promising optimization techniques, they face significant limitations when applied to task scheduling. BOA often experiences early convergence and can become trapped in local optima due to its limited ability to fully utilize the optimal solution. Additionally, BOA's phase-switching mechanism may become disoriented, deviating from the best global solutions. Similarly, FPA, despite its strength in balancing exploration and exploitation, can suffer from reduced diversity over time, limiting its capacity to explore novel solutions. To overcome these challenges, this study introduces a novel hybrid algorithm that integrates BOA and FPA through a mutualism mechanism inspired by ecological interactions.

In this context, the strengths of one algorithm offset the weaknesses of the other, creating a synergistic optimization

process. Furthermore, we propose an adaptive switching probability mechanism, a key innovation of this study, which dynamically adjusts the balance between the exploitation and exploration phases. This unique combination enhances the search process, improves convergence, and significantly optimizes cloud-based task scheduling, marking a substantial contribution to cloud computing optimization.

The remainder of the paper is organized as follows: The state-of-the-art review is outlined in Section II, about different existing cloud task-scheduling approaches as well as different meta-heuristic algorithms. This is followed by describing, in Section III, the problem statement, which includes the challenges and objectives that characterize cloud task scheduling. Section IV outlines our hybrid novel algorithm, illustrating its various components, including the mechanism behind the mutualism and switching probability adaptation process. Section V describes the experimental setup and discusses the results of the simulation. The implications of the findings are discussed in detail in Section VI. Finally, the paper concludes by summarizing the contributions in Section VII and presenting possible further research.

II. RELATED WORK

This section summarizes recent advancements in cloud task scheduling algorithms, as summarized in Table I. Various hybrid and metaheuristic approaches are highlighted, focusing on optimizing makespan, resource utilization, and load balancing to handle scheduling challenges in cloud computing environments.

TABLE I. SUMMARY OF RELATED WORKS ON CLOUD TASK SCHEDULING ALGORITHMS

Research	Description	Performance metrics	Key Findings
[15]	Genetic algorithm and multi-verse optimization are integrated to optimize task scheduling, focusing on bandwidth, virtualization, task counts, and sizes in cloud environments.	Time minimization and task transfer efficiency	Shows promising results in minimizing time for massive tasks by optimizing resource allocation.
[16]	Combination of genetic algorithm and thermodynamic simulated annealing, with crossover operator and thermodynamic mechanisms for balanced exploration and exploitation.	Effectiveness, speedup, schedule duration, and makespan	Effective in balancing exploration and exploitation and reducing makespan compared to other approaches.
[17]	Multiple objective task scheduling using grey wolf optimization, prioritizing tasks based on resource status and demand using HPC2N and NASA workload archives.	Makespan and resource allocation efficiency	Achieves significant improvements in scheduling parameters and adapts well to workload variability.
[18]	A novel method merging particle swarm optimization and genetic algorithms using phagocytosis-inspired merging for population diversity with a feedback mechanism.	Task completion time and convergence accuracy	Enhances task completion time and accuracy by guiding population movement toward optimal solutions.
[19]	Hybrid grey wolf optimization and genetic algorithm	Makespan, cost, and energy consumption	Outperforms GWO, GA, and PSO in minimizing makespan, energy use, and cost for large scheduling tasks.
[20]	The chameleon and remora search optimization algorithm integrates CSA and RSOA with a greedy approach focusing on MIPS and network bandwidth.	Makespan, load balancing, and cost	Effectively minimizes completion time and balances VM load, outperforming baseline approaches.
[21]	Uses dense spatial clustering to schedule tasks, aiming to optimize execution time and enhance the quality of service for user tasks.	Execution time, average start time, and completion time	Achieved a 13% reduction in execution time and a 49% improvement in start and completion times over ACO and PSO algorithms.

Abualigah and Alkhrabsheh [15] presented MVO-GA, a hybrid multi-verse optimizer and genetic algorithm to optimize task scheduling. In such a way, this approach enhances task transfer efficiency in a cloud system by investigating various aspects of cloud assets, including bandwidth, virtualization, task counts, and task sizes. The technique has shown promising results in minimizing the time used for massive cloud tasks.

Tanha, et al. [16] developed a combined meta-heuristic algorithm using the thermodynamic simulated annealing and genetic algorithms to resolve the cloud task scheduling issue. The performance of the algorithm is improved by a crossover operator and thermodynamic simulated annealing. In this approach, there is a reasonable equilibrium between exploration and exploitation.

Mangalampalli, et al. [17] suggested the multi-objective task scheduling grey wolf optimization algorithm, MOTSGWO, in which tasks are prioritized based on cloud resource status and workload demand. This approach is implemented in the Cloudsim toolkit with workloads generated from the HPC2N and NASA parallel workload archives. The experiments show the outstanding performance of MOTSGWO.

Fu, et al. [18] created a novel methodology using phagocytosis combined with particle swarm optimization and genetic algorithms. The method divides particles, adjusts their positions, and merges subpopulations for diversity. It uses a feedback mechanism to ensure the population moves towards the optimal solution. Simulations show it enhances cloud task completion time and convergence accuracy.

Behera and Sobhanayak [19] developed an algorithm that combines the Grey Wolf Optimization Algorithm (GWO) and Genetic Algorithm (GA) to improve cloud computing task scheduling. It aims to minimize cost, makespan, and energy usage while leveraging the GA-driven GWO algorithm's accelerated convergence in significant scheduling challenges.

Pabitha, et al. [20] developed a Chameleon and Remora Search Optimization Algorithm (CRSOA) to optimize cloud task scheduling by considering MIPS and network bandwidth. The CRSOA model, a multi-objective model, integrates the strengths of the Chameleon Search Algorithm (CSA) and Remora Search Optimization Algorithm (RSOA) through a greedy strategy. Simulation results showed that the CRSOA approach minimizes completion time and effectively handles load balancing between available VMs. The experimental investigation confirmed its effectiveness in reducing makespan, cost, and imbalance levels over baseline approaches.

Mustapha and Gupta [21] designed an approach based on DBSCAN (Density-Based Spatial Clustering) for task scheduling to ensure optimal effectiveness. DBSCAN-based task scheduling methodology enhances user satisfaction and optimises execution times, average start times, and end times. The experimental results reveal that the suggested model surpasses the current PSO and ACO, demonstrating 15% better execution times and 48% better start and completion times.

III. PROBLEM STATEMENT

The cloud task scheduling problem revolves around efficiently assigning multiple tasks to Virtual Machines (VMs) within a Cloud System (CS) to achieve the shortest possible execution time [22]. An overview of the proposed task scheduling system for a CS is shown in Fig. 1. The task manager component in this system collects tasks from different users. Upon receiving user tasks, the task manager arranges and forwards them to the scheduler component. The task manager also knows the basis for VMs' information. As a result, it supplies the task scheduler component with information about the status of VMs and task requests. In this context, the CS is represented by multiple Physical Machines (PMs), each housing several VMs [23], expressed as follows:

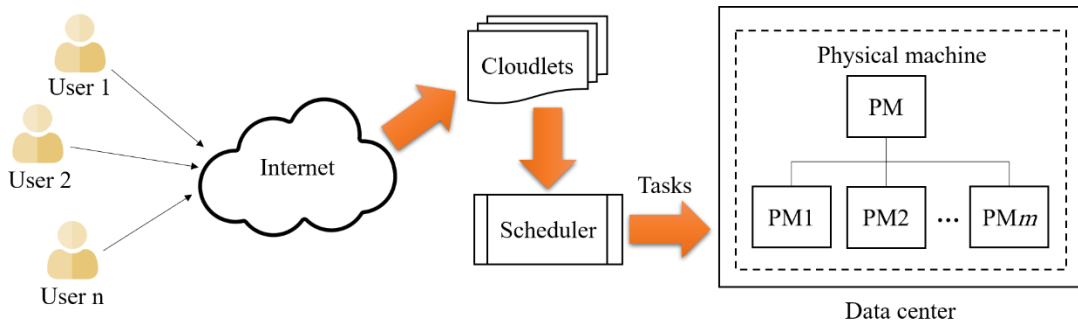


Fig. 1. System model for cloud task scheduling.

$$CS = \{PM_1, PM_2, \dots, PM_i, \dots, PM_m\} \quad (1)$$

Where m reflects the number of PMs in the system, and each PM i comprises a set of VMs as follows:

$$PM = \{VM_1, VM_2, \dots, VM_k, \dots, VM_n\} \quad (2)$$

Where n denotes the number of VMs within a particular PM. Each VM_k is defined by its processing speed $MIPS_k$ (measured in millions of instructions per second) and unique

identifier SID_{VMk} [24]. The tasks to be scheduled in the cloud are detailed as follows:

$$T = \{T_1, T_2, \dots, T_l, \dots, T_z\} \quad (3)$$

Where z stands for the total number of tasks, and each task T_l is described by an identifier SID_{Tl} , its length in terms of instructions (task_length), the expected completion time ECT_l , and priority PI . The expected completion time for a task T_l on VM_k is calculated using Eq. (4) [25].

$$ECT_{lk} = \frac{T \cdot length_l}{MIPS_k} \quad (4)$$

This scheduling problem is, therefore, formulated as an optimization problem. The objective is to distribute tasks across VMs to minimize total execution time, thereby maximizing resource utilization. The objective function to shorten the overall makespan can be expressed as:

$$fit = \min \left(\max_{k=1,2,\dots,n} \sum_{l=1}^z ECT_{lk} \right) \quad (5)$$

This approach aims to balance the load across virtual machines and optimize resource usage within the cloud infrastructure.

IV. METHODOLOGY

A. Butterfly Optimization Algorithm

BOA is a metaheuristic algorithm inspired by butterflies' cooperative foraging behavior. In the BOA, butterflies can find optimal solutions based on a fragrance function, influenced by parameters such as power exponent (a) and sensory modality (c) [26]. This fragrance, which represents the butterfly's fitness, is defined by Eq. (6).

$$f(t) = c \cdot I(t)^a \quad (6)$$

Where $I(t)$ denotes the stimulus intensity at a given step t , controlled by the sensory modality and power exponent. The fragrance emitted by each butterfly attracts others and guides their movement through the solution space.

BOA operates in three primary phases: initialization, iteration, and finalization [27]. During the initialization process, the objective function and the solution area are defined, generating a population of butterflies. Each butterfly's position is set, and fitness and fragrance scores are computed. In the iteration stage, BOA alternates between global and local key search strategies. Based on Eq. (7), each butterfly is guided toward the fittest solution $*$ in the global search.

$$x_i^{t+1} = x_i^t + (r^2 \cdot (g^* - x_i^t)) \cdot f_i \quad (7)$$

Where x_i^t represents the position of the i^{th} butterfly at iteration t , g^* denotes the best current solution, f_i is the fragrance of the i^{th} butterfly, and r is a random number between 0 and 1. In the local search mode, butterflies move based on the influence of nearby individuals. The position update is given by:

$$x_i^{t+1} = x_i^t + (r^2 \cdot (x_j^t - x_k^t)) \cdot f_i \quad (8)$$

Where x_j^t and x_k^t are positions of two randomly selected butterflies in the population. This local interaction allows BOA to explore diverse regions within the solution space.

A switch probability (p) controls the balance between global and local searches, enabling the algorithm to dynamically shift from broad exploration to intense local refinement. This adaptive strategy helps BOA avoid premature convergence and enhances its ability to find optimal solutions effectively, making it suitable for complex optimization tasks such as cloud scheduling.

Generally, BOA involves five steps. The first step is initializing all BOA and problem parameters. BOA has five parameters: population size (N), number of iterations (Itr), c , a , and p . In the second step, the BOA generates all solutions randomly. The solutions take the form of length vectors with the same dimension as the problem dimension d . A matrix containing all the solutions creates the population as follows.

The BOA typically consists of five sequential steps. The initial step involves setting up all BOA-related parameters and problem-specific variables. BOA utilizes five key parameters: population size (N), number of iterations (Itr), and the constants c , a , and p . During the second step, BOA generates all solutions randomly. These solutions are represented as vectors of equal length, corresponding to the dimensionality of the problem (d). The collection of these solution vectors forms a population matrix, structured as follows.

$$Population = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_d^1 \\ x_1^2 & x_2^2 & \dots & x_d^2 \\ \vdots & \vdots & \dots & \vdots \\ x_1^N & x_2^N & \dots & x_d^N \end{bmatrix} \quad (9)$$

The optimization problem's objective function serves to assess all potential solutions during the third step. Subsequently, the best-performing solution is designated as g^* . In the fourth step, all solutions are revised according to the fitness values determined in the previous phase. To guide the search process locally or globally, a random number r is generated and compared to the threshold p . If r is smaller than p , the butterfly executes a global movement following Equation 7; otherwise, it performs a local movement based on Equation 8. If the newly generated solution outperforms the previous one, it replaces the earlier solution. The value of g^* is then updated accordingly. Finally, the termination condition is evaluated. The pseudo-code outlining the general steps of the BOA is presented in Fig. 2.

Step 1:

Set up the problem-specific parameters.
Initialize BOA parameters, including the maximum iterations Itr , population size N , sensory modality c , power exponent a , and switch probability p .

Step 2:

Create an initial population matrix for butterfly positions.

Step 3:

Begin the main loop: repeat until the maximum number of iterations is reached.

For each solution in the population:

- Calculate its fitness value.
- Identify the current best solution g^* .

Step 4:

For each butterfly in the population:

- Generate a random number r within the range $[0, 1]$.
- If $r < p$:
 - Update the butterfly's position using Eq. 7.
- Otherwise:
 - Update the butterfly's position using Eq. 8.
- If the updated solution improves, update the population with this new solution.
- Adjust the sensory modality c if necessary.

Step 5:

Check if the iteration count has reached the maximum limit:

If not, increment the iteration counter by one and continue the loop.

End the loop when Itr is reached.

Return the best solution g^* found by the algorithm.

Fig. 2. Pseudo-code of BOA.

B. Flower Pollination Algorithm

FPA is a metaheuristic technique designed to solve complex optimization problems mimicking natural pollination. FPA follows the principles of two types of pollination found in nature: local and global pollination. Global pollination promotes exploration, allowing the algorithm to escape local optima, while local pollination emphasizes exploitation, speeding up convergence. The algorithm performs exploration or exploitation for each iteration based on a switching probability p , ensuring optimal solutions are found efficiently.

The algorithm searches for the global most attractive flower for each candidate solution in FPA, representing a flower in a d -dimensional space. This search is carried out to minimize the fitness function and locate the flower with the lowest fitness score, corresponding to the optimal solution. Four main steps are involved in the FPA's operation: initialization, fitness evaluation, pollination process, and selection. At first, the population of flowers F is defined using Eq. (10). Then, each solution X_{ij} within the defined search bounds is initialized using Eq. (11).

$$F = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{pmatrix} = \begin{pmatrix} x_{1,1} & \dots & x_{1,d} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \dots & x_{n,d} \end{pmatrix} \quad (10)$$

$$X_{ij} = x_{min} + (x_{max} - x_{min}) \cdot \mu \quad (11)$$

Where μ varies between 0 and 1. Eq. (12) refers to the objective function to evaluate fitness.

$$f(x), \quad X = (x_1, x_2, \dots, x_d) \quad (12)$$

The fitness of each flower is determined, and the current optimal solution g^* is identified, which has the lowest fitness value among all flowers. A random number $rand$ is determined by a uniform distribution between (0,1) for each flower. If $rand > p$, global pollination is accomplished as follows:

$$X_i^{t+1} = X_i^t + L \cdot (X_i^t - g^*) \quad (13)$$

Where L follows a Lévy flight distribution to simulate long-distance pollination, represented as follows:

$$L(\lambda) \sim \frac{\lambda \Gamma(\lambda) \sin(\pi\lambda/2)}{\pi} \cdot \frac{1}{s^{1+\lambda}} \quad (14)$$

With $\lambda=1.5$ and $s>0$ as the step size. If $rand \leq p$, local pollination occurs, and the flower's position is updated based on the positions of two randomly selected solutions as follows:

$$X_i^{t+1} = X_i^t + \epsilon \cdot (X_j^t - X_k^t) \quad (15)$$

Where ϵ comes from the [0,1] range, and X_j and X_k correspond to randomly chosen flowers. Each flower is rounded up to the closest valid position. The new positions are evaluated for fitness and each flower is updated if fitness has improved. The best solution g^* is also updated if a better solution is found.

C. Mutualism-based Hybrid Approach

MHA aims to enhance BOA's exploration and exploitation abilities by combining this algorithm with the FPA. Previous studies, such as BOA/DE and BOA/ABC, have shown that hybridizing BOA with other algorithms can balance exploration

(exploring the solution space broadly) and exploitation (improving solutions locally). However, these approaches still need more diversity, as they can become overly focused on high-performing solutions early on, leading to premature convergence.

The effectiveness of metaheuristic algorithms is determined by their capacity to maintain harmony throughout exploration and exploitation. Exploration refers to the search for solutions across the entire space while exploitation fine-tunes solutions around promising areas. Our approach introduces mutualism, a cooperative interaction between BOA and FPA to address this balance. This interaction allows the two algorithms to complement each other, with BOA providing global search capability and FPA enhancing local search.

MHA divides the population into two subgroups: butterflies and flowers. Each subgroup evolves independently, benefiting from BOA and FPA search properties. Dynamic switching probability is applied to determine when individuals should alternate between global and local searches, adapting based on the current optimization stage.

Mutualism in this context refers to the mutual benefit observed between butterflies and flowers in ecosystems. For instance, butterflies aid in pollination, while flowers provide nectar, benefiting both species. The hybrid algorithm simulates this mutualism using the Symbiotic Organisms Search (SOS) algorithm, which models ecosystem cooperation through mutualism, communalism, and parasitism. The mutualism stage, in particular, facilitates collaboration between two solutions by averaging their traits as follows:

$$Mutual_{agent} = \frac{x_i^t + x_j^t}{2} \quad (16)$$

The positions of the two interacting solutions X_i and X_j are then updated as follows:

$$X_i^{t+1} = X_i^t + rand[0,1] \times (g^* - Mutual_{agent} \times BF1) \quad (17)$$

$$X_j^{t+1} = X_j^t + rand[0,1] \times (g^* - Mutual_{agent} \times BF2) \quad (18)$$

Where g^* represents the most optimal solution in the population, $BF1$ and $BF2$ refer to attraction variables, and $rand[0,1]$ gives a random value to introduce variability. In addition, a dynamic switching probability p regulates the equilibrium between exploration and exploitation, calculated by Eq. (19).

$$p = 0.8 - 0.1 \times \frac{(Max_{iter} - iter)}{Max_{iter}} \quad (19)$$

Where Max_{iter} indicates the total number of iterations and $iter$ denotes the ongoing iteration. This probability decreases over time, favoring exploration early in the search and shifting towards exploitation as the algorithm progresses.

As shown in Fig. 3, through mutualism and adaptive switching, this hybrid strategy effectively incorporates the best features of both BOA and FPA, enhancing the diversity and convergence rate of the solution population. This hybrid method improves task scheduling performance by ensuring a comprehensive search in the solution space and optimizing the allocation of resources in cloud computing.

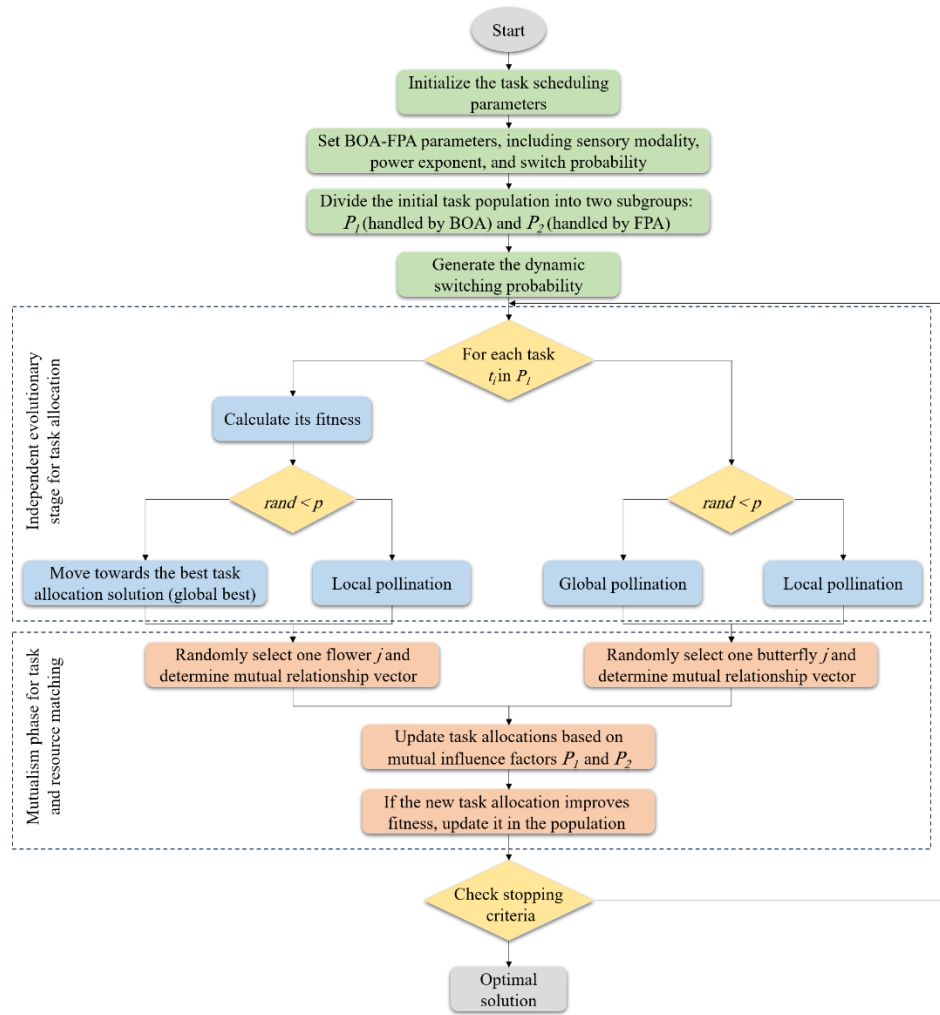


Fig. 3. Flowchart for proposed hybrid algorithm.

V. PERFORMANCE EVALUATION

To assess the performance of the algorithm, MHA, for cloud environments, simulations were performed on a synthetic dataset using MATLAB2018b on a PC powered by an Intel Core i5 3.5GHz CPU and 8GB RAM, running Windows 10. The experimental configuration, including the range of parameter values, is detailed in Table II. MHA was benchmarked against other algorithms, such as the Whale Optimization Algorithm (WOA), BOA, and FBA, using several key metrics: Makespan, Resource Utilization (RU), and imbalance degree.

TABLE II. EXPERIMENTAL CONFIGURATION AND PARAMETER RANGES

Parameter	Value range
Bandwidth	500 Mbps
VM Memory (RAM)	1 GB
VM CPU Speed (MIPS)	3,000 to 5,000
No. of VMs	20
Task Size (Million instructions)	1,000 to 20,000
No. of tasks	100 to 1,000

Makespan, the interval between task starts and endpoints, measures scheduling efficiency. Fig. 4 compares average makespan values across different algorithms. For 100 tasks, MHA achieved an average makespan of approximately 15.2, outperforming comparative algorithms. MHA maintained lower makespan values as task sizes increased to 500 and 1000, with 71.4 and 140.2, respectively. Regarding large-scale cloud scheduling, MHA is significantly better at handling larger task sets. Imbalance degree measures the stability and balance of workload distribution across VMs. A lower value indicates better balance, reducing overload risk. This metric is calculated as follows:

$$ID = \frac{ET_{max} - ET_{min}}{ET_{avg}} \quad (20)$$

Where ET_{min} and ET_{max} are the minimum and maximum execution times across VMs, and ET_{avg} is the average execution time. Fig. 5 shows ID comparisons for different algorithms. For 100 tasks, MHA achieved an imbalance degree of 0.71, lower than other algorithms. This trend of lower imbalance degree persisted as the number of tasks increased, demonstrating MHA's ability to maintain balanced workloads

across VMs. RU measures the extent of VM utilization during task scheduling and is given by:

$$RU = \frac{\sum_{j=1}^n ET_j}{\text{makespan} \times m} \quad (21)$$

Where ET_j refers to the execution time of each VM and m represents the number of VMs. Fig. 6 illustrates that MHA achieved the highest RU values, indicating better resource usage.

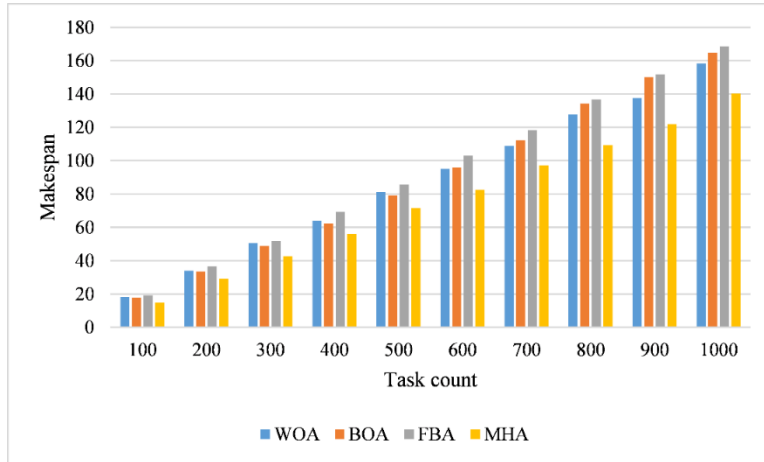


Fig. 4. Makespan results.

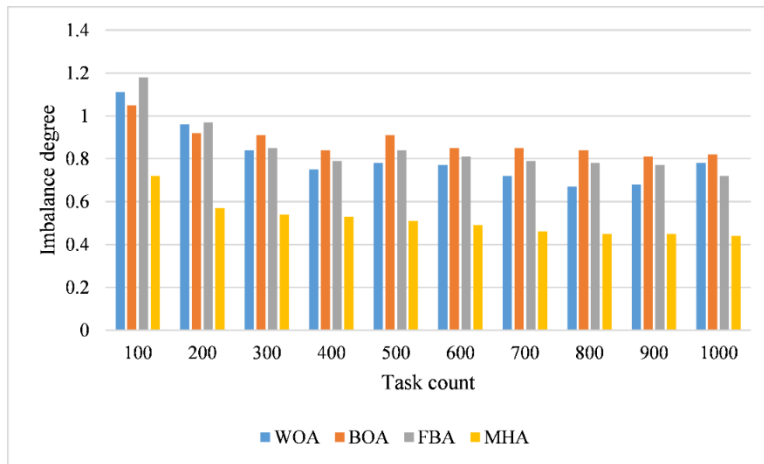


Fig. 5. Imbalance degree results.

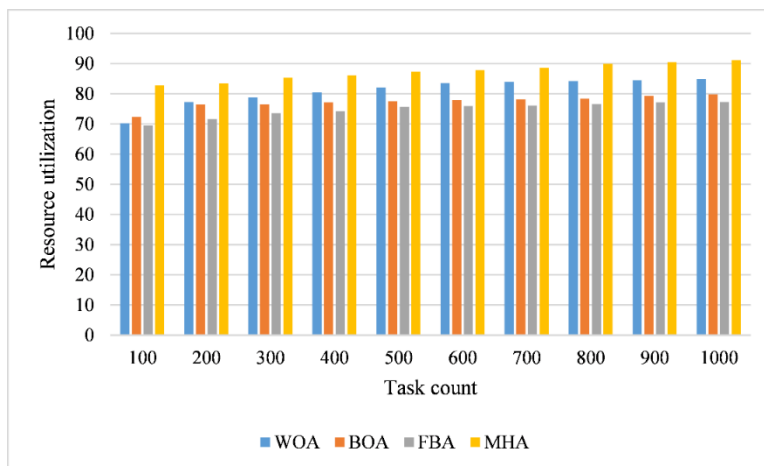


Fig. 6. Resource utilization results.

VI. DISCUSSION

The proposed hybrid algorithm significantly enhanced the challenge of cloud-based task scheduling. This section presents the implications of the results, the strengths of the proposed approach, and future avenues.

The hybridization of BOA and FPA through a mutualism-inspired mechanism addresses the individual limitations of each algorithm. BOA's tendency to converge prematurely is mitigated by FPA's enhanced diversity in exploration, while FPA's limited exploitation capabilities are bolstered by BOA's local search strengths. The introduction of an adaptive switching probability further enhances the balance between exploration and exploitation, allowing the algorithm to dynamically adjust its search strategy based on the progress of the optimization process. This innovative mechanism ensures robust performance, reducing the likelihood of stagnation in local optima and improving convergence speed.

Experimental results showed that the proposed hybrid algorithm achieved much better performance when compared to the benchmark methods on important metrics in this area, namely, makespan, resource utilization, and load balancing. Significance statistical tests are carried out that further establish the proposed algorithm's efficiency in handling diversified and dynamic challenges of task scheduling problems in cloud environments. The algorithm was found to work well with increased loads and much better performance according to scalability tests, hence its applicability to real-world applications.

The ability of the hybrid algorithm to optimize task scheduling has great implications for cloud computing environments. It can reduce operational costs, improve user satisfaction, and enhance system performance by efficiently allocating resources. Besides, it is a promising solution due to its adaptability and scalability in heterogeneous and dynamic cloud infrastructures.

Although the proposed algorithm outperforms the other methods by a great margin, some limitations must be conceded. Its performance is related to some parameters that might be fine-tuned in some situations. Therefore, Future work could address the proposition of automatic parameter-tuning mechanisms to make them more usable. Future work might also explore the effectiveness of the proposed hybrid approach for other optimization problems, such as load balancing in a distributed system or energy-aware scheduling.

VII. CONCLUSION

The paper proposed a new hybrid task scheduling algorithm called MHA, which combines BOA and FPA within a mutualism-based mechanism. MHA can effectively meet the most important challenges during Cloud platforms for effective task scheduling, such as minimizing makespan, maintaining workload balance across virtual machines, maximizing resource utilization, and improving overall scheduling performance. It achieves an excellent balance between exploration and exploitation through effective exploitation of BOA and FPA, with a guaranteed optimal distribution while the scheduling tasks continue to increase in scale and complexity. As reported from simulations, results show the outperformance

of the MHA compared to traditional algorithms, proven by comparisons, which always have the best makespan with reduced imbalance degree and resource utilization. More specifically, the rate of performance improvement proves that MHA has considerably improved scheduling efficiency.

The adaptive dynamic switching probability in MHA enables the algorithm to scale up efficiently for large task sizes, presenting a robust approach for real-world cloud computing environments where dynamic and efficient task allocation is paramount. These results reflect that MHA can solve the current cloud scheduling requirements and provide a base for further enhancements in resource allocation strategies. Future research will be done on further hybridization with other metaheuristic algorithms, deep learning usage in the process for predictive scheduling, or even extending MHA to multi-objective optimization frameworks. These will eventually enhance scalability, adaptability, and efficiency in task scheduling in complex cloud computing scenarios.

REFERENCES

- [1] V. Hayyolalam, B. Pourghebleh, M. R. Chehrehzad, and A. A. Pourhaji Kazem, "Single-objective service composition methods in cloud manufacturing systems: Recent techniques, classification, and future trends," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 5, p. e6698, 2022.
- [2] B. Pourghebleh, A. Aghaei Anvigh, A. R. Ramtin, and B. Mohammadi, "The importance of nature-inspired meta-heuristic algorithms for solving virtual machine consolidation problem in cloud environments," *Cluster Computing*, vol. 24, no. 3, pp. 2673-2696, 2021.
- [3] A. Katal, S. Dahiya, and T. Choudhury, "Energy efficiency in cloud computing data centers: a survey on software technologies," *Cluster Computing*, vol. 26, no. 3, pp. 1845-1875, 2023.
- [4] D. Mušić, J. Hribar, and C. Fortuna, "Digital transformation with a lightweight on-premise PaaS," *Future Generation Computer Systems*, 2024.
- [5] M. Saleem, M. Warsi, and S. Islam, "Secure information processing for multimedia forensics using zero-trust security model for large scale data analytics in SaaS cloud computing environment," *Journal of Information Security and Applications*, vol. 72, p. 103389, 2023.
- [6] V. Hayyolalam, B. Pourghebleh, A. A. Pourhaji Kazem, and A. Ghaffari, "Exploring the state-of-the-art service composition approaches in cloud manufacturing systems to enhance upcoming techniques," *The International Journal of Advanced Manufacturing Technology*, vol. 105, pp. 471-498, 2019.
- [7] K. Saidi and D. Bardou, "Task scheduling and VM placement to resource allocation in Cloud computing: challenges and opportunities," *Cluster Computing*, vol. 26, no. 5, pp. 3069-3087, 2023.
- [8] F. S. Prity, M. H. Gazi, and K. A. Uddin, "A review of task scheduling in cloud computing based on nature-inspired optimization algorithm," *Cluster computing*, vol. 26, no. 5, pp. 3037-3067, 2023.
- [9] M.-L. Chiang, H.-C. Hsieh, Y.-H. Cheng, W.-L. Lin, and B.-H. Zeng, "Improvement of tasks scheduling algorithm based on load balancing candidate method under cloud computing environment," *Expert Systems with Applications*, vol. 212, p. 118714, 2023.
- [10] G. Saravanan, S. Neelakandan, P. Ezhumalai, and S. Maurya, "Improved wild horse optimization with levy flight algorithm for effective task scheduling in cloud computing," *Journal of Cloud Computing*, vol. 12, no. 1, p. 24, 2023.
- [11] M. D. Tezerjani, M. Khoshnazar, M. Tangestanizadeh, and Q. Yang, "A Survey on Reinforcement Learning Applications in SLAM," *arXiv preprint arXiv:2408.14518*, 2024, doi: <https://doi.org/10.48550/arXiv.2408.14518>.
- [12] A. N. Malti, B. Benmammar, and M. Hakem, "Task Scheduling Optimization in Cloud Computing: A Comparative Study Between Flower Pollination and Butterfly Optimization Algorithms," in *2023 5th*

- International Conference on Pattern Analysis and Intelligent Systems (PAIS), 2023: IEEE, pp. 1-7.
- [13] S. Arora and S. Singh, "Butterfly optimization algorithm: a novel approach for global optimization," *Soft computing*, vol. 23, pp. 715-734, 2019.
- [14] X.-S. Yang, M. Karamanoglu, and X. He, "Flower pollination algorithm: a novel approach for multiobjective optimization," *Engineering optimization*, vol. 46, no. 9, pp. 1222-1237, 2014.
- [15] L. Abualigah and M. Alkhrabsheh, "Amended hybrid multi-verse optimizer with genetic algorithm for solving task scheduling problem in cloud computing," *The Journal of Supercomputing*, vol. 78, no. 1, pp. 740-765, 2022.
- [16] M. Tanha, M. Hosseini Shirvani, and A. M. Rahmani, "A hybrid meta-heuristic task scheduling algorithm based on genetic and thermodynamic simulated annealing algorithms in cloud computing environments," *Neural Computing and Applications*, vol. 33, pp. 16951-16984, 2021.
- [17] S. Mangalampalli, G. R. Karri, and U. Kose, "Multi Objective Trust aware task scheduling algorithm in cloud computing using Whale Optimization," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 2, pp. 791-809, 2023.
- [18] X. Fu, Y. Sun, H. Wang, and H. Li, "Task scheduling of cloud computing based on hybrid particle swarm algorithm and genetic algorithm," *Cluster Computing*, vol. 26, no. 5, pp. 2479-2488, 2023.
- [19] I. Behera and S. Sobhanayak, "Task scheduling optimization in heterogeneous cloud computing environments: A hybrid GA-GWO approach," *Journal of Parallel and Distributed Computing*, vol. 183, p. 104766, 2024.
- [20] P. Pabitha, K. Nivitha, C. Gunavathi, and B. Panjavaram, "A chameleon and remora search optimization algorithm for handling task scheduling uncertainty problem in cloud computing," *Sustainable Computing: Informatics and Systems*, vol. 41, p. 100944, 2024.
- [21] S. D. S. Mustapha and P. Gupta, "DBSCAN inspired task scheduling algorithm for cloud infrastructure," *Internet of Things and Cyber-Physical Systems*, vol. 4, pp. 32-39, 2024.
- [22] S. Gupta and S. Tripathi, "A comprehensive survey on cloud computing scheduling techniques," *Multimedia Tools and Applications*, vol. 83, no. 18, pp. 53581-53634, 2024.
- [23] O. L. Abraham, M. A. Ngadi, J. B. M. Sharif, and M. K. M. Sidik, "Multi-Objective Optimization Techniques in Cloud Task Scheduling: A Systematic Literature Review," *IEEE Access*, 2025.
- [24] S. Durairaj and R. Sridhar, "Coherent virtual machine provisioning based on balanced optimization using entropy-based conjectured scheduling in cloud environment," *Engineering Applications of Artificial Intelligence*, vol. 132, p. 108423, 2024.
- [25] S. A. Murad et al., "Priority based job scheduling technique that utilizes gaps to increase the efficiency of job distribution in cloud computing," *Sustainable Computing: Informatics and Systems*, vol. 41, p. 100942, 2024.
- [26] M. Alweshah, S. A. Khalailah, B. B. Gupta, A. Almomani, A. I. Hammouri, and M. A. Al-Betar, "The monarch butterfly optimization algorithm for solving feature selection problems," *Neural Computing and Applications*, pp. 1-15, 2022.
- [27] P. Chakraborty, S. Sharma, and A. K. Saha, "Convergence analysis of butterfly optimization algorithm," *Soft Computing*, vol. 27, no. 11, pp. 7245-7257, 2023.

Task Scheduling in Fog Computing-Powered Internet of Things Networks: A Review on Recent Techniques, Classification, and Upcoming Trends

Dongge TIAN

Department of Information Engineering, Hebei Chemical & Pharmaceutical College, Shijiazhuang 050000, China

Abstract—The Internet of Things (IoT) phenomenon influences daily activities by transforming physical equipment into smart objects. The IoT has achieved a wealth of technological innovations that were previously unimaginable. IoT application areas cover various sectors, including medical care, home automation, smart grids, and industrial operations. The massive growth of IoT applications causes network congestion because of the large volume of IoT tasks pushed to the cloud. Fog computing mitigates these transfers by placing resources near the edge. However, new challenges arise, such as limited computing power, high complexity, and the distributed characteristics of fog devices, negatively affecting the Quality of Service (QoS). Much research has been conducted to address these challenges in designing QoS-aware task scheduling optimization techniques. This paper comprehensively reviews task scheduling techniques in fog computing-powered IoT networks. We classify these techniques into heuristic-based, metaheuristic-based, and machine learning-based algorithms, evaluating their objectives, advantages, weaknesses, and performance metrics. Additionally, we highlight research gaps and propose actionable recommendations to address emerging challenges. Our findings offer a structured framework for researchers and practitioners to develop efficient, QoS-aware task scheduling solutions in fog computing environments.

Keywords—Internet of Things; task scheduling; fog computing; quality of service; network congestion; optimization

I. INTRODUCTION

The Internet of Things (IoT) phenomenon has changed the real world into a smart environment by turning everyday objects into smart objects/agents. This is accomplished by integrating sensors or microchips into these devices, along with internet connectivity [1]. These smart objects can independently interact and collect data, performing assigned duties [2]. The IoT is perceived as the upcoming model for ubiquitous computing and communication in the current technological environment. This ever-evolving environment is a network of billions of diverse, smart, connected devices that can revolutionize applications [3]. The applications of IoT range from individual home automation (smart homes) to overall city management (smart cities) [4]. It encompasses a variety of applications, including tracking high-precision agriculture and large-scale agricultural operations [5], monitoring individual building energy usage [6], and analyzing intelligent power grids [7].

The IoT also affects healthcare, providing personalized services for patients and the general public [8]. In addition, it

also drives automation in industries and provides business information [9]. Moreover, IoT applications exist in weather forecasting and monitoring tools locally and remotely [10]. Cloud, fog, and edge technologies enable deploying distributed data processing solutions essential for the IoT paradigm [11].

Fog computing is a variation of cloud computing that distributes data across several geographical regions [12]. It positions the processing and communication resources closer to the network boundary, where several fog hubs are located. Proximity to end-users and IoT devices enhances performance and responsiveness [13]. Many applications are limited by cloud-centric architectures, particularly those demanding real-time performance within smart cities and buildings [14]. In such conditions, most data require processing, analysis, and storage on remote cloud servers. This dependency on remote resources may have detrimental effects on response time, privacy, elasticity, and system integrity [15].

The emergence of delay-sensitive and location-aware applications similarly reveals the weaknesses of cloud-based approaches, which often fail to meet their demands for high efficiency and low latency [16]. The presence of fog layers near IoT objects within a smart city environment decreases latency. This feature enables fog computing to meet demanding latency criteria effectively. Fog computing functions as a supplementary layer to the cloud, allowing advanced applications and services to be created and implemented [17].

Integrating IoT, fog, and cloud computing paradigms requires efficient task scheduling techniques. While these intricate and vast ecosystems emerge, it becomes essential to improve the entire tone of the environment, reduce delays, and efficiently utilize resources. Task scheduling solutions are required to control and allocate computational processes in clouds and other processing nodes, such as edges and IoT gadgets. As the demand for such environments increases, developing innovative and highly effective task scheduling initiatives to improve the performance and efficiency of IoT, fog, and cloud-based systems becomes crucial. To tackle this challenge, the current research adopts a multifaceted approach. The main contributions of this paper are as follows:

- We offer a detailed classification of task scheduling algorithms based on their impact on Quality of Service (QoS) for both users and fog service providers, providing a clear framework for understanding various approaches.

- Our study includes an extensive review of existing research, evaluating objectives, advantages, weaknesses, performance metrics, computing environments, and future work, thus providing a holistic view of the field.
- We identify current research gaps in the area of QoS-aware task scheduling in fog computing, offering a clear direction for future studies to address these gaps and advance the field.
- We provide explicit and actionable recommendations for future research based on identified trends and gaps, guiding scholars and practitioners in their efforts to develop more effective and efficient task scheduling algorithms.

Various performance metrics evaluate task-scheduling techniques, resulting in optimal resource utilization and efficient system performance. The most important factors are makespan, representing the total time taken to execute tasks; energy consumption, evaluating the power efficiency of the system; throughput, measuring how many tasks can be efficiently processed within a certain amount of time; reliability, guaranteeing no failure during the execution of a task; and latency, reflecting the reflection of delay time during the processing of tasks. These metrics are important in developing the effectiveness of task-scheduling approaches for IoT networks powered by fog computing and serve as a basis for our evaluation and classification in the study.

This investigation will reveal emerging research issues and future research prospects. To ensure this methodology remains coherent and specific, the following questions serve as a guideline for this research.

- What are the key performance metrics used to assess different task scheduling techniques?
- What are the recent trends and potential prospects in research on task scheduling for fog computing?
- What approaches and parameters can be used to tune fog computing task scheduling algorithms to achieve optimal resource utilization, throughput, and reliability without compromising energy consumption, makespan, and delay?

The remaining sections are organized in the following manner. Section II presents the basic insights on the effect of the IoT and a brief introduction to fog computing. Section III presents the classification of task scheduling techniques, including heuristic-based, metaheuristic-based, and machine learning-based algorithms. Section IV summarizes results from previous studies and their implications. Section V discusses potential future research for bridging identified gaps. Finally, Section VI concludes with actionable recommendations to further advance task scheduling for fog computing.

II. BACKGROUND

A. IoT and its Impact

Numerous components, such as sensors, actuators, smartphones, and intelligent vehicles, are equipped with unique

identities in IoT environments [18]. IoT optimizes daily activities by utilizing data to facilitate remote access control and configuration via cloud-based platforms. However, the increasing number of IoT devices poses a challenge: effective load balancing across these devices is essential to ensure optimal network performance [19]. This task is complicated primarily due to changing traffic patterns and the lack of a centralized network structure. Therefore, non-optimal load balancing is a major concern, which has motivated researchers to develop IoT load balancing and routing solutions.

The rapid proliferation of gadgets in modern networks and infrastructures has created a highly complex digital world. These systems create content having various packet sizes, inter-packet arrival intervals, and transmission lengths [20]. Therefore, managing and controlling traffic have emerged as significant concerns in many areas, such as healthcare, data centers, big data, smart cities, and other fields. Different communication protocols have been adopted to accommodate these diverse networks. Nevertheless, the lack of defined data formats and protocols poses a major obstacle to traffic control in traditional IoT architectures. The absence of consistency in communication protocols used by different IoT devices impedes the comprehensive analysis of data gathered from several sources [21].

The main responsibilities involve coordinating data transfer, ensuring it is timely, and optimizing network utilization while reducing bottlenecks, delays, and inefficiencies. Traffic management extends beyond particular settings and includes intelligent healthcare systems, urban environments, big data applications, and numerous other areas. Urban settings and smart healthcare utilize networked sensors and gadgets to acquire and collect relevant information. Therefore, it is necessary to deploy smart traffic control strategies to satisfy QoS criteria for these diverse technologies. To effectively control traffic flows in healthcare and automated transportation systems while also considering environmental sustainability, it is necessary to build complex algorithms and real-time data analytics to handle mobile IoT devices.

The IoT grows through the interconnectedness of numerous devices and sensors, resulting in ever-increasing data communication. The large amount of data might cause network congestion, especially in wireless networks with a natural tendency to encounter transmission obstacles [22]. Congestion in an IoT environment arises when the quantities of transmitted data are larger than the capacities of available transmission resources. This phenomenon has significant implications for load balancing routing protocols. There are two main forms of congestion: link-level and node-level.

At the link level, congestion can be defined as the arrival rate of the packets is more than the rate at which the packets are served, resulting in buffer overflow. This scenario is similar to one in which the amount of water pouring in exceeds the drainage system capacity [23]. On the other hand, node-level congestion happens when many active sensor nodes transmit packets simultaneously on the same channel, causing interference and preventing successful transmission [24].

The limited energy resources of devices present a substantial obstacle to implementing IoT networks. Energy-

conscious routing protocols must confront complex issues associated with energy usage during sensing, data transmission, and receiving [25]. Data aggregation techniques can help reduce transmissions, but establishing the best level of aggregation is a complex operation that requires efficient solutions to maximize network device lifespan [26]. The natural diversity of IoT devices causes energy usage problems. These gadgets demonstrate notable disparities in computing power, energy profiles, and connectivity capabilities. Integrating these varied attributes into routing decisions is a considerable obstacle.

B. Introduction to Fog Computing

IoT deployment benefits from cloud computing in terms of computation, storage resources, and QoS constraints. It involves moving data to other servers at data centers, which is processed and returned to the end user [27]. Cloud computing also provides centralized virtual servers for data processing, storage, and analysis. These services are available on demand and are categorized as Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).

Nevertheless, IoT and the cloud components exchange more tasks and data, affecting the overall response time and resulting in network latency. The substantial physical separation between cloud-based IoT devices causes security problems. These delays can have severe consequences, especially in extremely sensitive tasks, such as real-time health monitoring, posing a danger to patients. Therefore, architectural plans are shifting from centralized data centers to distributed computing devices at the edge of the network. The decentralized nature of this mechanism also aims to eliminate the mentioned delays and security concerns.

Researchers have recently focused on investigating fog computing, a novel paradigm bridging the gap between IoT platforms and cloud computing. Fog computing leverages a decentralized architecture to reduce task transmission delays while upholding QoS requirements. Thus, this approach has developed into a sensible strategy for time-constrained operations within the IoT context. Unlike cloud computing, which relies on centralized servers to perform computations and relay outcomes to IoT devices, fog computing distributes these processes to fog nodes close to IoT devices to reduce latencies and enable increased response times in IoT applications.

The proximity of fog nodes to IoT devices provides further advantages, such as reduced total delay and increased protection of transmitted information. However, the resource limitations of fog devices necessitate offloading computationally intensive workloads to the cloud. Fig. 1 illustrates the overall structure, depicting the placement of IoT and edge devices with varying computational capabilities, cloud computing, and fog layers.

The goal of fog computing is to bring storage, transmission, and computing services closer to the network's edge. The proximity of the data center facilitates efficient data processing, low delays, and minimal bandwidth consumption. Fog devices, conversely, are characterized by lower processing, storage, and bandwidth capabilities due to their compact size. It is, therefore, important that all these constraints are considered when

scheduling tasks to have an effective scheduling strategy. This problem is defined as non-deterministic polynomial-time hard (NP-hard) due to the optimization of the problem with many parameters, including energy consumption, task deadlines, cost, makespan, and response time. This designation indicates that finding the optimal solution becomes computationally intractable as the problem grows.

Certain constraints are crucial for both end-users and system designers. For instance, task deadlines represent a critical QoS parameter for end-users, while energy usage of fog nodes is a QoS requirement of concern for the fog service provider. Consequently, assigning IoT tasks to fog nodes and the cloud necessitates careful consideration of these diverse constraints.

III. CLASSIFICATION OF TASK SCHEDULING TECHNIQUES

This research adopts a systematic review approach to analyzing and classifying various task scheduling techniques in IoT networks powered by fog computing. Related literature concerning task scheduling and QoS improvements was searched from reputable databases such as IEEE Xplore, Springer, and Elsevier. Techniques identified were further classified into heuristic-based, metaheuristic-based, and machine learning-based approaches and assessed based on the key performance metrics of latency, energy consumption, throughput, and reliability. Critical reviews have been done regarding the objectives, advantages, and limitations of each technique. Research gaps were identified, with recommendations for improvement also given for actionable purposes.

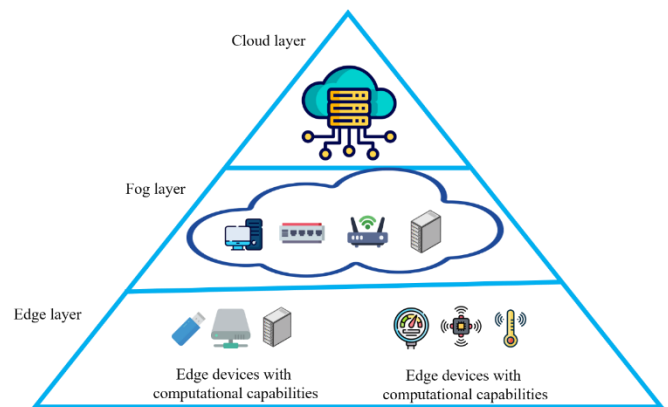


Fig. 1. IoT, edge, fog computing, and cloud computing architecture.

A. Heuristic-based Algorithms

As summarized in Table I, heuristic-based algorithms offer approximate solutions for computationally complex task scheduling problems within fog computing environments. These algorithms are specifically designed to efficiently manage the dynamic and heterogeneous characteristics inherent to fog environments.

Krivic, et al. [28] established the classification of IoT services, a crucial factor that directly influences scheduling algorithms. In addition, they introduced an innovative scheduling method that considers service context, user context, and processing devices. This approach allows for the efficient calculation of the most efficient schedule for executing service

components in a distributed fog-to-cloud context. The effectiveness of the suggested algorithm was confirmed by simulations, in which its distinguishing innovation and dynamic scheduling were particularly highlighted.

Aburukba, et al. [29] formulated scheduling IoT service queries as an optimization problem to shorten the total service request latency. They employed integer programming to model the problem; however, due to its NP-hard nature, this approach becomes impractical for large-scale scenarios. To solve this problem, they combined an individualized version of the Genetic Algorithm (GA) as an efficient heuristic for scheduling IoT tasks, considering holistic latency optimization. They evaluated the performance of the designed GA using an evolutionary simulation model, which reflects the inherent dynamism of real IoT environments.

Ibrahim, et al. [30] introduce a load-balanced and delay-aware scheduling model for fog computing environments, particularly for critical IoT applications. The developed mechanism prioritizes minimizing task execution delays and maximizing task acceptance rates. Furthermore, the mechanism is designed to output an optimal outcome by ensuring uniform and minimal load imbalances to fog resources to improve resource utilization and lower the average response time.

Wireless Sensor Networks (WSNs) generate many tasks with varying priorities and durations in healthcare monitoring,

transmitting them concurrently to fog computing platforms. This necessitates the implementation of an effective task scheduling algorithm capable of accurately prioritizing tasks according to their priority, regardless of their duration. Aladwani [31] introduced the Tasks Classification and Virtual Machines Categorization (TCVC) approach to improve the performance of static task scheduling algorithms. It classifies tasks by importance to patient health. The new method divides incoming IoT tasks into three levels of importance: high, average, and low.

Effective resource management is paramount for achieving optimal system performance, particularly concerning latency, within fog-cloud computing environments. Resource planning in such environments presents a computationally complex problem, classified as NP-hard. Khezri, et al. [32] investigated the optimization challenges associated with scheduling data-intensive jobs within fog-cloud based IoT systems, specifically focusing on maximizing job longevity. The proposed method starts by formulating the problem into an Integer Linear Programming (ILP) optimization scheme. Subsequently, a heuristic algorithm, Data-Locality Aware Job Scheduling in Fog-Cloud (DLJSF), is designed. Performance evaluations demonstrated that the proposed DLJSF algorithm achieves results closely approximating those obtained through the ILP model, with an average deviation of only 13.

TABLE I. SUMMARY OF HEURISTIC-BASED ALGORITHMS

Reference	Approach	Advantage	Disadvantage
[28]	Dynamic scheduling algorithm considering processing devices, user context, and service context	Significant increase in performance efficiency; adaptability to time-varying network conditions and QoS parameter changes	Complexity in implementation due to the need for constant monitoring and adjustments to QoS parameters
[29]	Individualized genetic algorithm for minimizing overall service request latency	Reduced overall latency	The genetic algorithm might be computationally expensive for large-scale scenarios
[30]	Load balancing mechanism prioritizing task execution delays and task acceptance rate	Improved resource utilization and lower average response time	Potential overhead in managing load balancing and ensuring task acceptance rates
[31]	Tasks classification and virtual machines categorization using the MAX-MIN scheduling algorithm	Improved performance in algorithm complexity, resource availability, execution time, waiting time, and finish time	The static nature of the approach might not handle highly dynamic environments effectively
[32]	Data-locality aware task scheduling in fog-cloud derived from an ILP optimization model	Results closely approximate the ILP model with a 13% average deviation; outperforms local processing by 99.16%	The ILP-based model might be complex, computationally intensive, and requires effective data locality awareness.
[33]	Priority-Aware Semi-Greedy (PSG) and PSG with Multistart (PSG-M) procedures	High performance on makespan, deadline violation time, energy consumption, and deadline compliance	Balancing multiple objectives (energy usage and QoS) can be challenging and may require fine-tuning for different scenarios
[34]	Heuristic for dynamic resource scheduling and allocation of real-time IoT workflows	Superior performance compared to static provisioning; real-time data awareness	Complexity in implementing dynamic resource provisioning and maintaining real-time data awareness.

Azizi, et al. [33] explored the issue of task scheduling in fog computing to find a compromise between reducing energy consumption in fog points and maintaining the QoS standards for IoT operations. They mathematically model the problem of optimizing these conflicting criteria as a multi-objective optimization problem. They also focused on minimizing deadline violation time in their approach, which they handled by proposing two new semi-greedy based algorithms: Priority-Aware Semi-Greedy (PSG) algorithm and a PSG with Multistart (PSG-M) procedure.

Stavrinides and Karatza [34] proposed a dynamic resource provisioning mechanism for cloud resources within a three-layer IoT-fog-cloud framework. This approach prioritizes real-time data awareness and dynamic scaling to optimize resource allocation. Additionally, they introduced a heuristic for scheduling instant IoT tasks. The efficacy of the suggested scheme was measured through experiments that compared its performance against a static provisioning strategy. These simulations employed various workload patterns to assess the impact on the framework's performance under different provisioning scenarios.

B. Metaheuristic-based Algorithms

Building upon heuristic approaches, metaheuristic-based algorithms leverage advanced strategies to efficiently explore and exploit the search space. These algorithms strive to identify near-optimal solutions while exhibiting improved convergence rates, as shown in Table II.

Abdel-Basset, et al. [35] introduced an energy-conscious task scheduling method for fog environments based on the Harris Hawks Optimization (HHO) algorithm integrated with Local Search, called HHOLS. HHOLS optimizes the QoS in IoT applications by focusing on energy efficiency. Their work commences with a detailed description of the highly virtualized layered fog computing model, emphasizing its heterogeneous architectural characteristics. To address the non-linear character of the task scheduling problem, they incorporated a scaling and normalization stage to adapt the standard Harris Hawks optimization algorithm.

Service execution plays a vital role in IoT networks, which is also an important problem in scheduling services in fog

computing. Consequently, fog infrastructure provides the execution environment for devices with limited computational capability. A fog environment comprises many fog nodes that can be some-edge servers, cloudlets, small-size ISPs, and caching nodes offering user-requested services. Najafizadeh, et al. [36] offered a privacy-preserving task scheduling architecture for IoT systems based on service execution to overcome these problems. In this design, a multi-objective algorithm is proposed to lower both service cost and execution time simultaneously.

Abd Elaziz, et al. [37] suggested AEOSSA, an alternative task scheduling approach for managing IoT tasks within a cloud-fog computing context. This approach builds upon a modified Artificial Ecosystem-Based Optimization (AEO) algorithm. The modified algorithm incorporates operators derived from the Salp Swarm Algorithm (SSA) to augment the exploitation capabilities of AEO in the search for optimum results for the task scheduling problem. An evaluation of AEOSSA is performed on some synthetic and real datasets that include a variety of computation sizes.

TABLE II. SUMMARY OF METAHEURISTIC-BASED ALGORITHMS

Reference	Approach	Advantage	Disadvantage
[35]	Harris hawks optimization with local search	Optimizes QoS in IoT applications focusing on energy efficiency	Complexity due to normalization, scaling, and local search phases
[36]	Multi-objective algorithm for privacy-preserving task scheduling	Minimizes service execution time and cost; maintains the privacy of IoT devices; performs well across different service composition complexities	Higher computational overhead due to multi-objective optimization
[37]	Modified artificial ecosystem-based optimization with salp swarm algorithm	Superior performance in makespan and throughput; effective for synthetic and real datasets	Potentially higher resource consumption due to extensive exploitation capabilities
[38]	Energy-aware model with arithmetic optimization algorithm and marine predators algorithm	Significant reductions in energy consumption and makespan	Increased complexity and potential for higher computational cost
[39]	Multi-cloud to multi-fog architecture with dynamic threshold strategy	Decreases service latency and increases fog node efficiency; achieves energy balance	Complexity in implementation and real-time dynamic scheduling
[40]	CHMPAD algorithm combining marine predators algorithm and disruption operator	Prevents local optimization; improves exploitation properties; significant reductions in makespan and throughput	Increased complexity and resource demands
[41]	Two-tiered approach with PSO and particle swarm genetic joint optimization artificial bee colony	Optimal load balancing and task scheduling; lower delay and energy consumption	Higher computational complexity due to multi-tiered strategy
[42]	Multi-tiered scheduling framework with Naïve Bayes classifier	Effective task classification and placement; enhances QoS parameters	Requires precise training data for classifier accuracy
[43]	Directed non-dominated sorting genetic algorithm	Minimizes energy consumption and response times; balances exploration and exploitation	Potential for higher computational overhead
[44]	Hunger games search with marine predators algorithm	Reduces energy consumption and makespan; effective for various workload traces	Complexity in algorithm integration and evaluation
[45]	Multi-objective gravitational search algorithm with star-quake operator	Reduces makespan, energy consumption, and cost; prevents local optima	Increased computational complexity and resource demands
[46]	Various algorithms, including machine learning and nature-inspired metaheuristics	Optimizes resource allocation, minimizes energy consumption and latency, and meets deadlines; consistent performance improvements	Varied complexity depending on the specific algorithm used

Abd Elaziz, et al. [38] overcame the task scheduling issue in fog computing by proposing an energy-aware model using a variant of the Arithmetic Optimization Algorithm (AOA) known as AOAM. Optimizing the makespan metric, they ensured that user QoS was the top priority. The authors integrated search operators inspired by the Marine Predators Algorithm (MPA) to overcome the limitations of the traditional

AOA. This modification encourages a wider range of solutions and avoids becoming stuck in suboptimal solutions. The efficacy of the proposed AOAM was validated through simulations that employed various parameters.

Luo, et al. [39] have suggested a unique multi-cloud to multi-fog model that involves two service models with

containerization technology, aiming to optimize fog resource usage and control service latency. On the other hand, the task scheduling algorithm provided in their proposal is particularly suitable for achieving an energy balance. Additionally, the algorithm takes the terminal device transmission energy requirements into account. It applies a flexible threshold algorithm to achieve real-time request scheduling while ensuring an energy balance state of the terminal device, thereby effectively avoiding transmission delays.

Attiya, et al. [40] presented a novel fog computing application-aware task scheduling algorithm called CHMPAD. It overcomes existing issues with the Chimp Optimization Algorithm by combining two key components of different algorithms: the Marine Predators Algorithm and a disruption operator. CHMPAD aims to prevent local optimization and improve the exploitation properties of the base ChOA algorithm. The applicability and effectiveness of CHMPAD are evaluated through extensive experiments performed on synthetic and real-world workloads.

Liu, et al. [41] developed a novel resource scheduling strategy for fog computing environments. This two-tiered approach optimizes load balancing and task scheduling to decrease energy consumption and execution time. The first tier leverages the Particle Swarm Optimization (PSO) algorithm for balancing loads within a fog cluster. This optimization seeks to identify the ideal distribution of tasks across fog nodes, minimizing computation time and energy usage. Building upon this foundation, the authors propose a novel Particle Swarm Genetic Joint Optimization Artificial Bee Colony (PGABC) algorithm. PGABC tackles the challenge of task scheduling across multiple fog clusters, utilizing the time and energy consumption data obtained from the initial load balancing phase.

Kaur, et al. [42] proposed a multi-tiered scheduling framework for managing IoT application tasks. This framework prioritizes QoS parameters to achieve optimal task placement. The framework operates on two levels: fog environment and fog node selection. The specific fog environment in which the task will be executed is set at the first level. Several factors, such as availability, physical distance, latency, and throughput, are used to choose an environment. After choosing the fog environment, a particular fog node is selected for analysis. They implemented a Naïve Bayes classifier to classify the task category (Compute-intensive, Memory-intensive, or GPU-intensive) based on the probability triad (C, M, G).

Mousavi, et al. [43] formulated a constrained bi-objective optimization problem for task scheduling in fog computing environments. This formulation aims to achieve two critical goals simultaneously: minimizing server energy consumption and reducing overall response times. To address this challenge, the authors proposed a novel Directed Non-dominated Sorting Genetic Algorithm (D-NSGA-II). This algorithm builds upon the foundation of NSGA-II, a well-established multi-objective optimization technique. The key innovation lies in the introduction of a new recombination operator. This operator empowers D-NSGA-II to regulate the selection pressure exerted on candidate solutions, thereby striking a balance

between the algorithm's exploration and exploitation capabilities.

Attiya, et al. [44] proposed a novel task scheduling algorithm, HGSMMPA, specifically designed for cloud-fog computing environments within the IoT domain. Their approach leverages the Hunger Games Search (HGS) algorithm as a foundation. The authors incorporated elements from the MPA to enhance the exploitation capabilities inherent in HGS. The efficacy of HGSMMPA was validated through experimental evaluations that employed various workload traces, both synthetic and real-world. The results convincingly demonstrate the superiority of HGSMMPA compared to existing scheduling algorithms.

Ahmadabadi, et al. [45] introduced a novel multi-objective task scheduling approach for fog-cloud computing systems. Their approach addresses three critical objectives simultaneously: minimizing monetary cost, energy consumption, and makespan. To achieve these goals, the authors proposed a new multi-objective function that incorporates all three objectives. Furthermore, they introduced a novel operator called star-quake, specifically designed for the Multi-Objective Gravitational Search Algorithm (MOGSA). This operator balances the algorithm's capabilities, such as selection pressure, exploration, and exploitation.

Alsamarai, et al. [46] have significantly improved task scheduling in fog-cloud computing environments for IoT applications. Their proposed algorithms address various challenges, including optimizing resource allocation (e.g., CHMPAD, DLJSF), minimizing energy consumption and latency (e.g., PGABC, Quality-aware Energy Efficient Scheduling), and meeting task deadlines (e.g., Bandwidth-Deadline Algorithm). They leverage a variety of techniques, including machine learning (PSO, ANN), heuristic approaches (genetic algorithms), and nature-inspired metaheuristics (Gravitational Search Algorithm, Ant Colony Optimization) to achieve these improvements.

C. Machine Learning-based Algorithms

Machine learning-based algorithms exploit historical data and learning models to predict near-optimal scheduling decisions [47]. Techniques such as reinforcement learning and neural networks have demonstrated significant potential in adapting to the dynamic characteristics of fog computing environments [48], as shown in Table III.

Bhatia, et al. [49] proposed a novel quantized approach for scheduling heterogeneous tasks within fog computing applications. The approach is built on a node-specific metric, the Node Computing Index (NCI), used to measure individual fog nodes' computational capability. They also proposed a QCI-Neural Network Model that forecasts the best available fog node for real-time execution of heterogeneous tasks. To validate the proposed approach, the authors conducted simulations in different scenarios.

Ali, et al. [50] tackled enhancing the overall efficiency of executing tasks for IoT applications. Their methodology revolves around selecting real-time jobs well-suited for execution at the fog layer. A fuzzy logic-based task scheduling algorithm is modeled for a fog-cloud computing environment.

This algorithm offers a smart scheme of allocating submitted tasks to the processing units within the fog layer. Heterogeneous resources can be found in fog.

Lim [51] addressed the low latency task execution in small-scale fog computing deployments. Their approach is based on a novel task scheduling strategy using partitioned Artificial Neural Networks (ANNs). Such partition allows parallel learning and hyperparameter optimization across different edge servers. This parallelism significantly reduces scheduling times and contributes to achieving desired service level objectives.

Aburukba, et al. [52] proposed a task scheduling solution for a three-tier fog computing architecture. This approach prioritizes maximizing the number of requests that meet their deadline requirements. To achieve this goal, the authors introduce an optimization model formulated using Mixed Integer Programming (MIP). This model aims to minimize the number of missed deadlines. The efficacy of the model was validated using an exact solution technique. However, the authors acknowledge that the task scheduling problem is NP-hard, rendering exact solutions impractical for typical fog computing environments due to problem size.

TABLE III. SUMMARY OF MACHINE LEARNING-BASED ALGORITHMS

Reference	Approach	Advantage	Disadvantage
[49]	Node computing index and QCI-neural network model	Significantly improved performance in execution delay, sensitivity, and precision; suitable for heterogeneous tasks	High computational complexity may require substantial training data.
[50]	Fuzzy logic-based task scheduling algorithm	Outperformed existing algorithms in task success ratio, makespan, average turnaround time, and delay rate; efficient for heterogeneous resources	Potential complexity in defining fuzzy rules may not scale well with large task sets.
[51]	Partitioned artificial neural network	Reduced scheduling times, maintained low energy consumption, achieved desired service level objectives	Limited scalability to larger fog computing environments, potential overhead in partitioning.
[52]	Mixed Integer Programming and genetic algorithm	Significant reduction in missed deadlines, effective for NP-hard scheduling problems; superior performance compared to round-robin and priority scheduling	The exact solution technique is impractical for large problem sizes, and the heuristic approach may not always find the global optimum.
[53]	Statistical techniques (moving averages, Heikin-Ashi patterns)	Enhanced precision in scheduling times, optimized task allocation across edge and fog nodes	The applicability of financial patterns to computing tasks may not be universally effective, and there is potential for increased computational overhead.
[54]	K-Means clustering and fuzzy logic	Accurate identification of groups, adapts to changing task distributions, improved execution time, response time, and network usage	Complexity in implementation and potential high computational overhead in large-scale dynamic environments
[55]	Distributed deep reinforcement learning with asynchronous proximal policy optimization	Fast convergence rate, high flexibility, upgradeability, and better time complexity in execution	Greedy nature of existing techniques and complexity in managing distributed experience trajectories
[56]	A2C-DRL based real-time task scheduling for edge-cloud environments	Simultaneous learning at multiple servers, flexibility with assignable hyperparameters, and superior load balancing	Complexity in defining reward functions and update policies
[57]	DRL-based algorithm for scheduling IoT applications	Adaptive response time, load balancing, significant cost reduction in execution and load balancing	Initial training phase might be resource-intensive, complexity in implementation on different platforms

The rapid rise in bandwidth requirements and computational load of the IoT has created opportunities for fog computing. However, maintaining the QoS of the data transfer process at an efficient cost in fog-based IoT networks remains a significant challenge. Potu, et al. [53] propose a novel scheduling algorithm that optimizes task allocation across edge and fog nodes. The proposed model integrates various statistical techniques, including moving averages and Heikin-Ashi patterns frequently employed in financial markets to visualize trends.

The basic scheduling strategies designed for special global cloud model do not really cope with static nature, heterogeneity, and resource constraints of the fog nodes. Sheikh, et al. [54] have addressed these limitations through the development of a new machine learning based approach that is aimed at dynamically allocating tasks in respect of the evolving status in the fog environment. Their approach builds on basics of K-Means clustering algorithm substantiated by fuzzy logic,

which can be considered an example of an unsupervised learning. Overall, this approach economically categorizes fog nodes based on resource and workload distribution. The proposed method builds on the strong points of the K-Means clustering that provides accurate identification of groups and fuzzy logic that allows one to adapt to changes concerning the distribution of tasks among the fog nodes.

Deep Reinforcement Learning (DRL) has recently gained traction in addressing complex service offloading problems. However, existing techniques are greedy in nature and are primarily designed for centralized problem formulation which results in slow convergence towards the global solution. In addition, data dependencies that are preconceived and QoS requirements inherent within the service components do not facilitate offloading. To overcome such limitations, Goudarzi, et al. [55] developed a distributed DRL strategy formulated through an actor-critic architecture named Asynchronous Proximal Policy Optimization (APPO). Thereby, it contributes

to creating a multitude of possible distributed experience trajectories to take place. Moreover, the authors use off-policy correction methods we have reviewed include PPO clipping and V-trape so as to enhance the rate of convergence to the optimal service offloading solutions.

Resource management in mobile edge and cloud systems often presents complex online decision-making challenges. Effective solutions necessitate real-time understanding of both workload and environment to facilitate the efficient utilization of distributed resources. However, geographically dispersed resources, limited capacity, unpredictable task characteristics, and network hierarchy inherent to edge environments significantly hinder efficient job scheduling. The above dynamic scenarios make heuristic-based methods inadequate since they are not easy to generalize or modify as would be required at times. One such unutilized yet potentially very beneficial technique is the DRL that names Advantage Actor-Critic (A2C). A2C learns quickly in environments with little data while DRL gains its knowledge from situations within the environment and applies them to make a decision. To address these challenges, Lu, et al. [56] propose an A2C-DRL based real-time task scheduling technique specifically designed for stochastic edge-cloud environments.

Wang, et al. [57] introduced a Deep Reinforcement Learning (DRL)-based algorithm for scheduling IoT applications, termed DRLIS. This approach is targeted to provide adaptive and efficient response time for wide range of IoT applications as well as the load balancing among the edge/fog servers. The authors incorporated DRLIS as an operational scheduler in the FogBus2, which is a function-as-a-service platform in the development of moving from edge to fog to cloud serverless computing model. The results of varied experiences indicate that the DRLIS bears a higher impact on the improvement of the execution cost of IoT applications.

IV. RESULTS AND DISCUSSION

A review and analysis of different task-scheduling techniques reveal tremendous variability in their performance based on the underlying methodologies and QoS metrics they try to optimize. Due to their simplicity, heuristic-based algorithms utilize low computational overhead; thus, they can be applied to only small-scale and resource-limited environments. Most of these techniques fail to optimize multiple QoS parameters like latency, energy consumption, and throughput simultaneously; hence, their usage is quite impractical in dynamic and large-scale fog computing scenarios. Contrarily, metaheuristic-based algorithms demonstrate much stronger adaptability in finding near-optimal solutions for complex scheduling problems. Despite the better performance, these algorithms usually introduce higher computation overhead, which may not be affordable for real-time applications.

While advanced algorithms based on machine learning leverage predictive and adaptive capabilities to optimize task scheduling dynamically, techniques such as reinforcement learning and deep neural networks have been promising in achieving significant reductions in latency and energy consumption while maintaining high throughput. For example, reinforcement learning-based models can predict the pattern of

task arrivals and resource availability to enable proactive scheduling. However, most machine learning techniques are implemented in a fog computing environment with extensive training in complex data computation resources and feature engineering, challenging widespread adaptation. Besides, machine learning interpretability may be one of the barriers if transparency in decision-making is essential.

Comparative analyses reveal that no method covers the fog computing-powered IoT network to date for all the challenges combined. Instead, heuristic and metaheuristic algorithms are more appropriate for scenarios with specific resource constraints, while machine learning-based approaches best apply to dynamic and complex environments. The hybridization of such techniques, though in very few instances, points out a bright future direction that can leverage the simplicity and efficiency of the heuristic approach together with adaptability and optimization capabilities from the metaheuristic and machine-learning-based techniques. Furthermore, much emphasis on standard frameworks is required to evaluate the benchmark techniques with regard to task scheduling properly; this ensures coherence in the performances reported through different scenarios. These insights emphasize the vital need for further research that should result in novel hybrid techniques applicable to the new demands put forward by IoT systems driven by fog computing.

Through extensive analysis of various methods for task scheduling in fog-cloud computing environments for IoT applications, several research gaps and limitations in prior studies have been identified. These limitations can include high run times, failure to meet the study's objectives, or negative impacts on other performance metrics. Common shortcomings include missing details on simulation parameters, comparisons with outdated algorithms, using small datasets for evaluation, omitting definitions of evaluation parameters and equations, neglecting relevant evaluation factors, and lacking results to support performance claims.

A critical limitation identified is using outdated benchmark algorithms for comparison in some studies. This makes it difficult to assess the efficiency of the proposed methods definitively. Additionally, several studies employed small datasets (fewer than 100 tasks) or omitted data set size information entirely. This raises concerns about the proposed algorithms' ability to handle real-world workloads with high throughput.

Our analysis also revealed a focus on specific objectives and performance metrics. Energy consumption emerged as the primary objective in many studies, followed by minimizing makespan, delay, and cost. Conversely, response time, resource utilization, deadline violation, security, and reliability received less attention. This focus is reflected in the most studied metrics: makespan, energy consumption, and cost. Most of the investigated algorithms were multi-objective, focusing on optimizing combinations like makespan and cost, makespan and energy, or delay and energy simultaneously.

V. FUTURE DIRECTIONS

As fog computing continues to evolve, several emerging trends and research directions are shaping the landscape of task

scheduling algorithms. These advancements aim to enhance the efficiency, scalability, and adaptability of fog computing systems to meet the growing demands of IoT applications. The future of task scheduling in fog computing is poised to leverage more sophisticated machine learning and Artificial Intelligence (AI) techniques. Reinforcement learning, deep learning, and federated learning are expected to play a significant role in developing more adaptive and intelligent scheduling algorithms. These approaches can enable real-time learning and decision-making, improving the allocation of resources and the overall performance of fog environments.

The collaboration between edge and cloud resources is anticipated to become more seamless, providing a hybrid model that optimally distributes tasks based on their computational and latency requirements. Future research will focus on developing algorithms that dynamically balance the load between edge and cloud, considering network conditions, energy consumption, and application-specific constraints. Energy efficiency will remain a critical concern in fog computing, particularly with the increasing number of connected devices and data-intensive applications. The research will continue to explore energy-aware scheduling algorithms that minimize power consumption without compromising performance. Sustainable computing practices, such as using renewable energy sources and energy-harvesting techniques, will also gain more attention.

With the proliferation of IoT devices and the sensitivity of the data they generate, ensuring security and privacy in task scheduling is paramount. Future research will delve into developing secure scheduling algorithms that incorporate encryption, anonymization, and other privacy-preserving techniques. These solutions must safeguard data integrity and confidentiality while maintaining efficient resource utilization. Integrating real-time analytics and predictive modeling into task scheduling algorithms will enhance responsiveness and accuracy. Using historical data and real-time monitoring, these algorithms can predict workload patterns, detect anomalies, and proactively adjust resource allocation, improving system reliability and performance.

Future scheduling algorithms will increasingly adopt multi-objective optimization techniques to balance various conflicting performance metrics, such as latency, throughput, energy consumption, and cost. Research will focus on developing algorithms that can effectively navigate the trade-offs between these objectives, providing optimal solutions that meet diverse application requirements. As quantum computing technologies mature, their integration into fog computing task scheduling could revolutionize the field. Quantum algorithms have the potential to solve complex optimization problems more efficiently than classical algorithms, offering unprecedented improvements in scheduling performance and resource utilization.

Developing standardized frameworks and protocols for task scheduling in fog computing ensures interoperability between different devices and platforms. Future research will explore ways to create universally accepted standards that facilitate seamless integration and collaboration across heterogeneous fog and edge environments. User-centric and context-aware

scheduling algorithms that consider the specific needs and preferences of end-users will become more prevalent. These algorithms will consider contextual information, such as user location, device capabilities, and application-specific requirements, to deliver personalized and efficient task scheduling solutions.

Federated learning, a decentralized machine learning approach where model training occurs locally on edge devices, will gain prominence in fog computing environments. The research will explore federated learning techniques for collaborative model training across distributed edge nodes, enabling privacy-preserving and resource-efficient machine learning. These approaches will empower edge devices to perform predictive analytics and decision-making tasks autonomously without relying heavily on centralized cloud servers. The integration of blockchain technology into task scheduling algorithms will enhance security, transparency, and trust in fog computing environments. Future research will investigate blockchain-based scheduling mechanisms that ensure verifiable task execution, prevent tampering or manipulation of scheduling decisions, and enable secure peer-to-peer transactions between edge devices. These blockchain-enabled solutions will facilitate decentralized task allocation and resource sharing while preserving data integrity and privacy.

VI. CONCLUSION

IoT technology applies to promote immense interconnectedness to data-driven functions like smart homes, cities, industrial automation, and health services. At the time of growth, this explosive creation of data created newer challenges for traditional models in cloud computing: latencies at high traffic volume and constricts scalability. Emerging as complementary mechanisms to remedy several limitations brought upon by traditional clouds in view of the huge impact created by IoT sensors continuously generating huge loads of information that requires computation, sometimes really urgent, Fog Computing extends the computational processes further toward network boundaries.

The present research gave an extensive review of different techniques for task scheduling in fog computing environments and broadly classified these techniques into three main categories, namely heuristic-based approaches, metaheuristic-based methods, and machine learning-based approaches. This research further analyzed the effects of different techniques on key QoS metrics related to latency, energy consumption, makespan, and reliability and, therefore, provided pragmatic insights into strengths, weaknesses, and applicability. The results revealed that these techniques can adapt dynamically to changing network conditions and workload demands, optimizing resource utilization and service quality in fog-enabled IoT systems.

Apart from indicating gaps, the study also identified several innovative solutions with regard to needs in proposals on hybrid techniques, along with standardized frameworks concerning the evaluation perspective. These lessons then provided concrete guidelines for both the researcher and the practitioner in creating algorithms on next-generation task scheduling at fog computing-powered IoT with efficiency and scalability, thus

guaranteeing enhanced QoS toward paving the right path for R-IoT.

REFERENCES

- [1] B. Pourghebleh, V. Hayyolalam, and A. A. Anvigh, "Service discovery in the Internet of Things: review of current trends and research challenges," *Wireless Networks*, vol. 26, no. 7, pp. 5371-5391, 2020.
- [2] B. Pourghebleh and N. J. Navimipour, "Data aggregation mechanisms in the Internet of things: A systematic review of the literature and recommendations for future research," *Journal of Network and Computer Applications*, vol. 97, pp. 23-34, 2017, doi: <https://doi.org/10.1016/j.jnca.2017.08.006>.
- [3] B. Pourghebleh, K. Wakil, and N. J. Navimipour, "A comprehensive study on the trust management techniques in the Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9326-9337, 2019, doi: <https://doi.org/10.1109/JIOT.2019.2933518>.
- [4] A. Ullah et al., "Smart cities: The role of Internet of Things and machine learning in realizing a data-centric smart environment," *Complex & Intelligent Systems*, vol. 10, no. 1, pp. 1607-1637, 2024.
- [5] S. Atalla et al., "Iot-enabled precision agriculture: Developing an ecosystem for optimized crop management," *Information*, vol. 14, no. 4, p. 205, 2023.
- [6] R. Selvaraj, V. M. Kuthadi, and S. Baskar, "Smart building energy management and monitoring system based on artificial intelligence in smart city," *Sustainable Energy Technologies and Assessments*, vol. 56, p. 103090, 2023.
- [7] N. Renugadevi, S. Saravanan, and C. N. Sudha, "IoT based smart energy grid for sustainable cities," *Materials Today: Proceedings*, vol. 81, pp. 98-104, 2023.
- [8] F. Kamalov, B. Pourghebleh, M. Gheisari, Y. Liu, and S. Moussa, "Internet of medical things privacy and security: Challenges, solutions, and future trends from a new perspective," *Sustainability*, vol. 15, no. 4, p. 3317, 2023, doi: <https://doi.org/10.3390/su15043317>.
- [9] K. C. Rath, A. Khang, and D. Roy, "The Role of Internet of Things (IoT) Technology in Industry 4.0 Economy," in *Advanced IoT Technologies and Applications in the Industry 4.0 Digital Economy*: CRC Press, 2024, pp. 1-28.
- [10] Ş. M. Kaya, B. İşler, A. M. Abu-Mahfouz, J. Rasheed, and A. AlShammari, "An intelligent anomaly detection approach for accurate and reliable weather forecasting at IoT edges: A case study," *Sensors*, vol. 23, no. 5, p. 2426, 2023.
- [11] S. Ahmad, I. Shakeel, S. Mehruz, and J. Ahmad, "Deep learning models for cloud, edge, fog, and IoT computing paradigms: Survey, recent advances, and future directions," *Computer Science Review*, vol. 49, p. 100568, 2023.
- [12] K. Behravan, N. Farzaneh, M. Jahanshahi, and S. A. H. Seno, "A comprehensive survey on using fog computing in vehicular networks," *Vehicular Communications*, p. 100604, 2023.
- [13] N. Keshari, D. Singh, and A. K. Maurya, "A survey on Vehicular Fog Computing: Current state-of-the-art and future directions," *Vehicular Communications*, vol. 38, p. 100512, 2022.
- [14] V. Hayyolalam, B. Pourghebleh, A. A. P. Kazem, and A. Ghaffari, "Exploring the state-of-the-art service composition approaches in cloud manufacturing systems to enhance upcoming techniques," *The International Journal of Advanced Manufacturing Technology*, vol. 105, no. 1-4, pp. 471-498, 2019.
- [15] V. Hayyolalam, B. Pourghebleh, and A. A. Pourhaji Kazem, "Trust management of services (TMOs): investigating the current mechanisms," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 10, p. e4063, 2020.
- [16] F. Yunlong and L. Jie, "Incentive approaches for cloud computing: challenges and solutions," *Journal of Engineering and Applied Science*, vol. 71, no. 1, p. 51, 2024.
- [17] G. Javadzadeh and A. M. Rahmani, "Fog computing applications in smart cities: A systematic survey," *Wireless Networks*, vol. 26, no. 2, pp. 1433-1457, 2020.
- [18] B. Pourghebleh and V. Hayyolalam, "A comprehensive and systematic review of the load balancing mechanisms in the Internet of Things," *Cluster Computing*, pp. 1-21, 2019.
- [19] B. Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy-efficient data fusion methods in the Internet of Things," *Concurrency and Computation: Practice and Experience*, p. e6959, 2022, doi: <https://doi.org/10.1002/cpe.6959>.
- [20] A. A. Anvigh, Y. Khavan, and B. Pourghebleh, "Transforming Vehicular Networks: How 6G can Revolutionize Intelligent Transportation?," *Science, Engineering and Technology*, vol. 4, no. 1, 2024.
- [21] B. A. Begum and S. V. Nandury, "Data aggregation protocols for WSN and IoT applications—A comprehensive survey," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 2, pp. 651-681, 2023.
- [22] M. A. Rahim, M. A. Rahman, M. M. Rahman, A. T. Asyhari, M. Z. A. Bhuiyan, and D. Ramasamy, "Evolution of IoT-enabled connectivity and applications in automotive industry: A review," *Vehicular Communications*, vol. 27, p. 100285, 2021.
- [23] H. Farag and Č. Stefanović, "Congestion-aware routing in dynamic IoT networks: A reinforcement learning approach," in *2021 IEEE Global Communications Conference (GLOBECOM)*, 2021: IEEE, pp. 1-6.
- [24] S. Bansal and D. Kumar, "Distance-based congestion control mechanism for CoAP in IoT," *IET Communications*, vol. 14, no. 19, pp. 3512-3520, 2020.
- [25] O. L. López et al., "Energy-sustainable IoT connectivity: Vision, technological enablers, challenges, and future directions," *IEEE Open Journal of the Communications Society*, 2023.
- [26] M. Zhang, Y. Li, Y. Ding, and B. Yang, "A Lightweight and Robust Multi-Dimensional Data Aggregation Scheme for IoT," *IEEE Internet of Things Journal*, 2023.
- [27] V. Hayyolalam, B. Pourghebleh, M. R. Chehrehzad, and A. A. Pourhaji Kazem, "Single-objective service composition methods in cloud manufacturing systems: Recent techniques, classification, and future trends," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 5, p. e6698, 2022.
- [28] P. Krivic, M. Kusek, I. Cavrak, and P. Skocir, "Dynamic scheduling of contextually categorised internet of things services in fog computing environment," *Sensors*, vol. 22, no. 2, p. 465, 2022.
- [29] R. O. Aburukba, M. AliKarrar, T. Landolsi, and K. El-Fakih, "Scheduling Internet of Things requests to minimize latency in hybrid Fog-Cloud computing," *Future Generation Computer Systems*, vol. 111, pp. 539-551, 2020.
- [30] M. Ibrahim, Y. Lee, and D.-H. Kim, "DALBFog: Deadline-Aware and Load-Balanced Task Scheduling for the Internet of Things in Fog Computing," *IEEE Systems, Man, and Cybernetics Magazine*, vol. 10, no. 1, pp. 62-71, 2024.
- [31] T. Aladwani, "Scheduling IoT healthcare tasks in fog computing based on their importance," *Procedia Computer Science*, vol. 163, pp. 560-569, 2019.
- [32] E. Khezri, R. O. Yahya, H. Hassanzadeh, M. Mohaidat, S. Ahmadi, and M. Trik, "DLJSF: Data-Locality Aware Job Scheduling IoT tasks in fog-cloud computing environments," *Results in Engineering*, vol. 21, p. 101780, 2024.
- [33] S. Azizi, M. Shojafar, J. Abawajy, and R. Buyya, "Deadline-aware and energy-efficient IoT task scheduling in fog computing systems: A semi-greedy approach," *Journal of network and computer applications*, vol. 201, p. 103333, 2022.
- [34] G. L. Stavrinides and H. D. Karatza, "Scheduling real-time IoT workflows in a fog computing environment utilizing cloud resources with data-aware elasticity," in *2021 Sixth International Conference on Fog and Mobile Edge Computing (FMEC)*, 2021: IEEE, pp. 1-8.
- [35] M. Abdel-Basset, D. El-Shahat, M. Elhoseny, and H. Song, "Energy-aware metaheuristic algorithm for industrial-Internet-of-Things task scheduling problems in fog computing applications," *IEEE Internet of Things Journal*, vol. 8, no. 16, pp. 12638-12649, 2020.
- [36] A. Najafzadeh, A. Salajegheh, A. M. Rahmani, and A. Sahafi, "Privacy-preserving for the internet of things in multi-objective task scheduling in cloud-fog computing using goal programming approach," *Peer-to-Peer Networking and Applications*, vol. 14, pp. 3865-3890, 2021.

- [37] M. Abd Elaziz, L. Abualigah, and I. Attiya, "Advanced optimization technique for scheduling IoT tasks in cloud-fog computing environments," *Future Generation Computer Systems*, vol. 124, pp. 142-154, 2021.
- [38] M. Abd Elaziz, L. Abualigah, R. A. Ibrahim, and I. Attiya, "IoT workflow scheduling using intelligent arithmetic optimization algorithm in fog computing," *Computational intelligence and neuroscience*, vol. 2021, pp. 1-14, 2021.
- [39] J. Luo et al., "Container-based fog computing architecture and energy-balancing scheduling algorithm for energy IoT," *Future Generation Computer Systems*, vol. 97, pp. 50-60, 2019.
- [40] I. Attiya, L. Abualigah, D. Elsadek, S. A. Chelloug, and M. Abd Elaziz, "An intelligent chimp optimizer for scheduling of IoT application tasks in fog computing," *Mathematics*, vol. 10, no. 7, p. 1100, 2022.
- [41] W. Liu, C. Li, A. Zheng, Z. Zheng, Z. Zhang, and Y. Xiao, "Fog computing resource-scheduling strategy in IoT based on artificial bee colony algorithm," *Electronics*, vol. 12, no. 7, p. 1511, 2023.
- [42] M. Kaur, R. Sandhu, and R. Mohana, "A framework for scheduling IoT application jobs on fog computing infrastructure based on QoS parameters," *International Journal of Pervasive Computing and Communications*, vol. 19, no. 3, pp. 364-385, 2023.
- [43] S. Mousavi, S. E. Mood, A. Souri, and M. M. Javidi, "Directed search: a new operator in NSGA-II for task scheduling in IoT based on cloud-fog computing," *IEEE Transactions on Cloud Computing*, 2022.
- [44] I. Attiya, M. Abd Elaziz, and I. Issawi, "An improved hunger game search optimizer based IoT task scheduling in cloud-fog computing," *Internet of Things*, p. 101196, 2024.
- [45] J. Z. Ahmadabadi, S. E. Mood, and A. Souri, "Star-quake: A new operator in multi-objective gravitational search algorithm for task scheduling in IoT based cloud-fog computing system," *IEEE Transactions on Consumer Electronics*, 2023.
- [46] N. A. Alsamurai, O. N. Uçan, and O. F. Khalaf, "Bandwidth-deadline IoT task scheduling in fog-cloud computing environment based on the task bandwidth," *Wireless Personal Communications*, pp. 1-17, 2023.
- [47] A. Azadi and M. Momayez, "Review on Constitutive Model for Simulation of Weak Rock Mass," *Geotechnics*, vol. 4, no. 3, pp. 872-892, 2024, doi: <https://doi.org/10.3390/geotechnics4030045>.
- [48] M. D. Tezerjani, M. Khoshnazar, M. Tangestanizadeh, and Q. Yang, "A Survey on Reinforcement Learning Applications in SLAM," *arXiv preprint arXiv:2408.14518*, 2024, doi: <https://doi.org/10.48550/arXiv.2408.14518>.
- [49] M. Bhatia, S. K. Sood, and S. Kaur, "Quantumized approach of load scheduling in fog computing environment for IoT applications," *Computing*, vol. 102, no. 5, pp. 1097-1115, 2020.
- [50] H. S. Ali, R. R. Rout, P. Parimi, and S. K. Das, "Real-time task scheduling in fog-cloud computing framework for IoT applications: A fuzzy logic based approach," in *2021 International Conference on Communication Systems & NETWORKS (COMSNETS)*, 2021: IEEE, pp. 556-564.
- [51] J. Lim, "Latency-aware task scheduling for IoT applications based on artificial intelligence with partitioning in small-scale fog computing environments," *Sensors*, vol. 22, no. 19, p. 7326, 2022.
- [52] R. O. Aburukba, T. Landolsi, and D. Omer, "A heuristic scheduling approach for fog-cloud computing environment with stationary IoT devices," *Journal of Network and Computer Applications*, vol. 180, p. 102994, 2021.
- [53] N. Potu, S. Bhukya, C. Jatoth, and P. Parvataneni, "Quality-aware energy efficient scheduling model for fog computing comprised IoT network," *Computers & Electrical Engineering*, vol. 97, p. 107603, 2022.
- [54] M. S. Sheikh, R. N. Enam, and R. I. Qureshi, "Machine learning-driven task scheduling with dynamic K-means based clustering algorithm using fuzzy logic in FOG environment," *Frontiers in Computer Science*, vol. 5, p. 1293209, 2023.
- [55] M. Goudarzi, M. A. Rodriguez, M. Sarvi, and R. Buyya, "\$\mu\$-DDRL: A QoS-Aware Distributed Deep Reinforcement Learning Technique for Service Offloading in Fog computing Environments," *IEEE Transactions on Services Computing*, 2023.
- [56] J. Lu et al., "A2C-DRL: Dynamic Scheduling for Stochastic Edge-Cloud Environments Using A2C and Deep Reinforcement Learning," *IEEE Internet of Things Journal*, 2024.
- [57] Z. Wang, M. Goudarzi, M. Gong, and R. Buyya, "Deep Reinforcement Learning-based scheduling for optimizing system load and response time in edge and fog computing environments," *Future Generation Computer Systems*, vol. 152, pp. 55-69, 2024.

A System Dynamics Model of Frozen Fish Supply Chain

Leni Herdiani¹, Maun Jamaludin², Iman Sudirman³, Widjajani⁴, Ismet Rohimat⁵

Department of Industrial Engineering, Universitas Langlangbuana, Bandung, Indonesia^{1,4,5}

Department of Business Administration, Universitas Pasundan, Bandung, Indonesia²

Department of Management Science, Post Graduate, Universitas Pasundan, Bandung, Indonesia³

Abstract—The system dynamics methodology examines the intricate behaviors of complex systems through time, incorporating inventories, transfers, feedback cycles, lookup functions, and temporal delays. In fisheries systems, the interaction between resources and management entities is intricate, with the dynamics of fisheries significantly influencing the formulation of effective policies. Fisheries hold a vital position in Indonesia's economy, contributing to food security, nutrition, and the welfare of fishermen. Under Law Number 7 of 2016, the fisheries sector covers all activities, from resource management to the marketing of marine products. With its rich fishery resources, Indramayu Regency is a major contributor to West Java's fish production. TPI Karangsong, the hub of fishing activities in Indramayu, is a key player in the frozen fish supply chain, relying heavily on cold storage facilities to ensure product quality. Consequently, the system dynamics approach proves valuable in understanding the frozen fish supply chain by modeling the interactions between different variables and evaluating the impact of policies to improve fish quality. The system dynamics model in this study consists of six sub-models: fish at TPI, cold storage, refrigerated cabinets, total revenue, cash, and cold trucks. The simulation results provide policy recommendations to improve the quality of frozen fish at TPI Karangsong, namely the baseline scenario, cold truck scenario, cold truck scenario, truck and cold storage integration scenario, cold storage and fish catch drop integration scenario.

Keywords—Supply chain; system dynamics; frozen fish; six sub-models; simulation; policy scenario

I. INTRODUCTION

As a maritime nation where 70% of its territory is comprised of seas, Indonesia has substantial potential in the tourism and marine sectors, particularly in fisheries. The fisheries sector serves as a crucial contributor to the national economy by generating employment and supporting food security. By Law Number 7 of 2016 [1], the fisheries sector spans activities from pre-production to marketing, integrated within the fisheries business system. The rich diversity of fish resources is a core strength of Indonesian fisheries, with fishing operations governed by Law Number 45 of 2009 [2]. Additionally, the management of marine resources, both renewable and non-renewable, is regulated under Law Number 32 of 2014 [3]. The Marine Affairs and Fisheries Minister Regulation [4] outlines processing and product safety standards to maintain the competitiveness of Indonesian fishery products in global markets. Innovations in fish processing, whether through traditional or modern methods, are essential for adding value to fishery products and enhancing their global market position

[5][6]. In Indramayu Regency, a major fisheries hub in West Java, over 35,000 fishermen produced 551,632.81 tons of fish in 2023, contributing 34.63% to West Java's total fishery production (BPS, 2022).

A company's competitive advantage can be improved through production efficiency, effective distribution, and the timely delivery of products to consumers [7][8]. Supply chain management (SCM) involves key factors such as technology utilization, customer satisfaction, supply chain unification, and inventory management are essential components, while competitive advantage is influenced by factors such as pricing, product quality, market readiness, and sales growth. A company's performance is evaluated through both financial and operational metrics [8]. While a stronger innovation strategy typically enhances operational performance, competitive advantage alone does not fully mediate the link between innovation strategy and overall company performance [9].

SCM is integral to the sustainability of food supply chains, particularly in minimizing food loss and addressing the challenges posed by climate change. The cold chain is essential for preserving the quality of perishable products, such as fish, during their transportation and storage [8]. Despite the significant obstacles faced by cold chain infrastructure in Indonesia, its successful implementation is essential for enhancing the fisheries sector's competitiveness in the global marketplace [10]. Moreover, the establishment of an efficient logistics system is imperative to bolster food security [11].

The complex structure of frozen fish distribution networks requires more than simplistic methods or single-cause solutions. Thus, there is a need for a framework that facilitates an understanding of the multifaceted issues within a systemic context [12] [13]. System dynamics is instrumental in illustrating the interconnections among suppliers, producers, and distribution networks. [14]. This approach underscores the importance of temporal factors in comprehending the overall behavior of a system, demonstrating how such a system can respond to external disturbances or align with the model's objectives (Coyle, 1997 in Wati et al., 2021). Furthermore, system dynamics highlights the influence of policies on system behavior (Richardson & Pugh III, 1997 in Wati et al., 2021) and facilitates the analysis of complex dynamics through the feedback mechanisms among system components [15]. The Causal Loop Diagram (CLD) is extensively used to design complex supply chains, including perishable products like frozen fish [16] [17] while [18] designed a sustainable supply

chain model for the fisheries sector using a system dynamics approach.

TPI Karangsong is a leading fish distribution center in West Java, where the fisheries sector plays a vital role in promoting local economic development. Therefore, adequate cooling facilities are crucial to maintaining the quality of the distributed fish. This study aims to develop a system dynamics model that maps the frozen fish supply chain, expecting the simulation results to propose various policy scenarios that can potentially improve the quality of frozen fish products.

II. RELATED WORKS

A. Supply Chain Management (SCM)

There is integrates various entities collaborating to source raw materials transform into finished goods, deliver them to stores and then consumers [19][20]. Heizer et al. [21] Explain how a distribution network involves suppliers, manufacturers, distributors, and retailers, while Chopra and Meindl [22] emphasize the role of all stakeholders in meeting consumer demands. According to Apriani et al. [23], the interrelationship between goods, finances, and information is crucial. To optimize operations and enhance customer satisfaction, the supply chain mechanism must be effectively implemented through strong relationships with suppliers, efficient production processes, and excellent service [24] [25]. SCM is an integrated approach that efficiently Oversees the movement of products, finances, and insights from upstream to downstream to improve product quality, profitability, and organizational performance [26] [27] [28] [29]. Additionally, SCM enhances competitive advantage by driving efficiency in both production and distribution. [7] [8] [30].

B. System Dynamics (SD)

SD is initially presented by Forrester in the 1950s [31], who constructed a model to demonstrate the way policies influence the soundness of manufacturing systems [32], as well an advanced framework applied to model sophisticated feedback structures [33], making it highly valuable for analyzing interdisciplinary concerns [34]. Besides that, it can also identified using nonlinear relationships and feedback loops between system components, making it suitable for modeling complex socio-economic phenomena [35]. It enables the analysis of interactions among various processes at different levels, providing a holistic understanding of system behavior [15][36]. As part of systems theory, SD helps model nonlinearity through structural modeling incorporating feedback loops and time delays [37]. Key principles in the use of system dynamics include (1) building model structures that define system behavior, (2) integrating soft variables, and (3) interpreting mental models that provide substantial influence [38]. SD offers various benefits, such as visualizing causal links between variables, recognizing the effects of delays, and examining system responses to various scenarios (Zapata et al., 2019). In ecotourism management, Sjaifuddin's [39] model integrates biophysical, social, and economic variables simulated under multiple scenarios, while Utami et al. [40] highlight key response variables such as ecotourism revenue and mangrove rehabilitation. In production planning and control, Karaz et al. [41] emphasize the significance of dynamic interactions in

evaluating the impact of dynamic management (MD) on construction projects. Similarly, Ismail et al. [42] focus on determining optimal Catch capacity to support the sustainability of Malaysian fisheries.

III. METHODOLOGY

The system dynamics model of the frozen fish supply chain in Karangsong is classified as a qualitative-quantitative research study and model-based simulation. This approach combines qualitative and quantitative analyses to understand complex systems and predict dynamic behaviors and feedback structures within the system. The study integrates qualitative and quantitative methods to capture system complexity and simulate time-series data [43] [44].

1) Qualitative research is employed to identify key variables, causal relationships, and behavioral dynamics that influence the frozen fish supply chain. Data are collected through interviews, observations, and discussions with stakeholders, such as fishermen, traders, and cooling facility managers.

2) A quantitative approach is applied during the modeling and simulation phase, using numerical data to construct an analytical model. This method consists of seven stages: problem identification, conceptualization, model formulation, behavior analysis, review, policy examination, and model application [17] [46].

3) The model-based simulation utilizes concepts such as stocks, flows, and feedback loops to describe the movement of materials, information, and policy influences within the system. The results of these simulations are used to evaluate policy scenarios that could enhance fish quality, stabilize supply, and increase income.

Integrating both qualitative and quantitative approaches enables this study to offer more comprehensive insights. By combining the narrative understanding from qualitative analysis with the predictive capabilities of quantitative simulation, the study provides more effective solutions for managing the sustainable frozen fish supply chain at TPI Karangsong. Using Vensim PLE 10.1.3 software, researchers can map variable relationships, analyze system structures, and simulate the dynamic behavior of the supply chain [29] [47].

The stages of building a system dynamics model in frozen fish supply chain system research are (Fig. 1):

1) *Needs analysis*: The initial step aims to identify the problems and needs of the system to be modeled.

2) *Causal loop diagram*: Used to map the cause-and-effect relationships between variables in the system, illustrating feedback loops that influence system dynamics.

3) *Stock and flow diagram*: A more detailed representation of the system, with stocks representing accumulation and flows regulating stock changes. This diagram models the quantitative dynamics of the system.

4) *Simulation models*: The development of system dynamics-based simulation models that enable the analysis of system changes over time.

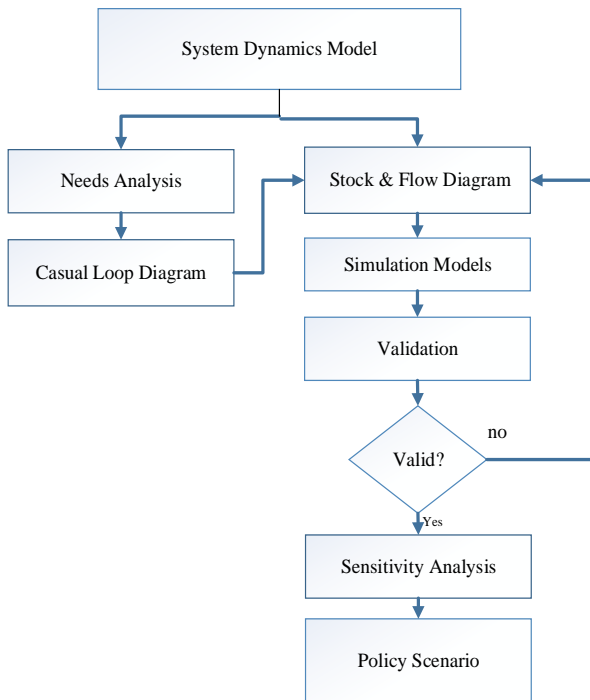


Fig. 1. Research steps.

5) *Validation*: Assessing the model's suitability against historical data or real-world conditions. The Pairwise Pearson Correlation method is used in system dynamics modeling or statistical models to evaluate the relationship between simulation results and actual data. Fish purchase data and fund allocation for fish purchases at one of the *Bakul* from 2021 to 2023 are used to calculate the validation. The formula is:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

This method ensures that the simulation results can represent dynamics that correspond to the actual conditions of the modeled system.

1) *Sensitivity analysis*: After model validation, sensitivity analysis evaluates the impact of parameter changes on the model's results.

2) *Policy scenario*: The final stage involves using the model to test various policy scenarios to determine the best decision for improving system performance.

IV. RESULT AND DISCUSSION

This section is the completion stage in building a system dynamics model on the frozen fish supply chain and provides a selection of several policy scenarios obtained from the simulation results.

A. Results

The system dynamics model is a system modeling tool developed by Jay W. Forrester, which emphasizes feedback (closed loops) to understand the behavior as a whole. This model assumes that the system is always changing, with various

activities influencing each other. Interrelated sub-models are used to achieve certain goals. The main variables in this model include Level (accumulated flow over time), Rate (flow rate), and Auxiliary (helping variables).

1) *Needs analysis*: This stage involves analyzing the needs of stakeholders in PPI Karangsong, such as:

a) *Fishing port*: The main location of fishing industry activities, equipped with safety and other supporting facilities.

b) *Juragan*: The main supplier of frozen fish and ship provider.

c) *TPI Karangsong*: fish auction place.

d) *Fishermen*: The main actors in fishing in the waters.

e) *Bakul*: Fish suppliers to ports, retailers, or consumers.

f) *Consumers*: Actors who utilize the catch.

g) *Government*: Through the Fisheries and Marine Service, regulates and manages fishing activities.

2) *System identification*: System identification is an approach used to comprehensively describe and analyze a system, often employing cause-and-effect diagrams or causal loop diagrams (CLD) to clarify the responses interactions across the variables within the system [48]. This methodology enables the identification of patterns or cycles in the system, facilitating more effective analysis and strategy design.

CLD are constructed based on research data, interviews, and literature reviews to accurately depict causal relationships [49]. In this study, the model is employed to understand of frozen fish supply chain dynamics, beginning with the auction at TPI Karangsong, continuing with storage in cold storage facilities, and culminating in distribution using refrigerated vehicles. This approach allows for analyzing the impact of decisions and related variables, ultimately aiming to improve supply chain efficiency.

The frozen fish supply chain system dynamics model simulation is conducted using Vensim PLE 10.1.13 software, and the CLD for frozen fish distribution is illustrated in Fig. 2.

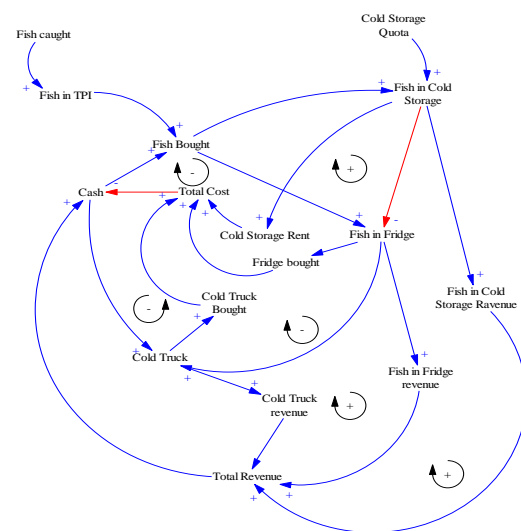


Fig. 2. Causal loop diagram.

Freezing and storage facilities for frozen fish products are crucial components in supporting the cold chain or supply chain system. These facilities help maintain the quality of fish products, enhance their added value, and ensure smooth distribution, price stability, and the availability of fishery products. As a result, the supply chain system can operate more efficiently and effectively.

3) *Boundary model:* The selection of model boundaries is a critical step in modeling, as it defines the scope of analysis relevant to the problem under investigation. These boundaries must encompass all cause-and-effect interactions within the frozen fish supply chain system to ensure a comprehensive representation. Effective decision-making in this supply chain depends on the model's accuracy and completeness, which directly influence the validity of the proposed decisions. The primary objective is to maintain product quality and safety throughout distribution, preventing any degradation that could negatively impact the selling price.

However, this model lacks a detailed analysis of the payback period, and risk analysis is not sufficiently addressed, limiting its ability to fully evaluate the financial feasibility and potential uncertainties associated with policy implementations.

4) *System dynamics simulation flowchart structure:* The simulation of system dynamics modeling results is used to observe behavior patterns and trends within the system, as well as the factors influencing it. However, Causal Loop Diagrams (CLD) have limitations in capturing the structure of stocks and flows in the system. To address this, a Stock & Flow Diagram (SFD) is employed, offering more detail and enabling more accurate simulations of system behavior. SFD incorporates the element of time, allowing for a more detailed analysis of interactions between variables. The frozen fish supply chain system is structured into six sub-models. These six sub-models enhance the system's overall analysis and simulation accuracy, as follows:

a) *Frozen Fish Sub Model at TPI:* Fig. 3 is a frozen fish sub-model at TPI Karangsong.

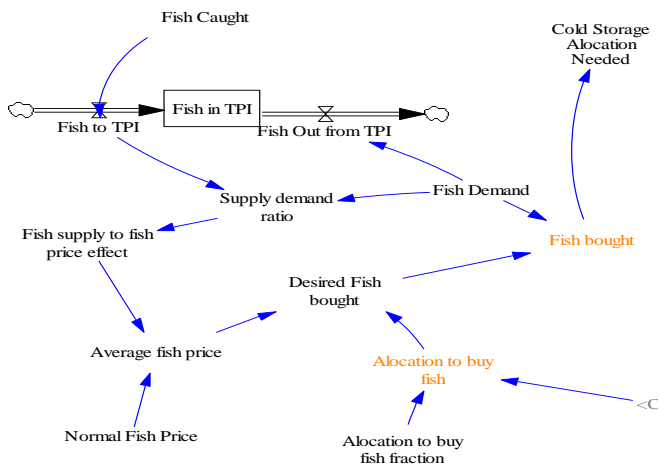


Fig. 3. Frozen fish sub model at TPI.

In this sub-model, frozen fish collected at the TPI represents a stock variable. Various types of frozen fish are auctioned at the TPI, and purchased by the *Bakul*, which reduces the stock of fish at the TPI. The supply-demand ratio is defined as the demand for fish relative to the availability of fish at the TPI. The average fish price is affected by the impact of fish supply on prices relative to the normal price level. The amount of fish purchased by the *Bakul* is determined by the allocation of funds available in the *Bakul's* cash reserves for buying frozen fish. The quantity of fish purchased then dictates the allocation of cold storage required.

b) *Cold storage sub model:* In this sub-model, the fish stored in cold storage is a stock variable. Fig. 4 is a sub-model of Cold storage at TPI Karangsong.

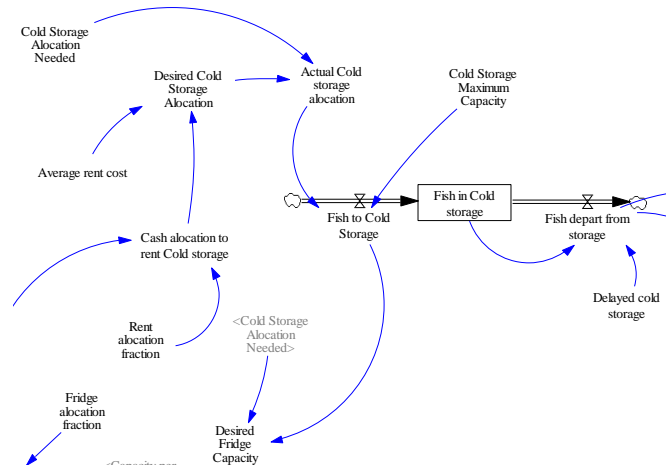


Fig. 4. Cold storage sub-model.

The purchased fish will initially be stored in cold storage. The allocation of fish that can be stored is determined by the cold storage's maximum capacity. The amount of fish stored will affect the cold storage rental cost, which will contribute to the total cost. If the cold storage reaches full capacity, the *Bakul* will store the excess frozen fish in a refrigerator, from which it will later be distributed.

c) *Fish fridge sub model:* In this sub-module, the stock variables are the number of refrigerators (Fish fridge) and fish stored in the Fish fridge. The simulation results of the Fish fridge sub-model are as shown in Fig. 5.

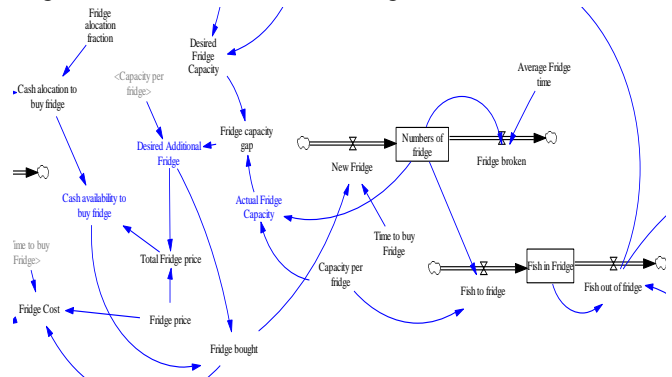


Fig. 5. Fish fridge sub model.

Fish that cannot be stored in cold storage must be placed in a refrigerator, although the quality will not be as well preserved as in cold storage. However, the *Bakul* must allocate funds to purchase the necessary refrigerators. The number of refrigerators required is determined by the amount of fish that exceeds the capacity of the cold storage. The need for additional refrigerators arises when a discrepancy exists between the desired and actual storage capacity to store the surplus fish.

d) *Total revenue sub-model*: Total income in this sub-model is calculated from how many frozen fish are in cold storage multiplied by the fish that come out of cold storage, plus income from fish in the fish fridge and those sent using cold trucks. The simulation results for total revenue are presented in Fig. 6.

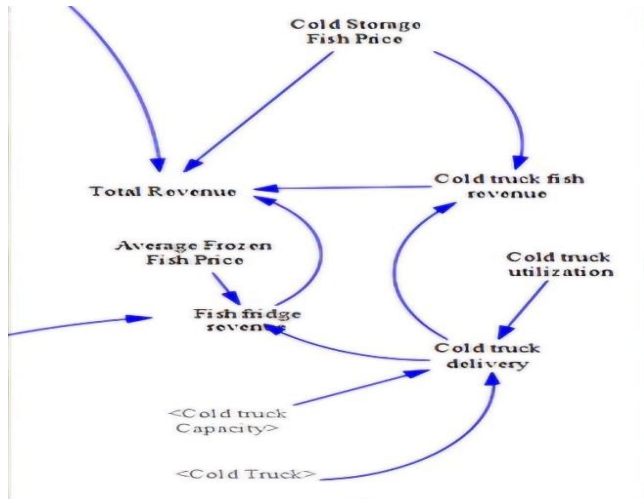


Fig. 6. Total revenue sub-model.

In this sub-model, it is assumed that all frozen fish purchased will be sold out. The assumption is that the price of frozen fish stored in cold storage is higher than that of fish stored in the Fish fridge. Then, frozen fish sent using a cold truck will maintain its quality to the maximum so that the selling price is the same as fish stored in the Fish fridge.

e) *Sub model cash*: Cash is a variable that increases due to income and decreases due to expenses. The total costs calculated in this model include cold storage rental costs, Fish fridge purchases and cold truck purchases (Fig. 7), so cash is income minus expenses.

f) *Sub Model Cold truck*: Cold trucks become stock variables in the sub-model (Fig. 8).

The cold truck is a model used for scenario simulations. In the baseline scenario, it is assumed that the *Bakul* does not own a cold truck, while in the subsequent scenario, the *Bakul* purchases a cold truck based on the amount of fish in the refrigerator and the availability of cash. The purchase of a cold truck depends on the cash allocation available to the *Bakul*. The procurement of the cold truck is determined by its capacity and the amount of fish in the refrigerator that needs to be transported. The price of the cold truck corresponds to the average market price, which varies depending on its specifications and capacity. The cold truck's role is to maintain the safety and quality of frozen fish during distribution, preventing quality degradation

that typically occurs when using non-refrigerated trucks. This ensures that the fish delivered to consumers remains in optimal condition, enhancing consumer satisfaction and loyalty. The overall model is represented by a stock and flow diagram (Fig. 9).

5) *Validation*: Is conducted to verify that a model comprehensively achieves its objectives and accurately represents real system conditions. This process involves evaluating and testing the accuracy of the quantitative framework by comparing the outcomes of the dynamic system simulation against historical data. In this study, model validation was performed using a sample of *Bakul* who passed the normality test from the total population of *Bakul* at TPI Karangsong, Indramayu Regency. The assumption is that the *Bakul* consistently purchase frozen fish. Using data from 2021 to 2023, The following is data on demand for frozen fish in one of the *Bakul*. And normality data is (Fig. 10).

Based on historical data, the forecast results using linear trends for the number of fish purchased and fund allocation are as shown in Fig. 11. Table I shows demand of frozen fish in *Bakul*.

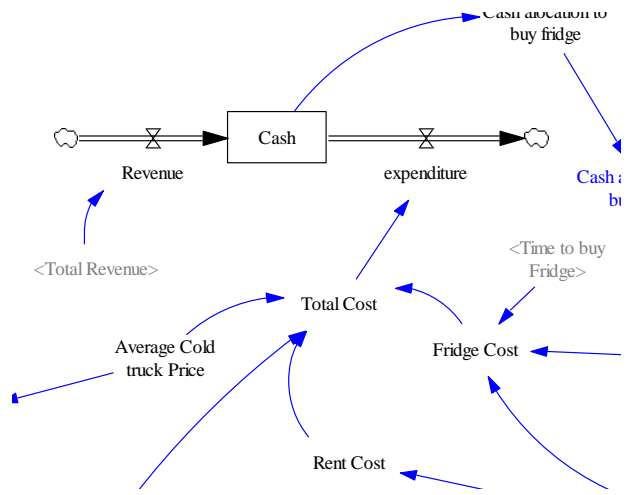


Fig. 7. Cash sub model.

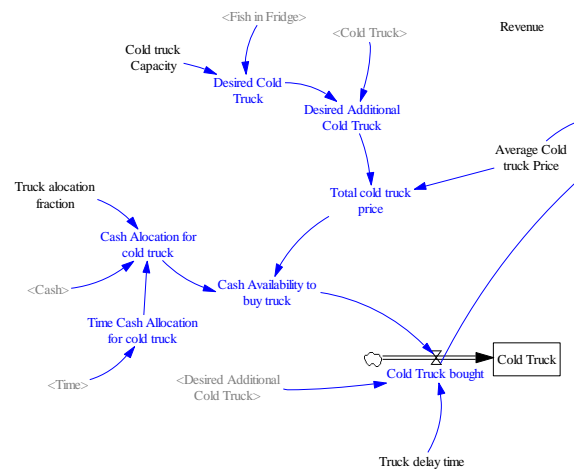


Fig. 8. Cold truck sub-model.

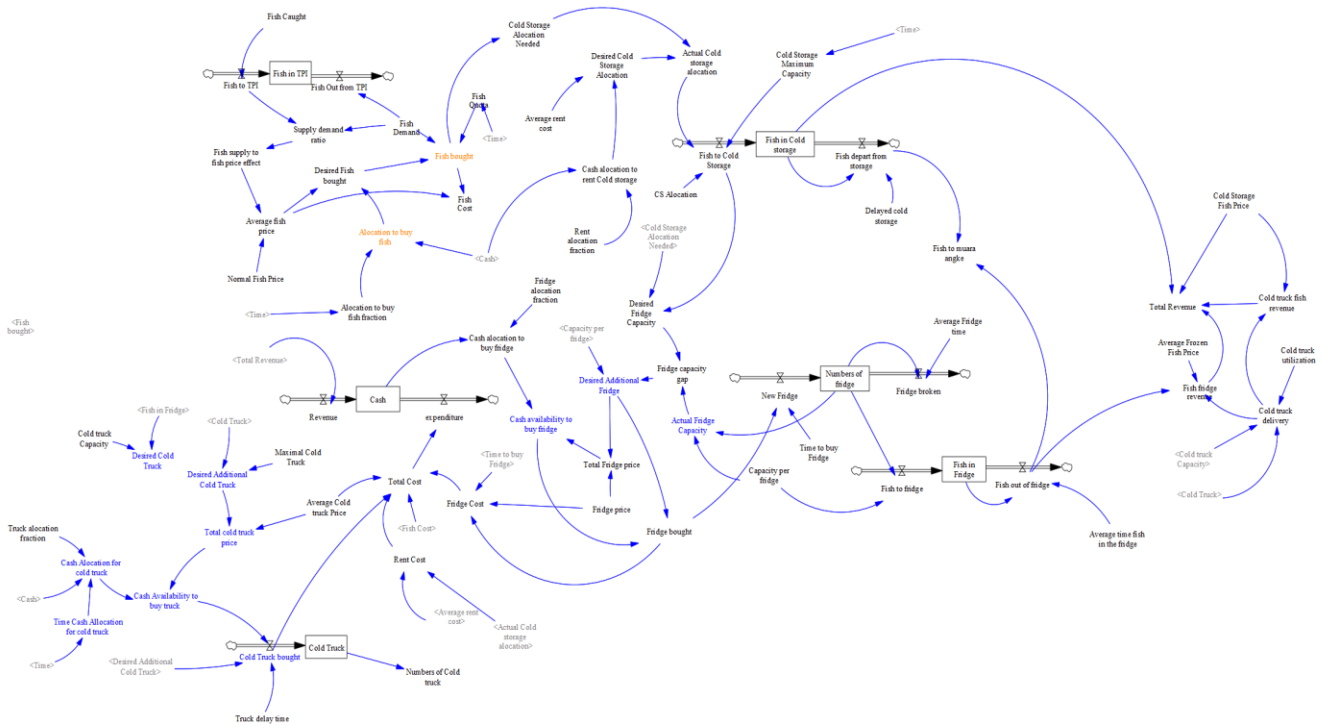


Fig. 9. SFD Frozen fish.

TABLE I. DEMAND OF FROZEN FISH IN BAKUL

Periode	Bought Fish (Kg.)	Forecast (Kg.)	Periode	Allocation to Buy Fish (Rp.)	Forecast (Rp.)
1	2,473	12,444	1	48,931,000	185,975,311
2	8,867	12,706	2	147,607,000	191,417,268
3	20,606	12,968	3	290,844,000	196,859,225
4	26,151	13,230	4	357,867,000	202,301,182
5	8,391	13,492	5	137,282,000	207,743,139
6	1,257	13,754	6	31,132,000	213,185,096
7	6,803	14,016	7	119,456,000	218,627,053
8	8,248	14,278	8	138,812,000	224,069,010
9	16,553	14,540	9	249,369,000	229,510,967
10	10,083	14,802	10	151,900,395	234,952,924
11	19,442	15,064	11	303,967,000	240,394,881
12	4,649	15,326	12	84,935,000	245,836,838
13	4,853	15,588	13	11,025,000	251,278,795
14	9,645	15,850	14	179,837,000	256,720,752
15	33,389	16,112	15	559,813,000	262,162,709
16	29,058	16,374	16	480,636,000	267,604,666
17	22,381	16,636	17	388,136,000	273,046,623
18	14,970	16,898	18	269,764,000	278,488,580
19	18,301	17,160	19	318,884,000	283,930,537
20	29,015	17,422	20	456,545,000	289,372,494
21	20,213	17,684	21	318,644,000	294,814,451
22	28,018	17,946	22	437,558,000	300,256,408
23	31,431	18,208	23	456,345,000	305,698,365
24	13,023	18,470	24	206,548,000	311,140,322
25	29,299	18,732	25	513,981,000	316,582,279

Periode	Bought Fish (Kg.)	Forecast (Kg.)	Periode	Allocation to Buy Fish (Rp.)	Forecast (Rp.)
26	28,172	18,994	26	532,323,000	322,024,236
27	30,358	19,256	27	511,786,000	327,466,193
28	21,777	19,518	28	392,037,000	332,908,150
29	9,240	19,780	29	171,705,000	338,350,107
30	20,146	20,042	30	358,203,000	343,792,064
31	20,327	20,304	31	360,188,000	349,234,021
32	10,767	20,566	32	189,555,000	354,675,978
33	7,267	20,828	33	125,478,000	360,117,935
34	4,768	21,090	34	75,005,000	365,559,892
35		21,344.3	35		371,001,849
36		21,606.1	36		376,443,806
37		21,867.9	37		381,885,763
38		22,129.7	38		387,327,720
39		22,391.4	39		392,769,677
40		22,653.2	40		398,211,634
41		22,915.0	41		403,653,591
42		23,176.8	42		409,095,548
43		23,438.6	43		414,537,505
44		23,700.4	44		419,979,462
45		23,962.2	45		425,421,419
46		24,224.0	46		430,863,376
47		24,485.8	47		436,305,333
48		24,747.5	48		441,747,290
49		25,009.3	49		447,189,247
50		25,271.1	50		452,631,204

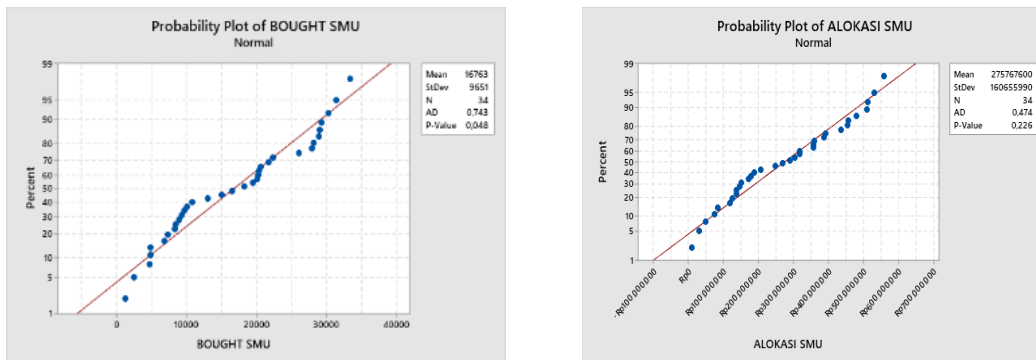


Fig. 10. Normal P Plot test.

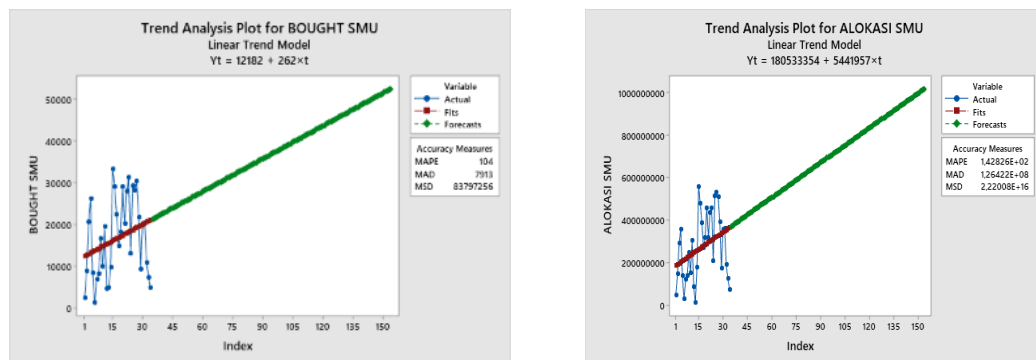


Fig. 11. Forecast results of the number of fish purchased and allocation of funds in one of the *Baku*.

TABLE II. FORECAST ERROR

Method	Value
Bought Fish	
MAPE	104
MAD	7913
MSD	83797256
Allocation to buy Fish	
MAPE	1,42826E+02
MAD	1,26422E+08
MSD	2,22008E+16

From the forecast (Table II) results, Bakul experienced an increase in both the amount and allocation of funds. The results of data processing using MINITAB 19.1.1.0: The following are forecasting errors, namely:

Using the pairwise Pearson correlation method with a 95% confidence and a 5% significance level, the P-value for bought fish was 0.122, and for allocation to buy was 0.051, indicating that the data is valid for predicting future values at *Bakul*.

6) *Sensitivity of system dynamics model*: Sensitivity tests are conducted to measure how sensitive the model is to changes in input parameters or model structure so that it can understand its impact on model output. The results of this sensitivity test are behavioral changes, which are used to analyze the effects of interventions on the model. Sensitivity tests in this study include:

a) *Functional intervention*: Functional intervention entails modifying a specific parameter within the model. This study applies it to the parameter that represents the fish quantity at the TPI. The quantity of fish caught influences the amount of frozen fish purchased by the *Bakul* at the TPI. An increase or decrease in the number of fish will affect the cash allocation available to the *Bakul*.

b) *Structural intervention*: Structural intervention refers to changes made to the model by modifying the relationships that form its structure, aiming to assess their effects on the model's variables. In this study, structural intervention is implemented in the cold truck and cold storage sub-models. The purpose of acquiring a cold truck is to preserve the quality of frozen fish during transportation. In contrast, the increase in cold storage capacity is designed to ensure sufficient storage space.

The (3) three sub-models, namely the fish sub-model at TPI (functional intervention the cold truck sub-model, and the cold storage sub-model (structural intervention) have output sensitivity.

Sensitivity tests help identify various modeling scenarios by adjusting the model's parameters or structure. Thus, sensitivity tests allow researchers to explore various conditions and see how changes in a model's parameters or structure can influence its output.

7) *Analysis of simulation results and policy scenarios*: System dynamics models are essential tools for supporting

practical decision-making, enabling policymakers to model different policy scenarios and assess the impact of each decision. This helps them select the most effective and efficient strategies for addressing complex and dynamic problems. The average price of frozen fish is often influenced by the availability of fish catches from the TPI. As fishing increases, the fish supply becomes more abundant. The harvested fish are then properly stored in cold storage to maintain the quality of frozen fish. Adequate cold storage allows frozen fish sales to occur throughout the year, stabilizing supply and creating more consistent demand and prices. The use of cold storage to preserve frozen fish quality is affected by several factors, including rental costs and storage capacity.

The simulation results produced three scenarios:

a) *Cold truck procurement scenario*: Simulation results in the Cold Truck Procurement Scenario as shown in Fig. 12.

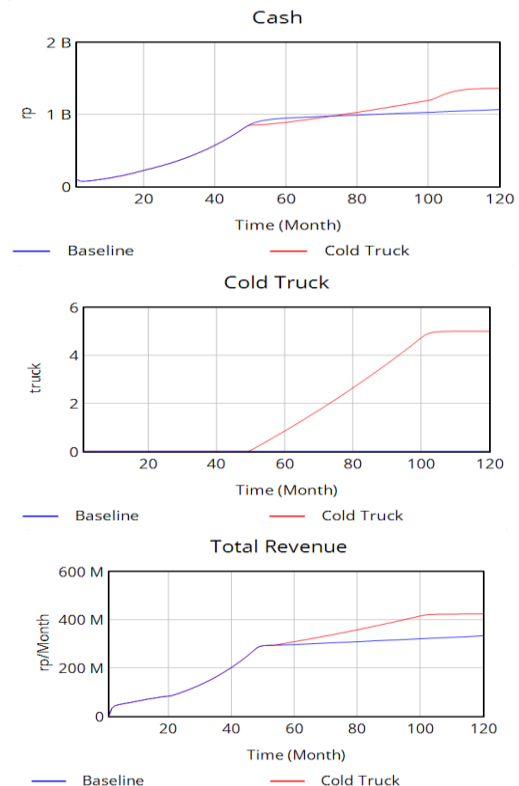


Fig. 12. Cold truck scenario simulation results.

The simulation results (Fig. 12) indicate that although there is a decrease in cash in the 50th month due to the purchase of cold trucks, revenue increases. Investing in cold trucks is essential to maintaining the quality of frozen fish during distribution, preventing price declines. Despite the high initial cost, cold trucks are a strategic investment that supports long-term cash growth through revenue from selling high-quality fish. With five cold trucks, the model positively impacts long-term financial performance, despite the initial outlay. Cold trucks ensure stable fish temperatures during transportation from cold storage to markets or restaurants, mitigating quality issues that arise from unrefrigerated shipments, particularly for travel distances exceeding four hours. Without cold trucks, fish quality deteriorates due to temperature fluctuations, thawing, microorganism growth, and texture damage, all of which reduce the fish's selling value, lead to economic losses, and diminish consumer satisfaction and trust. The procurement of cold trucks helps preserve fish quality, increase selling prices, and improve consumer satisfaction, making it a key solution for the sustainability of TPI Karangsong and the efficiency of distribution.

b) *Cold truck and cold storage integration scenario:* Simulation Results of Cold Truck and Cold Storage Integration Scenario as found in Fig. 13.

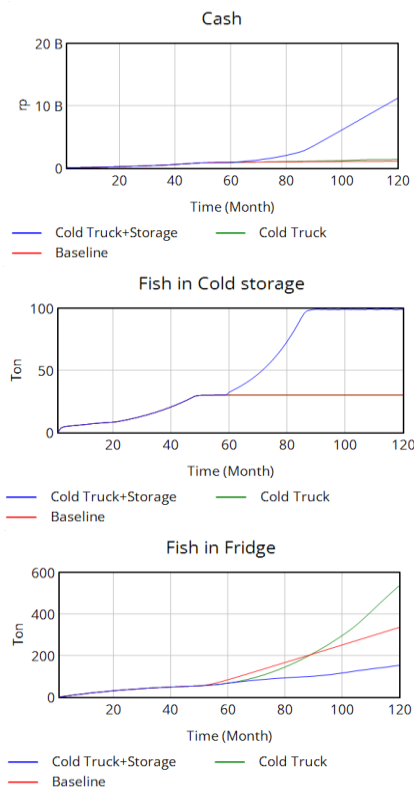


Fig. 13. Simulation results of cold truck and cold storage integration scenario.

This model illustrates how increasing cold storage capacity and using cold trucks for shipping can enhance storage efficiency, product quality, and revenue at TPI Karangsong. Expanding cold storage capacity and utilizing cold trucks are critical strategies for maintaining the quality of frozen fish during storage and distribution. Increased cold storage capacity

allows fish to be kept under optimal conditions for extended periods, minimizing the risk of quality degradation from temperature instability. Cold trucks, with their efficient cooling systems, ensure stable fish temperatures during transportation from cold storage to the final distribution point. The integration of cold storage and cold trucks ensures that frozen fish are stored and distributed under ideal conditions, preserving freshness and quality.

Key impacts of this strategy include improved product quality and safety, which are essential for consumer health and compliance with industry regulations. Revenue increases are driven by the higher selling value of high-quality products, boosting both income and profitability. Operational efficiency is also enhanced due to reduced product spoilage during storage and distribution, lowering damage-related costs. Customer satisfaction improves as well, evidenced by a decrease in return rates and higher satisfaction scores, resulting from consistent product quality. The integration of cold trucks and cold storage not only preserves product quality but also enhances operational efficiency and customer satisfaction, supporting business stability and profitability. Furthermore, this strategy strengthens TPI Karangsong's reputation as a high-quality fish provider, contributing to local fishermen's income and welfare.

c) *Cold truck, cold storage and fish catch drop integration scenario:* Fig. 14 is a simulation results of the integration scenario of cold trucks, cold storage, and fish catch drop.

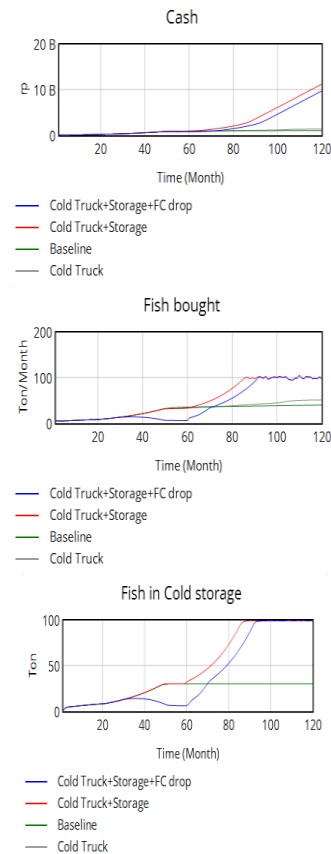


Fig. 14. Simulation results of the integration scenario of Cold Truck, Cold Storage, and Fish Catch Drop.

The simulation results indicate that a decrease in fish catch results in a decrease in the quantity of fish purchased, resulting in lower cold storage usage and reduced revenue. Although cash flow improved compared to the baseline, the increase was not as significant as it would be with stable fish catches. The integration of cold trucks and cold storage is a key strategy for optimizing the storage and distribution of frozen fish, ensuring product quality and safety from the point of catch to the final consumer. By increasing cold storage capacity, fish can be kept under optimal conditions for extended periods, minimizing the risk of quality degradation due to temperature fluctuations. Cold trucks maintain stable fish temperatures during distribution, preserving freshness and quality.

This strategy involves three main components: the fish catch drop, cold storage, and cold trucks. The fish catch drop refers to the initial phase where freshly caught fish are transferred to temperature-controlled facilities to slow spoilage. Cold storage maintains fish at low temperatures to preserve quality over time, reducing microbiological and enzymatic activity. Cold trucks ensure stable temperatures are maintained during transport, safeguarding product integrity.

While this integration enhances operational efficiency and product quality, the decline in fish catch still negatively impacts revenue and shipment volume. However, cash flow improved compared to the baseline, indicating that the strategy mitigates some of the negative effects of reduced fish catches, though overall revenue remains lower. The integration of cold trucks and cold storage is vital for maintaining product quality, improving supply chain efficiency, and enhancing consumer satisfaction, while also helping to reduce the impact of catch fluctuations on TPI Karangsong's revenue and operations.

B. Discussion

Additionally, this research highlights that the system dynamics model for the frozen fish supply chain generates several policy scenarios aimed at enhancing fish quality and improving fishermen's welfare. Policies should promote the effective, efficient, and sustainable utilization of all available natural resources within the country [30]. The policy scenarios are as follows:

1) *Scenario 1: Baseline:* Describes the current state without additional policies. It provides a picture of how variables develop without changes or interventions, helping to identify potential problems and the need for corrective action to achieve desired goals in the future.

2) *Scenario 2: Cold Truck Procurement:* The procurement of cold trucks is intended to maintain the stable temperature of frozen fish during transportation, ensuring the preservation of fish quality. With better product quality, the selling price of fish increases, and consumer satisfaction is ensured. Cold trucks are essential for transporting frozen fish from cold storage to markets, restaurants, or distributors. Trucks with a minimum capacity of 2.9 tons and a temperature range between -20°C and $+10^{\circ}\text{C}$ must be used for deliveries exceeding four hours, while non-refrigerated trucks are prohibited. Although this investment results in an initial cash reduction, it leads to increased revenue and maintains product quality, making it a

strategic decision. Routine maintenance is necessary to extend the lifespan of cold trucks to up to 10 years, ensuring sustainable operations.

The use of cold trucks is crucial for preserving product quality, reducing losses, and ensuring consumer satisfaction and trust.

3) *Scenario 3: Integration of Cold truck and Cold storage:* The integration of cold trucks and cold storage is a strategic policy in the distribution of frozen fish to maintain product quality and improve operational efficiency. This integration allows temperature stability from storage to delivery, reduces the risk of product damage, and increases customer satisfaction. Increasing cold storage capacity requires significant investment and government support. The use of cold trucks in distribution ensures that product temperatures are maintained throughout the supply chain, contributing directly to increased revenue and product sales value. Operational efficiency is improved through better coordination and integrated inventory management, reducing product transfer times and adjusting shipments to market demand. The success of this policy can be measured through increased revenue, reduced return rates, and customer satisfaction. In addition, this integration also reduces product damage, meets food safety standards, and improves business sustainability with better energy efficiency and supply chain management.

4) *Scenario 4: Integration of Cold Truck, Cold Storage, and Fish Catch Drop:* The integration of fish catch reduction, cold storage, and cold trucks aims to maintain optimal temperature and quality of frozen fish throughout the supply chain while enhancing operational efficiency and customer satisfaction. This policy combines all elements of the cold chain [45] to ensure product quality and safety from catch to consumer. The fish catch reduction strategy involves determining the timing and location of catches based on fish population data, along with using monitoring technology to improve catch efficiency. After processing, the fish are stored in cold storage at sub-zero temperatures, focusing on maintaining stable temperatures throughout the facility. Cold trucks are employed during distribution to preserve temperatures during transport, using efficient cooling systems and regular maintenance. Companies must also adhere to food safety standards at every stage. The success of this policy will be evaluated based on product quality, operational efficiency, return rates, and customer satisfaction.

This study offers significant advantages in supporting policy decision-making for the frozen fish supply chain. First, applying a system dynamics approach enables holistic and integrated analysis, capturing interactions among supply chain components, such as the impact of investment in cold trucks and cold storage on fish quality and income. Second, the proposed policy simulations provide insights into various relevant scenarios, including a baseline scenario without intervention for comparison and a scenario involving cold truck and cold storage investments, demonstrating the potential for increased income and selling prices through improved fish quality. Third, this

model facilitates predicting the long-term effects of policies, particularly in scenarios involving fish catch declines, aiding in understanding income stability challenges amid reduced catches. Additionally, the approach supports formulating evidence-based improvement strategies tailored to the needs of local fisheries stakeholders, contributing to sustainable solutions. By integrating simulation results with model-based decision-making, this study enhances its relevance in promoting the sustainability of the fisheries supply chain in TPI Karangsong and surrounding areas.

V. CONCLUSION

The study utilizes a system dynamics approach to simulate the frozen fish supply chain, the simulation results in several policy scenarios. The baseline scenario without intervention shows no significant changes. The cold truck procurement scenario increases income, despite requiring initial investment. Meanwhile, the integration of cold trucks and cold storage helps maintain fish quality and increases selling prices. In the final scenario, which combines cold trucks, cold storage, and a reduction in fish catches, fish quality is preserved, though the income is still affected by the catch decline.

Future research will explore the potential of new technologies, such as IoT for real-time temperature monitoring and blockchain for transparent tracking, to improve the supply chain model's accuracy. Additionally, the sub-models should be supported by a microservice-based IT roadmap, starting with small programs that evolve into larger modules.

REFERENCES

- [1] Undang-Undang Republik Indonesia No. 7 tahun 2016 tentang Perlindungan Dan Pemberdayaan Nelayan, Pembudi Daya Ikan, Dan Petambak Garam.
- [2] Undang-undang Nomor 45 Tahun 2009 tentang Perubahan Atas Undang-Undang No. 31 Tahun 2004 Tentang Perikanan.
- [3] Undang – Undang N0. 32 tahun 2014 tentang Kelautan menggantikan dan mencabut UU 6 tahun 1996 tentang Perairan Indonesia.
- [4] Peraturan Menteri Kelautan Dan Perikanan Republik Indonesia Nomor 9 Tahun 2024 Tentang Pengelolaan Sistem Distribusi Ikan.
- [5] Hutauruk J., Tarigan K., Siahaan S., Sitohang M., & Sihombing D., (2018, November), Hayami method application in the evaluation process of farmers who produce wet and dry corn seeds. In IOP Conference Series: Earth and Environmental Science (Vol. 205, No. 1, p. 012009). IOP Publishing.
- [6] Herdiani, L., Jamaludin, M., Rizkinita, M. A., Sudirman, I., & Rohimat, I. (2024a), The Value Chain Analysis of Coffee Products in the Case of Bandung District, *Almana: Jurnal Manajemen dan Bisnis*, 8(2), 306-317.
- [7] Jamaludin, M. (2021a). Supply Chain Management Strategy In Small And Medium Enterprises (Smes) In The City Of Bandung, West Java. *Journal of Economic Empowerment Strategy (JEES)*, 4(2), 14-24. DOI: <https://doi.org/10.30740/jees.v4i2.121>
- [8] Jamaludin, M. (2021b). The influence of supply chain management on competitive advantage and company performance. *Uncertain Supply Chain Management*, 9(3), 696-704. DOI: 10.5267/j.uscm.2021.4.009
- [9] Jamaludin M., Kania T. N., Segarwati Y., Yuniarti Y., Martiawan R., Rustandi I., Gunawan I. (2023). SCM Practices and Innovation Strategies on Sme Competitive Advantage and Operational Performance, *Operational Research in Engineering Sciences: Theory and Applications* Vol. 6, Issue 4, 2023, pp. a head of print ISSN: 2620-1607 eISSN: 2620-1747 DOI: <https://doi.org/10.31181/oresta/060417>
- [10] Chen Y. H. (2020). Intelligent algorithms for cold chain logistics distribution optimization based on big data cloud computing analysis. *Journal of Cloud Computing*, 9(1), 37.
- [11] Perdana, T., Handayati, Y., Sadeli, A. H., Utomo, D. S., & Hermiatin, F. R. (2020). A Conceptual Model of Smart Supply Chain for Managing Rice Industry. *Mimbar Jurnal Sosial dan Pembangunan*, 36(1), 128-138.
- [12] Bastan M., Delshad Sisi S., Nikoonezhad Z., & Ahmadvand A. M. (2016). Sustainable development analysis of agriculture using system dynamics approach. In *The 34th international conference of the system dynamics society*.
- [13] Reinker M., & Gralla, E. (2018). A system dynamics model of the adoption of improved agricultural inputs in Uganda, with insights for systems approaches to development. *Systems*, 6(3), 31.
- [14] Nuñez Rodriguez, J., Andrade Sosa, H. H., Villarreal Archila, S. M., & Ortiz, A. (2021). System dynamics modeling in additive manufacturing supply chain management. *Processes*, 9(6), 982.
- [15] Guo, Q., Wang, E., Nie, Y., & Shen, J. (2018). Profit or environment? A system dynamic model analysis of waste electrical and electronic equipment management system in China. *Journal of Cleaner Production*, 194, 34-42.
- [16] Elsayah S., McLucas A., & Mazanov J., (2017). An empirical investigation into the learning effects of management flight simulators: A mental models approach. *European Journal of Operational Research*, 259(1), 262-272.
- [17] Hatta M., Mulyani, S., & Umar, N. A. (2020). Dynamic model of fisheries management system and maritime highway program in Makassar Strait. In *IOP Conference Series: Earth and Environmental Science* (Vol. 564, No. 1, p. 012062). IOP Publishing.
- [18] Prabakusuma A. S., Apriani I., Wardono B., Suwondo E., Widodo K. H., & Mareeh H. Y. S. (2020). Designing of Closed-Loop Supply Chain on Dry Land-Based Catfish Aquabusiness in Gunungkidul: A System Dynamics Approach. *ECSoFiM (Economic and Social of Fisheries and Marine Journal)*, 7(2), 212-227.
- [19] Darajat, Y., & Wuryaningtyas, E. (2017). Pengukuran Performansi Perusahaan dengan Menggunakan Metode Supply Chain Operation Reference (SCOR). In *Seminar dan Konferensi Nasional IDEC 2017 8-9 Mei 2017*.
- [20] Anititawati, & dkk. (2016, Mei 29), *Supply Chain Management*, Retrieved from Eprints: <https://Eprints.UMG.ac.id>
- [21] Heizer, J., Render, B. & Munson, C., 2016, *Operations Management: Sustainability and Supply Chain Management*, 12th Edition, Pearson, Boston.
- [22] Chopra, Sunil & Meindl, Peter, 2016, *Supply Chain Management: Strategy, Planning, and Operation* (6th ed). Pearson Education Inc.
- [23] Apriani, H., Erliana, C. I., & Zakaria, M. (2019, October). Analisis supply chain management (SCM) udang vaname di Desa Teupin Pukat Kabupaten Aceh Timur. In *Seminar Nasional Teknik Industri 2019* (Vol. 4, No. 1). Teknik Industri Universitas Malikussaleh.
- [24] Moktadir, M. A., Ali, S. M., Rajesh, R., & Paul, S. K., 2018, *Modeling the Interrelationships Among Barriers To Sustainable Supply Chain Management In Leather Industry*, *Journal of Cleaner Production*, 181, 631-651.
- [25] Maun Jamaludin, N. I. D. N., Fauzi, T. H., Deden Novan Setiawan Nugraha, D., & Latifah Adnani, N. I. D. N. (2022). Service supply chain management in the performance of national logistics agency in national food security. *International Journal of Supply Chain Management*, 9(3), 1080-1084.
- [26] Haudi, H., Rahadjeng, E., Santamoko, R., Putra, R., Purwoko, D., Nurjannah, D., ... & Purwanto, A. (2022). The role of e-marketing and e-CRM on e-loyalty of Indonesian companies during Covid pandemic and digital era. *Uncertain Supply Chain Management*, 10(1), 217-224.
- [27] Yusuf A., & Soediantono, D. (2022). Supply chain management and recommendations for implementation in the defense industry: a literature review. *International Journal of Social and Management Studies*, 3(3), 63-77.
- [28] Vistasusiyanti, V., Kindangen, P., & Palandeng, I. D. (2017). Analisis manajemen rantai pasokan spring bed pada PT. Massindo Sinar Pratama Kota Manado. *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis dan Akuntansi*, 5(2).
- [29] Herdiani L, Jamaludin M, Widjajani, Rohimat I, Putri F.W. (2024b), Mapping of the Frozen Fish Supply Chain System at TPI Karangsong,

- Indramayu, International Journal of Trend in Research and Development, Volume 11(4), ISSN: 2394-9333.
- [30] Jamaludin, M., Fauzi, T., & Nugraha, D. (2021c). A system dynamics approach for analyzing supply chain industry: Evidence from rice industry. *Uncertain Supply Chain Management*, 9 (1), 217-226. DOI: 10.5267/j.uscm.2020.7.007.
- [31] Yu H., Dong S., & Li F. (2019). A System Dynamics Approach to Eco-Industry System Effects and Trends. 28(3), 1469–1482. <https://doi.org/10.15244/pjoes/89508>
- [32] Cesar A., Pinha, H., & Sagawa, J. K. (2020). A system dynamics modeling approach for municipal solid waste management and financial analysis. *Journal of Cleaner Production*, 122350. <https://doi.org/10.1016/j.jclepro.2020.122350>.
- [33] Teng J., Xu C., Wang W., & Wu X. (2018). A system dynamics-based decision-making tool and strategy optimization simulation of green building development in China. *Clean technologies and environmental policy*, 20, 1259-1270.
- [34] Tan, W. J., Yang, C. F., Château, P. A., Lee, M. T., & Chang, Y. C. (2018). Integrated coastal-zone management for sustainable tourism using a decision support system based on system dynamics: A case study of Cijin, Kaohsiung, Taiwan. *Ocean and Coastal Management*, 153(December 2017), 131–139. <https://doi.org/10.1016/j.ocecoaman.2017.12.012>
- [35] Wit M. D., Heun M., & Crookes D. (2018). An overview of salient factors, relationships and values to support integrated energy-economic system dynamics modelling. *Journal of Energy in Southern Africa*, 29(4), 27-36.
- [36] Newman B. M. & Newman P. R., 2020, Theories of Adolescent Development, 2020.
- [37] Daneshzand F., Amin-Naseri, M. R., Asali, M., Elkamel, A., & Fowler, M. (2019). A system dynamics model for optimal allocation of natural gas to various demand sectors. *Computers & Chemical Engineering*, 128, 88-105.
- [38] Saavedra M. R., de O. Fontes, C. H., and Freires, F. G. M. (2018), Sustainable and renewable energy supply chain: A system dynamics overview, *Renewable and Sustainable Energy Reviews*, 82, 247-259.
- [39] Sjaifuddin S. (2020). Sustainable management of freshwater swamp forest as an ecotourism destination in Indonesia: a system dynamics modeling. *Entrepreneurship and Sustainability Issues*, 8(2), 64-85.
- [40] Utami T. N., Fattah M., & Iintyas C. A. (2022). The system dynamic of mangrove ecotourism of “Kampung Blekok” Situbondo East Java Indonesia: economic and ecological dimension. *Environmental Research, Engineering, and Management*, 78(2), 58-72.
- [41] Karaz, M., Teixeira, J. M. C., & Amaral, T. G. D. (2024). Mitigating Making-Do Practices Using the Last Planner System and BIM: A System Dynamic Analysis. *Buildings*, 14(8), 2314.
- [42] Ismail, I., Failler, P., March, A., & Thorpe, A. (2022). A System Dynamics Approach for Improved Management of the Indian Mackerel Fishery in Peninsular Malaysia. *Sustainability*, 14(21), 14190.
- [43] Khoi L.N.D., 2016, *The Conceptual Framework for Fish Quality Management*, International Journal of Science and Research (IJSR), ISSN: 2319-7064 Index Copernicus Value (2016): 79.57 | Impact Factor (2017): 7.296.
- [44] Syukhrani, S., Nurani, T. W., & Haluan, J., (2018). Model Konseptual Pengembangan Perikanan Tongkol Dan Cakalang Yang Didaratkan Di Kota Bengkulu. *Jurnal Teknologi Perikanan dan Kelautan*, 9(1), 1-11.
- [45] Pusporini, P., & Dahdah, S. S., (2020). The conceptual framework of cold chain for fishery products in Indonesia. *Journal of Food Science and Technology*, 8(2), 28-30.
- [46] Wardono B., Yusuf R., Ahmad F., Luhur E. S., & Arthatiani F. Y. (2021, October). Fisheries development model to increase fish consumption in Tabanan, Bali. In *IOP Conference Series: Earth and Environmental Science* (Vol. 860, No. 1, p. 012093). IOP Publishing.
- [47] Marzouk, M., & Fattouh, K. M. (2022). Modeling investment policies effect on environmental indicators in Egyptian construction sector using system dynamics. *Cleaner Engineering and Technology*, 6, 100368.
- [48] Kristianto, A. H., & Nadapdap, J. P. (2021). Dinamika Sistem Ekonomi Sirkular Berbasis Masyarakat Metode Causal Loop Diagram Kota Bengkayang. *Sebatik*, 25(1), 59-67.
- [49] Nugroho, A., Uehara, T., Herwangi, Y., 2019, Interpreting Daly’s Sustainability Criteria for Assessing the Sustainability of Marine Protected Areas: A System Dynamics Approach, *Sustainability* 11, 1-27.

Methodological Review of Social Engineering Policy Model for Digital Marketing

Wenni Syafitri, Zarina Shukur, Umi Asma' Mokhtar, Rossilawati Sulaiman
Center for Cyber Security-Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia, Bangi Selangor, Malaysia

Abstract—Social engineering attacks are recognized as human-based threats and continue to increase, despite studies focusing on prevention methods that do not rely on the human aspect. The impacts of these attacks are felt across various industries and organizations. To solve this issue, a social engineering policy model must be proposed for prevention in industrial settings, particularly emphasizing digital marketing activities, a crucial process in contemporary industries. However, hackers often exploit activities or information in these practices, necessitating an industry-specific policy to prevent these threats in digital marketing. As a result, a comprehensive review was conducted to identify critical methods for developing social engineering policy model. The review uses Bryman's method to determine effective approaches for designing a social engineering policy model tailored for digital marketing. Consequently, this review provided a method for crafting effective social engineering policy, providing valuable insights for enhancing digital marketing security.

Keywords—Digital marketing; social engineering attack prevention; review study; security policy model

I. INTRODUCTION

Social engineering attacks are often performed to expose private information through unauthorized actions [1]. Similarly, NIST describes social engineering attacks as a method of getting trust and confidence from victims [2]. Verizon defines social engineering as exploring human psychology and manipulating sensitive information to exploit people's vulnerability [3].

Various methods have been used to prevent social engineering attacks. For example, CISA recommends practices such as staying vigilant, verifying phone calls and emails, refraining from divulging private or organizational information, ensuring email safety for financial transactions, installing, and managing antivirus software, implementing email filtering and firewall protection, using anti-phishing features in emails and browser plugins, and using multi-factor authentication (CISA). SANS Institute also advocates for security awareness training to mitigate the impact of social engineering attacks (SANS). Moreover, it is crucial to be aware that the prevention methods proposed by CISA and SANS primarily address the technical aspects.

The term "social engineering" is applicable in the context of information security and marketing. Typically, marketing practices can be viewed as a form of social engineering [4]. In marketing, social engineering comprises applied methods for influencing social impact or change, signifying practices used

to influence people's decisions [5]. Consequently, activities in marketing can be termed social engineering [6].

The growing trend of organizations engaging in digital marketing to communicate with external parties makes organizational boundaries unclear. This complicates decisions regarding the information that can be shared with external partners [7]. For example, using social media for digital marketing, including advertising, introduces the risk of unintended information leakage.

According to the Weekly Threat Report dated April 12, 2021, published by NSCS, a data breach compromised 553 million social network users in 106 countries. This breach exposed private information such as IDs, gender, location, and date of birth (NSCS). Unauthorized individuals may exploit this information, manipulating human weaknesses to acquire more confidential data for financial gain. This deceptive tactic is executed through social engineering attack methods. The study conducted by [8] and [9] defined marketing studies based on the respective methods and scopes regarding social engineering. Unlimited exploitation of privacy can cause problems for both customers and companies, leading to a loss of customer trust and revenue when not properly managed by the company.

Cybersecurity policy has a significant influence in fostering cyber governance and cyber resilience in organizations [10], [11]. Some researchers try to build cybersecurity policies with several techniques, such as identifying appropriate security policies to be applied to cyberspace [12], identifying awareness of cybersecurity policies [13], and conducting comparisons between two countries in terms of governance aspects and security policies [14]. However, the policies are to be built by [12], [13], and [14] generalized against various attacks so that prevention against social attacks cannot be used. Therefore, [10] suggested that building policies against cyber-attacks should be specific, such as policies to prevent social engineering attacks.

Very few studies have built social engineering policies. The study in [15] examined recommendations for dealing with organizational members who fall prey to social engineering as an organizational policy issue. The results of this analysis showed that participants did not favor a punitive approach to security failures. Instead, they tended to favor education as a more pragmatic and humane solution. In contrast to [10], they combine the principles of raising security awareness and education in building a social engineering policy. The concept proposed by [10] requires organizational members to read and

learn how social engineering attacks work. In addition, social engineering attack awareness training is required to support the learning activities of organizational members, such as bringing in experts if they have the budget [16].

The studies in [15] and [10] specifically do not focus on building social engineering attack policies. They focus more on policies toward victims of social engineering attacks [15] and policies on how to enhance the social engineering knowledge of each member of the organization [10]. None of the researchers discussed how to build a social engineering policy. Therefore, to answer the gap that researchers have not resolved, this study propose a technical way to design a social engineering policy with a focus on digital marketing. The specialized policy aims to identify and implement relevant policy rules [10].

Developing and consistently updating information security policy is essential for enhancing an organization's security culture [17]. Numerous studies recommend adapting social engineering attack prevention methods at the organizational level by implementing robust social engineering attack policy. The study in [18] proposed the establishment of a strong information security culture through a policy aimed at preventing social engineering attacks. This implies that the crafted information security policy needs to anticipate recent trends in social engineering attacks [19]. Additionally, [15] advised focusing on content rather than just attack policy, taking into account factors such as avoiding harsh sanctions, providing employee education, offering incentives for positive behavior, and determining the appropriate timing for administering punishments.

A good organization should develop and evaluate security policy based on relevant standards and business processes to manage systems, applications, and information effectively. Implementation of social engineering policy by organizations can mitigate vulnerability to hacker attacks, thereby minimizing potential damage [20]. NIST SP 800-152 establishes a standard for Cryptographic Key Management Systems (CKMS), which categorizes security policy into two levels, namely a high-level policy for managing organizational information and a low-level policy consisting of rules to safeguard this information (NIST). CKMS standard is structured with three layers of security policy, including action management, information security, and cryptographic key management system security policy. NIST SP 800-53 Revision 4 defines security policy as a set of standards that support security services.

The search activity showed 31 studies on security policy, each categorized based on the security policy aspects. Each study used a unique method with the primary goal of developing a security policy, aiming to identify an appropriate phase for designing a security policy model to prevent social engineering attacks, particularly in digital marketing.

In the subsequent sections of this paper, studies on the security policy are explained in Section II. Meanwhile, Section III outlines the methodology, Section IV presents the results, Section V presents discussion and Section VI contains the conclusion.

II. RELATED WORK

This section explains some closely related studies, focusing on the topics, challenges, and recommendations relevant to the objective of this current study. The mentioned articles serve as references for designing a security policy model to prevent social engineering attacks in digital marketing. Consequently, the related studies are categorized into four sections as follows:

A. Social Engineering in Digital Marketing

Social engineering has a unique significance in digital marketing, where various marketing activities, such as content marketing, inbound marketing, influencer marketing, social media marketing, creative marketing, innovation marketing, customer journey marketing, conversational marketing, customized lifecycle marketing, performance marketing, and Marketing 4.0 & 5.0, are categorized as forms of social engineering [6].

Social marketing is loosely associated with various marketing methods (as shown in Table I), including non-profit marketing, charity marketing, cause-related marketing, public sector marketing, and government marketing. Additionally, it shares ties with more commercial activities including green marketing or branding for charitable causes, where a company aims to be recognized as a socially responsible entity [8]. When social marketing is used by governments, it does not carry the same negative connotation as totalitarian regimes' propaganda, even though it is a routine government activity [8]. Social engineering is often linked to the desired outcomes of a totalitarian state and is commonly associated with the oppression of citizens in the public perception [8]. Consequently, social engineering is typically viewed as unfavorable, while social marketing is seen as positive.

Digital marketing plays a crucial role in augmenting company income. However, improper use of digital marketing concerning information security can lead to substantial losses for the organization. [8] developed a conceptual model that described the factors related to social engineering and marketing. This model shows the effective and ineffective implementation of social engineering in government activities, such as education, policing, and funding. Similarly, [9] concluded that prioritizing privacy was a viable strategy for increasing hotel revenue. Hotels strategically use privacy measures to provide customers with appropriate services, such as spa and self-service amenities (mini-bars, vending machines, etc.), to enhance perceptions of room comfort and promote repeat visits.

Specific policies are required to address social engineering attacks in organizations, primarily in the aspect of digital marketing. These policies not only help protect sensitive data but also improve the overall effectiveness of marketing campaigns by fostering trust and security among consumers.

In addition, this policy will increase information security awareness as individuals can recognize the nature of cyber threats, how attacks are delivered, their impact on individual safety and business operations, recognize what behaviors can put organizations at risk, and what actions they need to implement when they are attacked [10].

TABLE I. RELATION BETWEEN SOCIAL ENGINEERING AND DIGITAL MARKETING COMPARISON OF A REVIEW STUDY

Term in digital marketing	Digital marketing activity	Related to social engineering activity
Personalization [21]	Marketers utilize browsing history, transaction history, and demographic information to build personalized ads that aim to increase competitive advantage.	Attackers make personalized message content contextually relevant to targets based on collected information [22].
Social Proof [23]	This technique utilizes feedback or ratings from other customers to influence customer decisions in purchasing a product or service.	Attackers create a false sense of security and collective behavior based on fake testimonials or statistics to increase credibility. In addition, the attacker may impersonate a trusted or respected figure in an organization and then say that his or her requests are in line with what everyone else is doing [24].
Scarcity and Urgency [25]	Marketers use the technique of conveying information on commodity unavailability or limited offer of a product in marketing activities.	Attackers using the scarcity principle refer to a persuasion approach using time-based constraints. This technique triggers feelings of anxiety about what will happen if no immediate action is taken.[26].
Storytelling [27]	Marketers create stories around products to create an emotional connection with consumers.	The attacker constructs a personalized story to get the attention of the target, such as a sad story or a victim of a crime or war [28].

Trust is paramount in digital marketing. A well-designed security policy will give customers the impression that the organization takes protecting customers from cybersecurity attacks seriously. The policy must contain an interactional approach to influencing user decisions through recommendations or responses from others as a preventive measure for social engineering attacks [29].

In addition, the social engineering attack policy that has been built is essential to be implemented in an organization. Regular training to raise awareness is an important aspect when implementing social engineering attack policies. Every organization must build social engineering policies carefully to reduce individuals becoming victims of social engineering attacks [15].

The social engineering attack policy should contain technical measures including incident management [30]. Every incident caused by social engineering must be immediately responded to by the organization, either automatically or manually, so as not to have a fatal impact on the management and finances of the organization [31]. Therefore, social engineering attack policies must contain incident management measures that are always up-to-date with social engineering attack patterns [32].

Therefore, building a social engineering attack policy is not only a preventive measure, but digital marketing aspects are an inseparable part of fostering customer trust, ensuring

compliance, and protecting the organization from evolving threats.

Very few researchers have modeled social engineering threats such as [33], [34], and [35]. One of the most famous social engineering attacks is phishing [34]. Threat modeling [33] is to build a phishing model consisting of the factors of threat detection, elaboration, phishing susceptibility, motivation to process, ability to process, and knowledge. Threat detection factors have a significant influence on reducing phishing attacks. Organizations should invest in mitigation measures that support users in detecting phishing threats [36]. In addition, the use of phishing threat modeling also has a significant impact on identifying and securing IoT device vulnerabilities during the initial design phase [34]. Reference [34] utilize detailed information about attacks from each stakeholder. After that, Authors in [34] built a Data Flow Diagram (DFD) to apply threat modeling techniques to identify potential threats in the underlying case using STRIDE threat modeling.

In contrast to [35], they predicted the occurrence of social engineering attacks based on data on the effectiveness of the modalities and principles of persuasion used in Social Engineering Threats (SETs). However, the prevention was carried out by [33], [34], and [35]. It is not fully maximized because it still focuses on technical prevention and evaluation of vulnerabilities that have the opportunity to be breached by social engineering.

Some researchers built threat modeling on social networks, such as Privacy Threat Modeling Language (PTMOL) [37] and DetThr model [38]. The study in [37] built PTMOL to model privacy threats in the Online Social Network (OSN) domain. PTMOL can be incorporated into software development during the design phase of OSNs [37] so that software developers can focus more on privacy protection when building OSNs. The DetThr model built by [38] uses the ThrNet semantic network. The study in [38] claimed that the DetThr Model performed very well in identifying threatening tweet messages. Similar to [33], [34], and [35], [38] and [37] have not been able to build a maximum prevention for social engineering attacks because they still focus on privacy and tweet threats technically.

Brand honesty, consumer trust, and economic security are all severely compromised by sophisticated cyberattacks targeting brand communication networks in today's digitally driven market [39].

Social engineering threats are increasingly dangerous today, but very few focus on prevention, especially in digital marketing. Some social engineering attacks that can be used by attackers in digital marketing are phishing, spear phishing, baiting, pretexting, vishing, smishing, and water-Holing [40]. Phishing utilizes human weaknesses such as time constraints, threats, and user habits to obtain important information by using emails that already contain malicious links. Similar to phishing, spearphishing is more targeted to potential victims; likewise, with vishing, phishing techniques utilize phone calls or the like to get victims, while smishing utilizes Short Message Service (SMS). Baiting utilizes the curiosity of potential victims, such as using a USB Stick with a specific

company logo that contains malicious code. At the same time, water-holing takes advantage of the weakness of an organization's website to insert malicious code to obtain important information when the victim accesses the website. Social engineering attacks can utilize digital marketing activities, as shown in Table II.

TABLE II. SOCIAL ENGINEERING WITH ATTACK VECTOR IN DIGITAL MARKETING

Social engineering attack	Attack vector	Potential digital marketing activity to exploit
Phishing	Email or message feature from Social Networking Sites contains a malicious link with a broader target	Email marketing, ads on social media, promotional landing pages, promotion-based instant messaging, Malicious SEO (Search Engine Optimization), Promotion in Online Groups or Forums, and QR Codes in Offline-Online Campaigns.
Spear Phishing	Email or message feature from Social Networking Sites contains a malicious link with a broader target	Personalized Offer Emails, Targeted Ads on social media, LinkedIn or Other Professional Platforms, Fake Events or Webinars, Targeted E-Commerce or Product Offers, Browsing Record-Based Phishing (Retargeting), Targeted Charity or Donation Campaigns, Surveys or Feedback Forms, Personalized WhatsApp or SMS Messages, and Use of Public Facts about Targets.
Vishing	Robocall or Malicious Call	Exclusive Promotional Offers by Phone, Fake Order Confirmations, Fake Charity or Donation Campaigns, Special Investment or Insurance Offers, Fake Surveys with Prizes, Customer Retention Scams, Lure of Prizes from social media or Online Contests, Confirmation of Changes to Customer Accounts or Data, Retargeting or Remarketing Based Scams, and Webinar or Online Event Registration Based Scams.
Baiting	Malicious USB Sticks or digital assets	Free Digital Content Promotions, Fake Giveaways, Fake Discounts or Coupons, Free E-Book or Educational Material Offers, Pop-Up Ads Offering Gifts or Services, Free Software or Plugin Offers, Fake "Try It Free" Campaigns, "Rare" or Exclusive File Promotions, Rewarded Surveys or Polls, and Rewarded QR Code Offers.
Smishing	The SMS contains a malicious link	Discount or Special Promo Offers, Order Confirmation or Package Delivery, Suspicious Activity Notifications on Accounts, Sweepstakes or Giveaway Winner Announcements, Surveys or Quizzes with Prizes, Service Upgrade Offers, Account Closure Announcements, Fake OTP Codes, Free Service Offers, and Personal Data Update Requests.
Water-Holing	Infected website with malicious code	Display Ads on Popular Sites, Manipulation of Affiliate or Partner Sites, Advertised Local Events or Events, and Attacks on Coupon or Discount Provider Sites.

Significantly few researchers have prevented social engineering attacks on digital marketing, such as preventing water-holing attacks by utilizing Remote Browser Isolation Technology [41] and building a Socio-Cyber-Physical System (SCPS) framework to protect digital marketing assets from the threat of cyber-attacks [39]. The study in [41] performs

isolation and protection of website security access by utilizing Remote Browser Isolation Technology. The concept proposed by [41] helps maintain customer trust, ensure data privacy, and protect brand reputation in the ever-evolving digital marketing landscape. In contrast to [41], [39] combines social behavior analysis, physical network monitoring, and powerful artificial intelligence to build a comprehensive and flexible security system to identify cyber-attacks in advertisements, such as phishing.

There was no one-size-fits-all solution for social engineering attacks, and not all mitigation or defense strategies were equally effective for every target. Therefore, methods for identifying and addressing differences in each target and existing social engineering attack models were needed to develop better prevention strategies [35].

B. Information Security Policy

Existing information security policies proposed by various studies and defined based on correlation topics among the policy and other scope include:

1) *General information security policy*: To build and implement an information security policy, [42] identified ten necessary elements, namely risk assessment, policy construction, implementation, compliance, management, employee support, and three input elements for policy development, including policy guidance standards, drivers, and current literature related to information security policy. Different elements proposed by [42] and [43] design information security policy to prevent insider threats in organizations. The study uses six elements, namely cyber threat intelligence, organizational commitment, security intelligence, information security investment, and misperceptions of information security. Meanwhile, [44] formulated information security policy to assist in organizational regulation and information system security. Information security forms consist of three main elements, such as archives of main directions or policy, standard aspects or policy elements, and activity procedures or technical directions. Additionally, [45] identified determining elements of information security policy, including policy components, objectives, actionable tools, consequences, educational concepts, general concepts, and additional sources. Similar to [45], [46] used seven elements, namely local, global, and integration elements, areas of information security policy expertise, ISP characterization, management, and critical player factors to build an information security model.

The elements proposed by [46], [45], [42], [44], and [43] shared a common relationship. However, a crucial step or procedure is more critical in the development of an information security policy [47], [48], [49]. Action methods, incorporating planning, action, and reflection are used to formulate information security policy in Small and Medium Enterprise (SME). The planning phase comprises identifying SME problems, the action phase consists of formulating an information system security policy, and the reflection phase assesses whether the policy is consistent with organizational

goals and resolves SME problems. This method was different from that of [49], which deployed a framework to realize information security policy for higher education. To safeguard an educational organization information asset from internal, external, intentional, or unintentional threats, an "Information Security Policy" must be developed [50]. The phases for elaborating information security policy, according to [49], include team development as a pre-development process, risk assessment, regulation drafting, validated policy documents, as a process of development, realization, control, and evaluation of policy as a utilization process.

Various studies adopted different perspectives when developing information security policy, such as ontologies and models. [51] used four stages to analyze the ontology, defining policy recognition to determine permission, obligation, or prohibition rules for users, determining action rules for accessible options in the system, identifying compliance factors with policy, and determining actors and accessors or recipients of information security policy practices. In comparison with [51], [52] used a policy-and human-oriented model consisting of three main factors, namely information security policy awareness, security training, and computer and security technology proficiency. Differing from the model proposed by [52], [53] used a formal model to determine security policy in companies by adopting the Chinese wall concept. Similarly, [54] developed a finite automaton policy model for implementation in network security systems. Meanwhile, [55] proposed a model discussing motivation mechanisms for employees to comply with the Information Systems Security Policy (ISSP). While studies presented various models to determine information security policy, no evaluation has been conducted on the proposed models in developing policy.

Numerous studies tried to assess existing information security policy to optimize development. Reference [56] scrutinized information security policy for implementing e-commerce in Saudi Arabia, while [91] explored the phases, context, and content of ISP information security policy development. Adjusting information security policy with business strategies is crucial for successful implementation, as identified by [57], which explored the assimilation of information security policy using a normative, mimetic, and coercive method. Evaluation of information security policy across various sectors, including business environments [58], education [59], and multiple entities in different countries [60], showed considerations such as non-compliance, promotion, management, policy updates, and biased policy areas. Recommendations for information security policy stem from risk analysis, industry guidelines, government legislation, and current organizational policy, yet [60] showed a lack of consistency in applying 'security controls' across policy.

Despite numerous studies on information security policy, several critical aspects require improvement. This includes the adoption of proposed policy applicable to various organizations, countries, and conditions, the evaluation of awareness surrounding developed policy, the lack of evaluation for policy acceptance, and the absence of technical procedures accompanying policy implementation.

2) *Information security policy in social engineering cases:* Social engineering is a cyber-attack with significant repercussions for organizations and individuals and has received limited attention in the aspect of information security policy. Reference [61] evaluated social engineering victims and provided policy recommendations for affected organizations. The study suggested that while education is highly appropriate for individuals affected by social engineering attacks, it may not suffice for members of organizations facing repeated attacks. The need for more suitable sanctions for organizational members was also discussed, advocating for a policy-based method to prevent social engineering attacks.

3) *Information security policy and formal model:* Reference [62] constructed a formal verification for information flow security with dynamic policy in a system. Furthermore, the study developed a general security model incorporating dynamic security policy, underscoring the importance of considering security policy in securing the flow of information.

In summary, various studies established diverse information security policies across different domains and scopes. Generally, investigation on information security policy has been developed based on unique methods and study scopes. However, no prior explorations outlined a social engineering security policy model for digital marketing, which is a gap the current study aims to fill.

C. Risk Assessment in Security Policy

Conducting risk assessments is recommended as a foundational step in developing information security policy. Social engineering arises due to hackers exploiting human vulnerability [63]. Therefore, risk assessments are necessary for organizations and individuals as an initial step in formulating policy to prevent social engineering attacks. [63] identified the nature and key factors of social engineering, conducted a risk assessment using a probabilistic model, and subsequently implemented mitigation strategies based on the assessment results.

Compared with [63], [64] verified the attack vector and prevention of social engineering using a formal model method. Risk assessment models prove valuable in situations characterized by high uncertainty and known facts [64]. The developed risk assessment model aids decision-makers in choosing the optimal solution for mitigating vulnerability and reducing risks.

D. Information Security Policy Evaluation

Previous studies proposed evaluating information security policy. For example, [65] described information security policy measurement using the readability factor. The study used sequential mixed methods to assess the readability of information security policy, although it did not delve into explaining the elements of information security policy. Similar to prior investigations, this study does not evaluate social engineering policy for digital marketing. The current review aims to address the gap in the existing literature by exploring a

security policy model to prevent social engineering attacks in digital marketing, an area that has received limited attention.

The study topic primarily focused on social engineering and information security policy as shown in Table III. It was observed that there was an absence of a study topic specifically addressing social engineering policy for digital marketing. Reference [8] determined the connection between social engineering and marketing, showing that social engineering could be used to reach customers in marketing strategies. However, it could not be directly applied to prevent the impact of social engineering in marketing activities. Addressing these gaps would be interesting to develop a social engineering policy model for digital marketing.

TABLE III. COMPARISON OF THE REVIEW STUDIES

Years	Studies	Social engineering	Information security policy	Marketing/digital marketing
2022	[50]	-	√	-
2020	[66]	√	-	-
2022	[67]	√	-	-
2021	[68]	√	-	-
2021	[69]	√	-	-
2021	[20]	√	-	-
2020	[70]	√	-	-
2020	[71]	√	-	-
2020	[45]	-	√	-
2020	[72]	√	-	-
2022	[73]	√	-	-
2022	[60]	-	√	-
2022	[74]	-	√	-
2020	[57]	-	√	-
2024	This Study	√	√	√

III. METHODOLOGY

In this section, the research outlines the approach employed to formulate a security policy model aimed at preventing social engineering attacks within the domain of digital marketing. This analysis was organized following the findings presented in study [75]. The systematic and structured steps are outlined in three main stages, namely Planning, Conducting, and Reporting and Dissemination, as illustrated in Fig. 1.

In the planning phase, this research methodology was constructed based on the significant contributions found in this source. These references provide the theoretical foundation and key insights that guided the selection and evaluation of relevant articles. Therefore, the structure of this review acknowledges the significant contributions of this literature, which plays a critical role in shaping this study approach.

Moving into the conducting phases, this research involved a systematic and comprehensive exploration of the selected literature. This phase included a meticulous analysis of the identified articles to extract relevant information pertaining to the design and implementation of social engineering policy

models in the realm of digital marketing. Data extraction methods were employed to categorize and synthesize key findings, enabling a detailed understanding of the methodologies, challenges, and outcomes presented in the literature. Additionally, during this phase, this research applied rigorous criteria to ensure the inclusion of studies that align closely with research question and objectives. The conducting phases were crucial in assembling a comprehensive overview of existing insights, paving the way for a nuanced and evidence-based evaluation of social engineering policy models in the context of digital marketing.

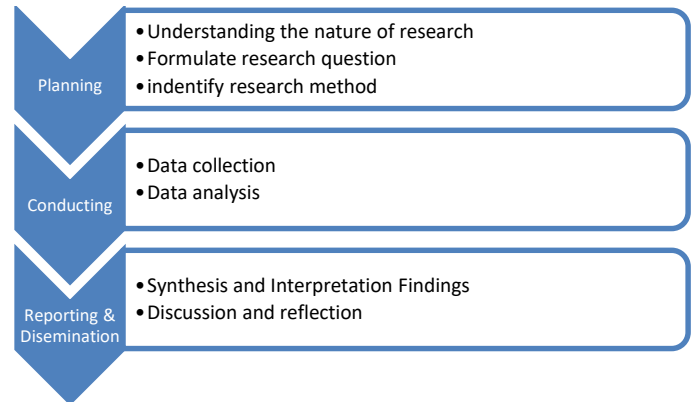


Fig. 1. The review study steps.

As this study transitioned into the reporting and dissemination phases, the focus shifted towards synthesizing the gathered information into a coherent narrative. This involved the compilation of a comprehensive report summarizing the key methodologies, findings, and insights obtained throughout the review. The report was structured to provide clarity and accessibility, ensuring that stakeholders and fellow researchers could easily comprehend the nuances of this study. Moreover, the dissemination aspect involved sharing research outcomes through appropriate channels, such as academic conferences, journals, and other platforms. This phase aimed to contribute to the broader scholarly conversation, fostering knowledge exchange and potentially influencing future research and practical applications in the field of social engineering policy models for digital marketing. Details of the criteria for the search, selection, and assessment process can be seen in Fig. 2.

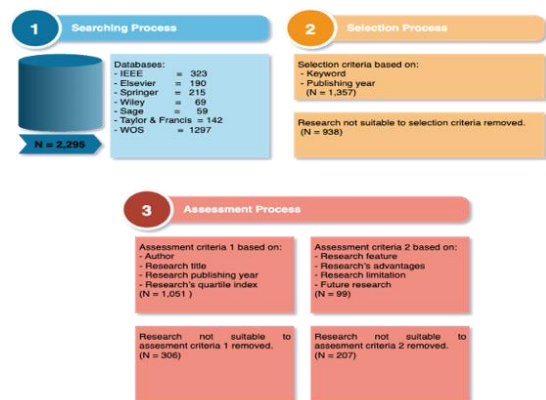


Fig. 2. Searching, Selection, and Assessment process.

A. Searching Process

The search process is a crucial component of this study, encompassing distinct selection and assessment phases. In the selection phase, the research objective was to identify and include studies that met predefined criteria essential to this study focus. Specific criteria, including aspects such as title, year of publication, and article type, were carefully applied to filter and include only studies relevant to this investigation.

To ensure a comprehensive exploration of the existing literature, the search was systematically conducted across nine renowned digital databases. These databases, namely ACM, Elsevier, Emerald, IEEE, Mdpi, Sage, Springer, Wiley, and the WOS index were selected for their prominence in hosting scholarly works related to the research domain. The inclusion of these diverse databases aimed to capture a broad spectrum of literature, enhancing the comprehensiveness and depth of this study.

Through this search process, this research sought to curate a robust collection of studies that would contribute meaningfully to understanding of social engineering policy models in the context of digital marketing. The emphasis on specific criteria and diverse databases ensures a rigorous and inclusive approach, allowing for a thorough examination of the available literature.

Aiming to address specific research question, “Which is the suitable method to design social engineering policy model for digital marketing?”, this study identifies, elucidates, and summarizes suitable methods. Article selection is based on specific criteria with relevant keywords such as social engineering, information security policy, risk assessment, and evaluation methods.

The outcome of the search process revealed 2,295 articles, including 323 articles indexed by IEEE, 190 articles indexed by Elsevier, 215 indexed by Springer, 69 articles indexed by Wiley, 59 articles indexed by Sage, 142 articles indexed by Taylor & Francis, and 1,297 indexed by WOS.

B. Selection Process

To ensure the inclusion of studies aligned with the research objectives, a meticulous selection process was undertaken. Specific criteria were defined, centering on keywords deemed relevant to the investigation, including social engineering, information security policy, risk assessment, and evaluation. These keywords were strategically chosen to encapsulate the essential components of research focus, allowing us to narrow down the pool of potential studies.

This comprehensive selection approach, which spanned multiple databases and publishers, aimed to capture a representative sample of the available literature, enriching the breadth and depth of this study. By adhering to specific criteria and surveying various databases, this study sought to ensure a thorough and well-rounded examination of the relevant studies in the field of social engineering policy models for digital marketing.

Articles were screened using criteria related to keywords and publication year, resulting in the identification of 1,357

articles during this phase. However, 938 articles did not meet the specified selection criteria.

C. Assessment Process

The next step involved assessment process, which was the final stage in this study. The process identified studies that were consistent with the objectives of this current study, using specific criteria and context. Furthermore, a matrix was developed to capture essential information such as author, title, publication year, quartile index, features, advantages, limitations, and potential future research.

This study employed two sets of assessment criteria. The initial criteria screened articles according to authorship, research title, publication year, and quartile index, yielding a total of 1,051 articles meeting these criteria. However, 306 articles did not align with the first set of assessment criteria and were excluded. The second set of criteria evaluated articles based on research features, advantages, limitations, and future research, resulting in the identification of 99 articles meeting these criteria. Additionally, 207 articles were deemed unsuitable based on the second set of criteria and were consequently excluded.

As a result, the selection process yielded a diverse set of results from various databases. Specifically, this study retrieved 2 papers from ACM, 13 from Elsevier, 9 from Emerald, 45 from IEEE, 1 from MDPI, 2 from Sage, 13 from Springer, 3 from Wiley, and 11 from other publishers indexed by the Web of Science (WOS). The results are presented in Table IV, encompassing the number of selected articles from each database.

TABLE IV. RESULT ASSESSMENT PROCESS

Databases	Search String					
	so- cial	Eng- ineer- ing	in- for- mation	Sec- urit- y	Ass- ess- ment	
ACM	2	-	-	-	-	2
Elsevier	9	4	-	-	-	13
Emerald	6	3	-	-	-	9
IEEE	32	11	1	1	-	45
Mdpi	1	-	-	-	-	1
Sage	1	1	-	-	-	2
Springer	7	6	-	-	-	13
Wiley	3	-	-	-	-	3
WOS	6	4	1	-	-	11
	67	29	2	1	-	99

IV. RESULT

This study was categorized into three main areas with four subcategories, as shown in Table V. The primary contributions of the selected study for developing social engineering policy model for digital marketing comprised methods to building risk assessment based on qualitative methods models [63] or quantitative methods [64] and an evaluation method for information security policy before the release to users and stakeholders [76]. Finally, the selected study adopted different methods, aiming to develop information security policy or

prevent social engineering in the field of study, hence contributing to the development of a social engineering policy model for digital marketing.

TABLE V. MAPPING METHODOLOGY REVIEW

Methodology	Keyword	Study
Qualitative	Social Engineering	[7], [8], [18], [19], [29], [67], [68], [69], [70], [71], [72], [73], [77], [78], [79], [80], [81], [82], [83], [84], [85], [86], [87], [88], [89], [90], [91], [92], [93], [94], [95], [96], [97], [98], [99], [100], [101], [102], [103], [104], [105], [106], [107], [108], [109], [110], [111], [112], [113], [114], [115], [116], [117], [118], [119], [120]
	Information Security Policy	[9], [15], [44], [45], [47], [48], [49], [51], [52], [54], [56], [57], [58], [59], [60], [62], [74], [121], [122], [123], [124]
	Risk Assessment	[63]
	Evaluation	-
Quantitative	Social Engineering	[125], [126], [127], [128], [129], [130], [131]
	Information Security Policy	[51], [132], [133]
	Risk Assessment	[64]
	Evaluation	[76]
Mix methods (Qualitative and Quantitative Methods)	Social Engineering	[66]
	Information Security Policy	[42], [43]
	Risk Assessment	-
	Evaluation	-

V. DISCUSSION

The previous studies offered recommendations for designing security policy model to prevent social engineering attacks in digital marketing. Although various methods existed for designing security policy models to prevent social engineering attacks, this current investigation, according to [57], followed three phases, namely identifying security policy requirements, developing security policy model, and validating security policy model.

However, previous ones fell short of comprehensively addressing all three phases of designing security policy models to prevent social engineering attacks. It should be acknowledged that each study tended to concentrate on a specific part. For example, [118] exclusively discussed the user-centric model without covering other phases of designing security policy model for preventing social engineering attacks in digital marketing. Meanwhile, it overlooked the phases of developing security policy models, risk assessment frameworks, formal methods, and evaluation methods. This implies that future works are needed to design a comprehensive approach to model security policies to prevent social engineering attacks in digital marketing and to implement them in various organizations to see the effectiveness of the policies.

This study experienced several limitations, including acquiring methods to evaluate social engineering attack policy, particularly before and after policy implementation. Several studies solely concentrated on building information security policies to counter information security threats. Additionally, some social engineering explorations were limited to evaluating perceptions about information security policy and using formal models to identify essential processes in information security policy. While readability methods served as an alternative for assessing policy effectiveness, challenges existed in determining how to assess better readability in an organizational context. The aspect of digital marketing for social engineering is relatively new, yet numerous social engineering activities inevitably occur, particularly in information gathering. Even though the term "social engineering" in marketing carries a positive connotation, when the activity deviates, it could cause a threat to digital marketing activities without realization.

VI. CONCLUSION

This review successfully identifies methodologies that can be used to build Social Engineering Policy Models, especially Digital Marketing. This review categorizes quality articles into three methodologies, namely Qualitative, Quantitative, and mixed methods. Each article is categorized into social engineering, information security policy, risk assessment, and evaluation. Based on the review’s findings, many researchers only build policies to prevent social engineering attacks but do not validate the policies used, especially researchers who use mixed methods. Therefore, one of the directions of future research development is to ensure that every social engineering attack policy built must be validated or assessed. Validation and assessment can use formal methods and risk assessment techniques.

ACKNOWLEDGMENT

The authors also would like to show gratitude to Faculty of Information Science and Technology, and Universiti Kebangsaan Malaysia for their support.

REFERENCES

- [1] European Union Agency For Network And Information Security, "Definition of Cybersecurity-Gaps and overlaps in standardisation," 2015. doi: 10.2824/4069.
- [2] P. A. Grassi, M. E. Garcia, and J. L. Fenton, "Digital identity guidelines: revision 3," Gaithersburg, MD, Jun. 2017. doi: 10.6028/NIST.SP.800-63-3.
- [3] Verizon, "DBIR 2023 Data Breach Investigations Report," 2023.
- [4] Chen, Chiang, and Storey, "Business Intelligence and Analytics: From Big Data to Big Impact," MIS Quarterly, vol. 36, no. 4, p. 1165, 2012, doi: 10.2307/41703503.
- [5] A. Kennedy and A. Parsons, "Macro-social marketing and social engineering: a systems approach," J Soc Mark, vol. 2, no. 1, pp. 37–51, Feb. 2012, doi: 10.1108/20426761211203247.
- [6] J. Lies, "Marketing Intelligence and Big Data: Digital Marketing Techniques on their Way to Becoming Social Engineering Techniques in Marketing," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 5, pp. 134–144, 2019, doi: 10.9781/ijimai.2019.05.002.
- [7] K. Krombholz, H. Hobel, M. Huber, and E. Weippl, "Advanced social engineering attacks," Journal of Information Security and Applications, vol. 22, pp. 113–122, Jun. 2015, doi: 10.1016/j.jisa.2014.09.005.

- [8] A. M. Kennedy and A. Parsons, "Social engineering and social marketing: Why is one 'good' and the other 'bad'?", *J Soc Mark*, vol. 4, no. 3, pp. 198–209, Sep. 2014, doi: 10.1108/JSOCM-01-2014-0006.
- [9] M. J. Magalhães, S. T. de Magalhães, K. Revett, and H. Jahankhani, "A review on privacy issues in hotels: A contribution to the definition of information security policies and marketing strategies," in *Communications in Computer and Information Science*, Springer Verlag, 2016, pp. 205–217. doi: 10.1007/978-3-319-51064-4_17.
- [10] E. Stavrou, A. Piki, and P. Varnava, "Merging Policy and Practice: Crafting Effective Social Engineering Awareness-Raising Policies," in *International Conference on Information Systems Security and Privacy*, Science and Technology Publications, Lda, 2024, pp. 179–186. doi: 10.5220/0012410300003648.
- [11] D. Singh, "Civil Servants Awareness Guideline Towards Computer Security Policy: A Case Study at the Manpower Department, Ministry of Human Resources," *Asia-Pacific Journal of Information Technology and Multimedia*, vol. 10, no. 01, pp. 86–99, Jun. 2021, doi: 10.17576/apjitm-2021-1001-08.
- [12] J. O. Oyelami and A. M. Kassim, "Cyber Security Defence Policies: A Proposed Guidelines for Organisations Cyber Security Practices," 2020. [Online]. Available: www.ijacsa.thesai.org
- [13] M. A. Pitchan and S. Z. Omar, "Cyber security policy: Review on netizen awareness and laws," *Jurnal Komunikasi: Malaysian Journal of Communication*, vol. 35, no. 1, pp. 103–119, 2019, doi: 10.17576/JKMJC-2019-3501-08.
- [14] N. Rawindaran et al., "Enhancing Cyber Security Governance and Policy for SMEs in Industry 5.0: A Comparative Study between Saudi Arabia and the United Kingdom," *Digital*, vol. 3, no. 3, pp. 200–231, Sep. 2023, doi: 10.3390/digital3030014.
- [15] K. F. Steinmetz and T. J. Holt, "Falling for Social Engineering: A Qualitative Analysis of Social Engineering Policy Recommendations," *Soc Sci Comput Rev*, vol. 41, no. 2, pp. 592–607, Apr. 2023, doi: 10.1177/08944393221117501.
- [16] N. A. Azam, A. Geogiana Buja, R. Ahmad, S. F. A. Latip, and N. M. Sahri, "An Analysis of the Deployment of Synergistic Cyber Security Awareness Model for the Elderly (SCSAM-Elderly) in Malaysia," *Akademika*, vol. 94, no. 3, pp. 90–107, 2024, doi: 10.17576/akad-2024-9403-06.
- [17] H. A. Aldawood and G. Skinner, "A Critical Appraisal of Contemporary Cyber Security Social Engineering Solutions: Measures, Policies, Tools and Applications," in *2018 26th International Conference on Systems Engineering (ICSEng)*, IEEE, Dec. 2018, pp. 1–6. doi: 10.1109/ICSENG.2018.8638166.
- [18] L. Pharris and B. Perez-Mira, "Preventing social engineering: a phenomenological inquiry," *Information and Computer Security*, vol. 31, no. 1, pp. 1–31, Feb. 2023, doi: 10.1108/ICS-09-2021-0137.
- [19] K. Matyokurehwa, N. Rudhumbu, C. Gombiro, and C. Chipfumbu-Kangara, "Enhanced social engineering framework mitigating against social engineering attacks in higher education," *SECURITY AND PRIVACY*, vol. 5, no. 5, Sep. 2022, doi: 10.1002/spy2.237.
- [20] B. Kotkova and M. Hromada, "Cyber Security and Social Engineering," in *Proceedings - 25th International Conference on Circuits, Systems, Communications and Computers, CSCC 2021*, Institute of Electrical and Electronics Engineers Inc., 2021, pp. 134–140. doi: 10.1109/CSCC53858.2021.00031.
- [21] N. Cavdar Aksoy, E. Tumer Kabadayi, C. Yilmaz, and A. Kocak Alan, "A typology of personalisation practices in marketing in the digital age," *Journal of Marketing Management*, vol. 37, no. 11–12, pp. 1091–1122, 2021, doi: 10.1080/0267257X.2020.1866647.
- [22] M. Schmitt and I. Flechais, "Digital deception: generative artificial intelligence in social engineering and phishing," *Artif Intell Rev*, vol. 57, no. 12, p. 324, Oct. 2024, doi: 10.1007/s10462-024-10973-2.
- [23] K. Roethke, J. Klumpe, M. Adam, and A. Benlian, "Social influence tactics in e-commerce onboarding: The role of social proof and reciprocity in affecting user registrations," *Decis Support Syst*, vol. 131, Apr. 2020, doi: 10.1016/j.dss.2020.113268.
- [24] A. Mollazehi, I. Abuelezz, M. Barhamgi, K. M. Khan, and R. Ali, "Do Cialdini's Persuasion Principles Still Influence Trust and Risk-Taking When Social Engineering is Knowingly Possible?," in *Lecture Notes in Business Information Processing*, vol. 513, Springer Science and Business Media Deutschland GmbH, 2024, pp. 273–288. doi: 10.1007/978-3-031-59465-6_17.
- [25] V. Pavlidou, J. Otterbacher, and S. Kleanthous, "User Perception of Algorithmic Digital Marketing in Conditions of Scarcity," in *Lecture Notes in Business Information Processing*, Springer Science and Business Media Deutschland GmbH, 2022, pp. 319–332. doi: 10.1007/978-3-030-95947-0_22.
- [26] M. A. Siddiqi, W. Pak, and M. A. Siddiqi, "A Study on the Psychology of Social Engineering-Based Cyberattacks and Existing Countermeasures," Jun. 01, 2022, MDPI. doi: 10.3390/app12126042.
- [27] E. Sung, D. I. D. Han, Y. K. Choi, B. Gillespie, A. Couperus, and M. Koppert, "Augmented digital human vs. human agents in storytelling marketing: Exploratory electroencephalography and experimental studies," *Psychol Mark*, vol. 40, no. 11, pp. 2428–2446, Nov. 2023, doi: 10.1002/mar.21898.
- [28] E. Dincelli and I. S. Chengalur-Smith, "Choose your own training adventure: designing a gamified SETA artefact for improving information security and privacy through interactive storytelling," *European Journal of Information Systems*, vol. 29, no. 6, pp. 669–687, 2020, doi: 10.1080/0960085X.2020.1797546.
- [29] Y. Kano and T. Nakajima, "Trust factors of social engineering attacks on social networking services," in *LifeTech 2021 - 2021 IEEE 3rd Global Conference on Life Sciences and Technologies*, Institute of Electrical and Electronics Engineers Inc., Mar. 2021, pp. 25–28. doi: 10.1109/LifeTech52111.2021.9391929.
- [30] S. S. I. Rahim, M. I. Mohd Huda, S. Sa'ad, and R. Moorthy, "Cyber Security Crisis/Threat: Analysis of Malaysia National Security Council (NSC) Involvement Through the Perceptions of Government, Private and People Based on the 3P Model," *e-Bangi Journal of Social Science and Humanities*, vol. 21, no. 2, May 2024, doi: 10.17576/ebangi.2024.2102.17.
- [31] S. Mamat, W. A. Wan Mahmud, and A. A. Azlan, "Trend Berkomunikasi dan Transaksi Dalam Talian: Selamatkah Data Peribadi Belia Malaysia?," *e-Bangi Journal of Social Science and Humanities*, vol. 20, no. 3, Aug. 2023, doi: 10.17576/ebangi.2023.2003.08.
- [32] S. Waelchli and Y. Walter, "Reducing the risk of social engineering attacks using SOAR measures in a real world environment: A case study," *Comput Secur*, vol. 148, Jan. 2025, doi: 10.1016/j.cose.2024.104137.
- [33] P. M. W. Musuva, K. W. Getao, and C. K. Chepken, "A new approach to modelling the effects of cognitive processing and threat detection on phishing susceptibility," *Comput Human Behav*, vol. 94, pp. 154–175, May 2019, doi: 10.1016/j.chb.2018.12.036.
- [34] S. G. Abbas et al., "Identifying and mitigating phishing attack threats in IoT use cases using a threat modelling approach," *Sensors*, vol. 21, no. 14, Jul. 2021, doi: 10.3390/s21144816.
- [35] M. Aijaz and M. Nazir, "Modelling and analysis of social engineering threats using the attack tree and the Markov model," *International Journal of Information Technology (Singapore)*, vol. 16, no. 2, pp. 1231–1238, Feb. 2024, doi: 10.1007/s41870-023-01540-z.
- [36] A. H. Shakiba, G. Ghadiri, and H. S. Karaki, "Iran's Legislative Policy in dealing with Fraud During COVID-19 Pandemic," *JURNAL UNDANG-UNDANG DAN MASYARAKAT*, vol. 33, pp. 103–118, Dec. 2023, doi: 10.17576/juum-2023-33-09.
- [37] A. Rodrigues, M. L. B. Villela, and E. L. Feitosa, "Privacy Threat Modeling Language," *IEEE Access*, vol. 11, pp. 24448–24471, 2023, doi: 10.1109/ACCESS.2023.3255548.
- [38] F. Fkih and G. Al-Turaif, "Threat Modelling and Detection Using Semantic Network for Improving Social Media Safety," *International Journal of Computer Network and Information Security*, vol. 15, no. 1, pp. 39–53, Feb. 2023, doi: 10.5815/ijcnis.2023.01.04.
- [39] S. Yang and H. Long, "Socio Cyber-Physical System for Cyber-Attack Detection in Brand Marketing Communication Network," *Wirel Pers Commun*, Jun. 2024, doi: 10.1007/s11277-024-11261-6.
- [40] Z. Wang, H. Zhu, P. Liu, and L. Sun, "Social engineering in cybersecurity: a domain ontology and knowledge graph application examples," *Cybersecurity*, vol. 4, no. 1, Dec. 2021, doi: 10.1186/s42400-021-00094-6.

- [41] J. Hu, H. Wang, and Y. Liu, "Strengthening Digital Marketing Security Website Threat Isolation and Protection Using Remote Browser Isolation Technology," *Comput Aided Des Appl*, pp. 56–74, Nov. 2023, doi: 10.14733/cadaps.2024.s4.56-74.
- [42] S. V. Flowerday and T. Tuyikeze, "Information security policy development and implementation: The what, how and who," *Comput Secur*, vol. 61, pp. 169–183, Aug. 2016, doi: 10.1016/j.cose.2016.06.002.
- [43] H. Stewart, "A systematic framework to explore the determinants of information security policy development and outcomes," *Information and Computer Security*, vol. 30, no. 4, pp. 490–516, Oct. 2022, doi: 10.1108/ICS-06-2021-0076.
- [44] E. L. G. Fontes and A. J. Balloni, "Information security policy: The regulatory basis for the protection of information systems," in *Laboratory Management Information Systems: Current Requirements and Future Perspectives*, IGI Global, 2014, pp. 95–117. doi: 10.4018/978-1-4666-6320-6.ch006.
- [45] E. Rostami, F. Karlsson, and S. Gao, "Requirements for computerized tools to design information security policies," *Comput Secur*, vol. 99, Dec. 2020, doi: 10.1016/j.cose.2020.102063.
- [46] A. Klaic and M. Golub, "Conceptual information modelling within the contemporary information security policies," 2013.
- [47] I. Lopes and P. Oliveira, "Implementation of information systems security policies: A survey in small and medium sized enterprises," in *Advances in Intelligent Systems and Computing*, Springer Verlag, 2015, pp. 459–468. doi: 10.1007/978-3-319-16486-1_45.
- [48] I. Lopes and P. Oliveira, "Applying action research in the formulation of information security policies," in *Advances in Intelligent Systems and Computing*, Springer Verlag, 2015, pp. 513–522. doi: 10.1007/978-3-319-16486-1_50.
- [49] W. B. W. Ismail, S. Widyarto, R. A. T. R. Ahmad, and K. A. Ghani, "A generic framework for information security policy development," in *2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, IEEE, Sep. 2017, pp. 1–6. doi: 10.1109/EECSI.2017.8239132.
- [50] P. Petrov, I. Kuyumdzhev, R. Malkawi, G. Dimitrov, and J. Jordanov, "Digitalization of Educational Services with Regard to Policy for Information Security," *TEM Journal*, vol. 11, no. 3, pp. 1093–1102, Aug. 2022, doi: 10.18421/TEM113-14.
- [51] D. Mandal and C. Mazumdar, "Towards an ontology for enterprise level information security policy analysis," in *ICISSP 2021 - Proceedings of the 7th International Conference on Information Systems Security and Privacy*, SciTePress, 2021, pp. 492–499. doi: 10.5220/0010248004920499.
- [52] K. E. H. A. Alhosani, S. K. A. Khalid, N. A. Samsudin, S. Jamel, and K. M. bin Mohamad, "A policy driven, human oriented information security model: a case study in UAE banking sector," in *2019 IEEE Conference on Application, Information and Network Security (AINS)*, IEEE, Nov. 2019, pp. 12–17. doi: 10.1109/AINS47559.2019.8968705.
- [53] T. Y. T. Y. Lin, "Chinese wall security policies information flows in business cloud," in *Proceedings - 2015 IEEE International Conference on Big Data*, IEEE Big Data 2015, Institute of Electrical and Electronics Engineers Inc., Dec. 2015, pp. 1603–1607. doi: 10.1109/BigData.2015.7363927.
- [54] D. Chernyavskiy and N. Miloslavskaya, "An Approach to Information Security Policy Modeling for Enterprise Networks," in *Communications and Multimedia Security*, B. De Decker and A. Zúquete, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 118–127.
- [55] H. P. Shih, X. Guo, K. H. Lai, and T. C. E. Cheng, "Taking promotion and prevention mechanisms matter for information systems security policy in Chinese SMEs," in *Proceedings of 2016 International Conference on Information Management, ICIM 2016*, Institute of Electrical and Electronics Engineers Inc., May 2016, pp. 110–115. doi: 10.1109/INFOMAN.2016.7477543.
- [56] K. Thakur, M. L. Ali, K. Gai, and M. Qiu, "Information Security Policy for E-Commerce in Saudi Arabia," in *2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity)*, IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), IEEE, Apr. 2016, pp. 187–190. doi: 10.1109/BigDataSecurity-HPSC-IDS.2016.14.
- [57] H. Paananen, M. Lapke, and M. Siponen, "State of the art in information security policy development," Jan. 01, 2020, Elsevier Ltd. doi: 10.1016/j.cose.2019.101608.
- [58] M. Alotaibi, S. Furnell, and N. Clarke, "Information security policies: A review of challenges and influencing factors," in *2016 11th International Conference for Internet Technology and Secured Transactions, ICITST 2016*, Institute of Electrical and Electronics Engineers Inc., Feb. 2017, pp. 352–358. doi: 10.1109/ICITST.2016.7856729.
- [59] L. G. Ording, S. Gao, and W. Chen, "The influence of inputs in the information security policy development: an institutional perspective," *Transforming Government: People, Process and Policy*, vol. 16, no. 4, pp. 418–435, Oct. 2022, doi: 10.1108/TG-03-2022-0030.
- [60] B. Ngoqo and K. Njenga, "The state of e-Government security in South Africa: Analysing the national information security policy," in *Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST*, Springer Verlag, 2018, pp. 29–46. doi: 10.1007/978-3-319-98827-6_3.
- [61] K. F. Steinmetz, T. J. Holt, and C. G. Brewer, "Developing and implementing social engineering-prevention policies: a qualitative study," *Security Journal*, Jun. 2023, doi: 10.1057/s41284-023-00385-2.
- [62] J. Sun, X. Long, and Y. Zhao, "A Verified Capability-Based Model for Information Flow Security With Dynamic Policies," *IEEE Access*, vol. 6, pp. 16395–16407, Mar. 2018, doi: 10.1109/ACCESS.2018.2815766.
- [63] T. Li, K. Wang, and J. Horkoff, "Towards effective assessment for social engineering attacks," in *Proceedings of the IEEE International Conference on Requirements Engineering*, IEEE Computer Society, Sep. 2019, pp. 392–397. doi: 10.1109/RE.2019.00051.
- [64] A. Șandor, G. Tont, and E. Simion, "A Mathematical Model for Risk Assessment of Social Engineering Attacks," *TEM Journal*, vol. 11, no. 1, pp. 334–338, Feb. 2022, doi: 10.18421/TEM111-42.
- [65] Y. Alkhurayyif and G. R. S. Weir, "Readability as a basis for information security policy assessment," in *2017 Seventh International Conference on Emerging Security Technologies (EST)*, IEEE, Sep. 2017, pp. 114–121. doi: 10.1109/EST.2017.8090409.
- [66] L. Bošnjak and B. Brumen, "Shoulder surfing experiments: A systematic literature review," Dec. 01, 2020, Elsevier Ltd. doi: 10.1016/j.cose.2020.102023.
- [67] J. E. McNealy, "Platforms as phish farms: Deceptive social engineering at scale," *New Media Soc*, vol. 24, no. 7, pp. 1677–1694, Jul. 2022, doi: 10.1177/14614448221099228.
- [68] Z. Wang, H. Zhu, and L. Sun, "Social engineering in cybersecurity: Effect mechanisms, human vulnerabilities and attack methods," *IEEE Access*, vol. 9, pp. 11895–11910, 2021, doi: 10.1109/ACCESS.2021.3051633.
- [69] M. Mattera and M. M. Chowdhury, "Social Engineering: The Looming Threat," in *IEEE International Conference on Electro Information Technology*, IEEE Computer Society, May 2021, pp. 56–61. doi: 10.1109/EIT51626.2021.9491884.
- [70] A. Yasin, R. Fatima, L. Liu, J. Wang, and R. Ali, "Understanding Social Engineers Strategies from the Perspective of Sun-Tzu Philosophy," in *Proceedings - 2020 IEEE 44th Annual Computers, Software, and Applications Conference, COMPSAC 2020*, Institute of Electrical and Electronics Engineers Inc., Jul. 2020, pp. 1773–1776. doi: 10.1109/COMPSAC48688.2020.00045.
- [71] M. R. Arabia-Obedoza, G. Rodriguez, A. Johnston, F. Salahdine, and N. Kaabouch, "Social Engineering Attacks A Reconnaissance Synthesis Analysis," in *2020 11th IEEE Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2020*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020, pp. 0843–0848. doi: 10.1109/UEMCON51285.2020.9298100.
- [72] K. S. Jones, M. E. Armstrong, M. K. Tornblad, and A. Siami Namin, "How social engineers use persuasion principles during vishing attacks," *Information and Computer Security*, vol. 29, no. 2, pp. 314–331, 2020, doi: 10.1108/ICS-07-2020-0113.
- [73] W. Fuertes et al., "Impact of Social Engineering Attacks: A Literature Review," in *Smart Innovation, Systems and Technologies*, Springer

- Science and Business Media Deutschland GmbH, 2022, pp. 25–35. doi: 10.1007/978-981-16-4884-7_3.
- [74] N. M. C. Galego, R. M. Pascoal, and P. R. Brandao, “BYOD: Impact in Architecture and Information Security Corporate Policy,” in 2022 17th Iberian Conference on Information Systems and Technologies (CISTI), IEEE, Jun. 2022, pp. 1–2. doi: 10.23919/CISTI54924.2022.9820043.
- [75] A. Bryman and E. Bell, *Business Research Methods*. Oxford University Press, 2015. [Online]. Available: <https://books.google.com.my/books?id=17u6BwAAQBAJ>
- [76] Y. Alkhurayyif and G. R. S. Weir, “Evaluating Readability as a Factor in Information Security Policies,” *International Journal of Trend in Research and Development*, pp. 20–22, 2017, Accessed: Aug. 25, 2024. [Online]. Available: <https://strathprints.strath.ac.uk/id/eprint/63070>
- [77] T. Koide, D. Chiba, M. Akiyama, K. Yoshioka, and T. Matsumoto, “To get lost is to learn the way: An analysis of multi-step social engineering attacks on the web,” *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E104A, no. 1, pp. 162–181, Jan. 2021, doi: 10.1587/transfun.2020CIP0005.
- [78] N. Tsinganos, P. Fouliras, G. Sakellariou, and I. Mavridis, “Towards an automated recognition system for chat-based social engineering attacks in enterprise environments,” in *ACM International Conference Proceeding Series*, Association for Computing Machinery, Aug. 2018. doi: 10.1145/3230833.3233277.
- [79] T. Nelms, R. Perdisci, M. Antonakakis, and M. Ahamad, “Towards measuring and mitigating social engineering software download attacks,” in *Proceedings of the 25th USENIX Conference on Security Symposium*, in SEC’16. USA: USENIX Association, 2016, pp. 773–789.
- [80] F. Mouton, L. Leenen, and H. S. Venter, “Social engineering attack examples, templates and scenarios,” *Comput Secur*, vol. 59, pp. 186–209, Jun. 2016, doi: 10.1016/j.cose.2016.03.004.
- [81] M. Edwards, R. Larson, B. Green, A. Rashid, and A. Baron, “Panning for gold: Automatically analysing online social engineering attack surfaces,” *Comput Secur*, vol. 69, pp. 18–34, Aug. 2017, doi: 10.1016/j.cose.2016.12.013.
- [82] H. Wilcox and M. Bhattacharya, “A framework to mitigate social engineering through social media within the enterprise,” in *Proceedings of the 2016 IEEE 11th Conference on Industrial Electronics and Applications*, ICIEA 2016, Institute of Electrical and Electronics Engineers Inc., Oct. 2016, pp. 1039–1044. doi: 10.1109/ICIEA.2016.7603735.
- [83] V. M. I. A. Hartl and U. Schmuntzsch, “Fraud protection for online banking: A user-centered approach on detecting typical double-dealings due to social engineering and inobservance whilst operating with personal login credentials,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer Verlag, 2016, pp. 37–47. doi: 10.1007/978-3-319-39381-0_4.
- [84] A. Jamil, K. Asif, Z. Ghulam, M. K. Nazir, S. Mudassar Alam, and R. Ashraf, “MPMPA: A Mitigation and Prevention Model for Social Engineering Based Phishing attacks on Facebook,” in 2018 IEEE International Conference on Big Data (Big Data), IEEE, Dec. 2018, pp. 5040–5048. doi: 10.1109/BigData.2018.8622505.
- [85] H. Aldawood and G. Skinner, “Analysis and Findings of Social Engineering Industry Experts Explorative Interviews: Perspectives on Measures, Tools, and Solutions,” *IEEE Access*, vol. 8, pp. 67321–67329, 2020, doi: 10.1109/ACCESS.2020.2983280.
- [86] F. Goodarzian, P. Ghasemi, V. Kumar, and A. Abraham, “A new modified social engineering optimizer algorithm for engineering applications,” *Soft comput*, vol. 26, no. 9, pp. 4333–4361, May 2022, doi: 10.1007/s00500-022-06837-y.
- [87] R. Heartfield and G. Loukas, “A Taxonomy of Attacks and a Survey of Defence Mechanisms for Semantic Social Engineering Attacks,” *ACM Comput Surv*, vol. 48, no. 3, pp. 1–39, Feb. 2016, doi: 10.1145/2835375.
- [88] C. Atwell, T. Blasi, and T. Hayajneh, “Reverse TCP and Social Engineering Attacks in the Era of Big Data,” in 2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS), IEEE, Apr. 2016, pp. 90–95. doi: 10.1109/BigDataSecurity-HPSC-IDS.2016.60.
- [89] Z. Wang, L. Sun, and H. Zhu, “Defining Social Engineering in Cybersecurity,” *IEEE Access*, vol. 8, pp. 85094–85115, 2020, doi: 10.1109/ACCESS.2020.2992807.
- [90] A. Algarni, Y. Xu, and T. Chan, “Social engineering in social networking sites: The art of impersonation,” in *Proceedings - 2014 IEEE International Conference on Services Computing, SCC 2014*, Institute of Electrical and Electronics Engineers Inc., Oct. 2014, pp. 797–804. doi: 10.1109/SCC.2014.108.
- [91] S. Uebelacker and S. Quiel, “The social engineering personality framework,” in *Proceedings - 4th Workshop on Socio-Technical Aspects in Security and Trust, STAST 2014 - Co-located with 27th IEEE Computer Security Foundations Symposium, CSF 2014 in the Vienna Summer of Logic 2014*, Institute of Electrical and Electronics Engineers Inc., Dec. 2014, pp. 24–30. doi: 10.1109/STAST.2014.12.
- [92] I. Del Pozo, M. Iturralde, and F. Restrepo, “Social engineering: Application of psychology to information security,” in *Proceedings - 2018 IEEE 6th International Conference on Future Internet of Things and Cloud Workshops, W-FiCloud 2018*, Institute of Electrical and Electronics Engineers Inc., Oct. 2018, pp. 108–114. doi: 10.1109/W-FiCloud.2018.00023.
- [93] F. Mouton, L. Leenen, and H. S. Venter, “Social Engineering Attack Detection Model: SEADMv2,” in *Proceedings - 2015 International Conference on Cyberworlds, CW 2015*, Institute of Electrical and Electronics Engineers Inc., Feb. 2015, pp. 216–223. doi: 10.1109/CW.2015.52.
- [94] A. Yasin, R. Fatima, L. Liu, J. Wang, R. Ali, and Z. Wei, “Understanding and deciphering of social engineering attack scenarios,” *Security and Privacy*, vol. 4, no. 4, Jul. 2021, doi: 10.1002/spy2.161.
- [95] A. M. Aroyo, F. Rea, G. Sandini, and A. Sciutti, “Trust and Social Engineering in Human Robot Interaction: Will a Robot Make You Disclose Sensitive Information, Conform to Its Recommendations or Gamble?,” *IEEE Robot Autom Lett*, vol. 3, no. 4, pp. 3701–3708, Oct. 2018, doi: 10.1109/LRA.2018.2856272.
- [96] A. Cullen and L. Armitage, “The social engineering attack spiral (SEAS),” in 2016 International Conference on Cyber Security and Protection of Digital Services, Cyber Security 2016, Institute of Electrical and Electronics Engineers Inc., Jun. 2016. doi: 10.1109/CyberSecPODS.2016.7502347.
- [97] F. Mouton, L. Leenen, M. M. Malan, and H. S. Venter, “Towards an Ontological Model Defining the Social Engineering Domain,” 2014, pp. 266–279. doi: 10.1007/978-3-662-44208-1_22.
- [98] K. Zheng, T. Wu, X. Wang, B. Wu, and C. Wu, “A Session and Dialogue-Based Social Engineering Framework,” *IEEE Access*, vol. 7, pp. 67781–67794, 2019, doi: 10.1109/ACCESS.2019.2919150.
- [99] F. Mouton, M. M. Malan, L. Leenen, and H. S. Venter, “Social engineering attack framework,” in 2014 Information Security for South Africa - Proceedings of the ISSA 2014 Conference, Institute of Electrical and Electronics Engineers Inc., Nov. 2014. doi: 10.1109/ISSA.2014.6950510.
- [100] C. Lekati, “Complexities in Investigating Cases of Social Engineering: How Reverse Engineering and Profiling can Assist in the Collection of Evidence,” in *Proceedings - 11th International Conference on IT Security Incident Management and IT Forensics, IMF 2018*, Institute of Electrical and Electronics Engineers Inc., Oct. 2018, pp. 107–109. doi: 10.1109/IMF.2018.00015.
- [101] P. Burda, L. Allodi, and N. Zannone, “Dissecting Social Engineering Attacks through the Lenses of Cognition,” in *Proceedings - 2021 IEEE European Symposium on Security and Privacy Workshops, Euro S and PW 2021*, Institute of Electrical and Electronics Engineers Inc., Sep. 2021, pp. 149–160. doi: 10.1109/EuroSPW54576.2021.00024.
- [102] A. Zingerle, “How to obtain passwords of online scammers by using social engineering methods,” in *Proceedings - 2014 International Conference on Cyberworlds, CW 2014*, Institute of Electrical and Electronics Engineers Inc., Dec. 2014, pp. 340–344. doi: 10.1109/CW.2014.54.
- [103] N. Tsinganos, I. Mavridis, and D. Gritzalis, “Utilizing Convolutional Neural Networks and Word Embeddings for Early-Stage Recognition of

- Persuasion in Chat-Based Social Engineering Attacks,” IEEE Access, vol. 10, pp. 108517–108529, 2022, doi: 10.1109/ACCESS.2022.3213681.
- [104] R. Heartfield, G. Loukas, and D. Gan, “You Are Probably Not the Weakest Link: Towards Practical Prediction of Susceptibility to Semantic Social Engineering Attacks,” IEEE Access, vol. 4, pp. 6910–6928, 2016, doi: 10.1109/ACCESS.2016.2616285.
- [105] A. Algarni, Y. Xu, and T. Chan, “Measuring source credibility of social engineering attackers on Facebook,” in Proceedings of the Annual Hawaii International Conference on System Sciences, IEEE Computer Society, Mar. 2016, pp. 3686–3695. doi: 10.1109/HICSS.2016.460.
- [106] M. Hijji and G. Alam, “A Multivocal Literature Review on Growing Social Engineering Based Cyber-Attacks/Threats during the COVID-19 Pandemic: Challenges and Prospective Solutions,” IEEE Access, vol. 9, pp. 7152–7169, 2021, doi: 10.1109/ACCESS.2020.3048839.
- [107] E. U. Osuagwu, G. A. Chukwudebe, T. Saliu, and V. N. Chukwudebe, “Mitigating social engineering for improved cybersecurity,” in CYBER-Abuja 2015 - International Conference on Cyberspace Governance: The Imperative for National and Economic Security - Proceedings, Institute of Electrical and Electronics Engineers Inc., Dec. 2015, pp. 91–100. doi: 10.1109/CYBER-Abuja.2015.7360515.
- [108] P. Schaab, K. Beckers, and S. Pape, “Social engineering defence mechanisms and counteracting training strategies,” 2017, Emerald Group Publishing Ltd. doi: 10.1108/ICS-04-2017-0022.
- [109] J. W. Bullee and M. Junger, “How effective are social engineering interventions? A meta-analysis,” Nov. 04, 2020, Emerald Group Holdings Ltd. doi: 10.1108/ICS-07-2019-0078.
- [110] N. Abe and M. Soltys, “Deploying Health Campaign Strategies to Defend Against Social Engineering Threats,” Procedia Comput Sci, vol. 159, pp. 824–831, 2019, doi: 10.1016/j.procs.2019.09.241.
- [111] C. C. Campbell, “Solutions for counteracting human deception in social engineering attacks,” Information Technology and People, vol. 32, no. 5, pp. 1130–1152, Sep. 2019, doi: 10.1108/ITP-12-2017-0422.
- [112] F. Salahdine and N. Kaabouch, “Social engineering attacks: A survey,” 2019, MDPI AG. doi: 10.3390/FII1040089.
- [113] H. Siadati, T. Nguyen, P. Gupta, M. Jakobsson, and N. Memon, “Mind your SMSes: Mitigating Social Engineering in Second Factor Authentication,” 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S016740481630116X>
- [114] F. Mouton, A. Nottingham, L. Leenen, and H. S. Venter, “Finite State Machine for the Social Engineering Attack Detection Model: SEADM,” SAIEE Africa Research Journal, vol. 109, no. 2, pp. 133–148, Jun. 2018, doi: 10.23919/SAIEE.2018.8531953.
- [115] R. J. Heartfield, “Utilising the concept of human-as-a-security-sensor for detecting semantic social engineering attacks,” PhD thesis, University of Greenwich, 2017.
- [116] A. Algarni, “Social Engineering in Social Networking Sites: Phase-Based and Source-Based Models,” International Journal of e-Education, e-Business, e-Management and e-Learning, 2013, doi: 10.7763/IJEEEE.2013.V3.278.
- [117] R. Heartfield and G. Loukas, “Detecting semantic social engineering attacks with the weakest link: Implementation and empirical evaluation of a human-as-a-security-sensor framework,” Comput Secur, vol. 76, pp. 101–127, Jul. 2018, doi: 10.1016/j.cose.2018.02.020.
- [118] S. Albladi and G. R. S. Weir, “Vulnerability to social engineering in social networks: a proposed user-centric framework,” in 2016 IEEE International Conference on Cybercrime and Computer Forensic (ICCCF), IEEE, Jun. 2016, pp. 1–6. doi: 10.1109/ICCCF.2016.7740435.
- [119] D. Al-dablan, A. Al-hamad, R. Al-Bahlal, and M. Altaib Badawi, “An Analysis of Various Social Engineering Attack in Social Network using Machine Learning Algorithm,” IJCSNS International Journal of Computer Science and Network Security, vol. 20, no. 10, p. 46, 2020, doi: 10.22937/IJCSNS.2020.20.10.7.
- [120] N. Mamedova, A. Urintsov, O. Staroverova, E. Ivanov, and D. Galahov, “Social engineering in the context of ensuring information security,” SHS Web of Conferences, vol. 69, p. 00073, 2019, doi: 10.1051/shsconf/20196900073.
- [121] R. Anand, S. Medhavi, V. Soni, C. Malhotra, and D. K. Banwet, “Transforming information security governance in India (A SAP-LAP based case study of security, IT policy and e-governance),” Information and Computer Security, vol. 26, no. 1, pp. 58–90, 2018, doi: 10.1108/ICS-12-2016-0090.
- [122] S. K. Jansen van Rensburg, “End-User Perceptions on Information Security,” Journal of Global Information Management, vol. 29, no. 6, pp. 1–16, Dec. 2021, doi: 10.4018/JGIM.293290.
- [123] E. Rostami, F. Karlsson, and S. Gao, “Policy Components - A Conceptual Model for Tailoring Information Security Policies,” in IFIP Advances in Information and Communication Technology, Springer Science and Business Media Deutschland GmbH, 2022, pp. 265–274. doi: 10.1007/978-3-031-12172-2_21.
- [124] M. Kang, T. Lee, and S. Um, “Establishment of Methods for Information Security System Policy Using Benchmarking,” in Proceedings - 29th IEEE International Symposium on Software Reliability Engineering Workshops, ISSREW 2018, Institute of Electrical and Electronics Engineers Inc., Nov. 2018, pp. 237–242. doi: 10.1109/ISSREW.2018.00012.
- [125] T. Grassegger and D. Nedbal, “The Role of Employees’ Information Security Awareness on the Intention to Resist Social Engineering,” Procedia Comput Sci, vol. 181, pp. 59–66, 2021, doi: 10.1016/j.procs.2021.01.103.
- [126] F. L. Greitzer, J. R. Strozer, S. Cohen, A. P. Moore, D. Mundie, and J. Cowley, “Analysis of unintentional insider threats deriving from social engineering exploits,” in Proceedings - IEEE Symposium on Security and Privacy, Institute of Electrical and Electronics Engineers Inc., Nov. 2014, pp. 236–250. doi: 10.1109/SPW.2014.39.
- [127] S. Vrhovec, I. Bernik, and B. Markelj, “Explaining information seeking intentions: Insights from a Slovenian social engineering awareness campaign,” Comput Secur, vol. 125, Feb. 2023, doi: 10.1016/j.cose.2022.103038.
- [128] S. M. Albladi and G. R. S. Weir, “Predicting individuals’ vulnerability to social engineering in social networks,” Cybersecurity, vol. 3, no. 1, Dec. 2020, doi: 10.1186/s42400-020-00047-5.
- [129] N. Wulandari, M. S. Adnan, and C. B. Wicaksono, “Are You a Soft Target for Cyber Attack? Drivers of Susceptibility to Social Engineering-Based Cyber Attack (SECA): A Case Study of Mobile Messaging Application,” Hum Behav Emerg Technol, vol. 2022, 2022, doi: 10.1155/2022/5738969.
- [130] A. Algarni, “What message characteristics make social engineering successful on Facebook: The role of central route, peripheral route, and perceived risk,” Information (Switzerland), vol. 10, no. 6, Jun. 2019, doi: 10.3390/info10060211.
- [131] L. Karadsheh, H. Alryalat, J. Alqatawna, S. F. Alhawari, and M. A. AL Jarrah, “The impact of social engineer attack phases on improved security countermeasures: Social engineer involvement as mediating variable,” International Journal of Digital Crime and Forensics, vol. 14, no. 1, pp. 1–26, Jan. 2022, doi: 10.4018/IJDCF.286762.
- [132] K. P. Gallagher, X. Zhang, and V. C. Gallagher, “Institutional drivers of assimilation of information security policies and procedures in U.S. firms: Test of an empirical model,” in Proceedings of the Annual Hawaii International Conference on System Sciences, IEEE Computer Society, Mar. 2015, pp. 4700–4709. doi: 10.1109/HICSS.2015.559.
- [133] S. Solak and Y. Zhuo, “Optimal policies for information sharing in information system security,” Eur J Oper Res, vol. 284, no. 3, pp. 934–950, Aug. 2020, doi: 10.1016/j.ejor.2019.12.016.

Comprehensive Bibliometric Literature Review of Chatbot Research: Trends, Frameworks, and Emerging Applications

Nazruddin Safaat Harahap¹, Aslina Saad^{2*}, Nor Hasbiah Ubaidullah³

The SIG of Information Systems and Technology Integration (ISTI)-Faculty of Computing and META-Technology,
Universiti Pendidikan Sultan Idris (UPSI), Malaysia^{1, 2, 3}

Teknik Informatika-Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia¹

Abstract—This study aims to conduct a comprehensive bibliometric literature review of chatbot research by examining key trends, frameworks, and influential applications across various domains. It seeks to map the evolution of chatbot technologies, identify influential works, and analyze how the research focus has shifted over time, particularly towards AI-driven chatbot frameworks. An expanded dataset was compiled from the Scopus database, and bibliometric analyses were conducted using n-gram reference analysis, network mapping, and temporal trend visualization. The analysis was performed using R Studio with Biblioshiny, allowing for the identification of thematic clusters and the progression from rule-based to advanced retrieval and generative language model paradigms in chatbot research. Chatbot research has grown significantly from 2020 to 2024, with rising publication volumes and increased global collaboration, led by contributions from the USA, China, and emerging regions, such as Southeast Asia. Thematic analysis highlights a shift from foundational AI and NLP technologies to specialized applications such as mental health chatbots and e-commerce systems, emphasizing practical and user-centered solutions. Advances in chatbot architectures, including generative AI, have demonstrated the field's interdisciplinary nature and trajectory towards sophisticated, context-aware conversational systems. The analysis primarily used data from Scopus, which may limit the breadth of the included research. Future studies are encouraged to integrate data from other sources, such as the Web of Science (WoS) and PubMed, for a more comprehensive understanding of the field.

Keywords—Chatbot research; bibliometric literature review; retrieval and generative; trend visualization

I. INTRODUCTION

Current technological trends indicate that chatbots are popular applications that use Artificial Intelligence [1], Machine Learning [2], Deep Learning [3], and Natural Language Processing [4]. They are used in various fields, including education, health, universities, schools, nursing, and business. Computers are becoming increasingly capable of performing tasks that humans perform. With artificial intelligence and machine learning technology, computer systems have become more compact and efficient in understanding voice and text interactions, leading to an effective speed [5].

Numerous prominent chatbots utilizing deep learning techniques have been created by major corporations, including ChatGPT by OpenAI, Alexa by Amazon, Siri by Apple, Google

Assistant by Google, Cortana by Microsoft, and Watson by IBM [6]. These chatbots serve as personal assistants in daily activities. They communicate through voice and integration with devices, such as smartphones, smartwatches, and cars. Owing to the rapid growth of chatbot applications, technical analysis methods are required, and a framework that is widely used in building chatbots today classifies chatbots into two distinct categories: task-oriented and non-task-oriented [7]. Task-oriented aspects operate according to human instructions, whereas non-task-oriented aspects have multiple objectives but cannot carry out specific activities. Non-task-oriented chatbot architectures can be classified into generative-based and retrieval-based. Chatbots are generally used to answer the questions asked by users. Typically, users initiate a conversation with this application and provide answers after analyzing the questions. Chatbot users must have convenient, flexible, and real-time access to ensure successful use [8].

This study investigates the trend of chatbot development, the sectors that use chatbots, and the frameworks and methodologies applied to chatbot development. Several research initiatives in the field of chatbots have used bibliographic analyses conducted by previous scholars. Io et al. [9] conducted an extensive bibliometric analysis of chatbot literature, emphasizing the growth of research beginning in 2015. This study suggests that future research should investigate new technologies and uses for chatbots, focusing on both user and business viewpoints to optimize chatbot functionality. The authors used traditional methodologies in chatbot development, focusing solely on Natural Language Processing techniques, such as natural voice, facial emotions, and body movements. Recent advances in deep learning require additional research to explore the potential applications of this cutting-edge technology for chatbot development across several domains.

Adomopoulou et al. [10] analyze the evolution of chatbots, focusing on their transition from simple rule-based systems to sophisticated machine-learning models. Researchers have critically analyzed various applications of chatbots in different domains, focusing on the importance of natural language processing and user interaction. It examines the current obstacles and potential future directions of chatbot technology, and suggests further research and development avenues to improve chatbot effectiveness and user satisfaction. The analysis conducted in this study reveals a persistent trend towards a high preference for basic authors. However, there is a

noticeable lack of comprehensive investigation into specific health topics, such as 'older people' or 'mHealth,' which are overshadowed by a broader emphasis on AI-based outcomes.

Pears et al.[11] conducted a bibliometric analysis to examine the increasing prevalence of chatbot applications in the healthcare sector. The increase in patient interactions has been attributed to advancements in AI and ML, particularly in NLP. Since 2016, there has been a substantial surge in research productivity, characterized by a transition from technology to studies centered around artificial intelligence and its practical applications. The research potential of advanced AI approaches has been emphasized by researchers who have specifically focused on developing user-friendly interfaces, tailored interactions, and integrating chatbots within various healthcare contexts. Furthermore, our analysis underscores the need to employ co-creation techniques that include stakeholders to enhance the efficacy of conversational agents designed for healthcare applications.

The objective of this study was to examine the trends and suggestions found in publications on chatbots, utilizing a bibliometric analysis approach. This study also presented visualizations of the current movement in chatbot development on different topics. The following research problems were addressed using data from the Scopus database.

RQ1: Who are the most prolific authors in chatbot, and what are their key research themes and topics?

RQ2: What are the most active countries in chatbot, and how does this vary across different regions and time periods?

RQ3: What are the prevalent keywords employed in the field of chatbots developed through bibliometric analysis?

RQ4: What research trends relate to chatbots (thematic and evolution trends), including journals, fields, countries, and universities?

RQ5: Which components, techniques, and frameworks are most widely used in chatbot development?

II. LITERATURE REVIEW

This literature review provides a comprehensive, updated overview of research developments in chatbot technology, focusing on recent innovations in AI, ML, NLP, and deep learning. This section surveys studies that address the expansion and application of chatbots across various domains, emphasizing the intellectual and social structures that shape current trends. By highlighting methodologies, frameworks, and analytical approaches from previous studies, this review aims to establish a foundation for understanding the gaps, advancements, and emerging directions in chatbot research, thereby framing the significance and context of the present study.

A. Historical Development

The history of chatbot development has evolved significantly from early rule-based systems to sophisticated AI-driven models. Over the years, advancements in machine learning, natural language processing, and neural networks have revolutionized chatbot capabilities. By the 2010s, major

companies such as Apple, Amazon, and Google had introduced intelligent virtual assistants—Siri, Alexa, and Google Assistant—that leveraged voice recognition and vast data processing to perform more complex tasks. These developments signaled a shift towards conversational AI with greater versatility, accuracy, and personalization. In recent years, deep learning and transformer models such as OpenAI's GPT series have marked significant milestones, enabling chatbots to generate human-like responses and comprehend contexts with remarkable accuracy. Studies often employ advanced research designs, such as bibliometric analyses, to map intellectual and social trends in chatbot development. This evolving landscape demonstrates a continued effort to optimize conversational agents for diverse applications, including healthcare, education, and customer service, with an emphasis on user-centric frameworks that balance functionality with ethical considerations.

B. Recent Development

Recent developments in chatbot research reflect a shift towards integrating emerging technologies, such as advanced NLP, deep learning, and transformer models, notably the architecture underlying OpenAI's GPT series. Researchers are increasingly examining chatbot applications in diverse fields, including healthcare, government, and education, where the demand for real-time, personalized, and secure interactions has grown. This expansion has led to new concerns, especially regarding data privacy and cybersecurity, because chatbots handle sensitive user data. In the healthcare sector, chatbots now assist in patient management and mental health support, whereas in the government, they provide accessible public service information. These applications underscore the importance of secure, efficient, and adaptable chatbot systems that can respond to domain-specific requirements. In terms of methodology, recent research has often used mixed methods, combining quantitative and qualitative data to capture user satisfaction and chatbot performance insights. Machine learning and NLP integration have become essential, enabling chatbots to process and understand complex queries more effectively. These trends indicate that future research will likely focus on enhancing chatbot trustworthiness, expanding their usability across contexts, and addressing ethical considerations in AI-driven interactions, ensuring that chatbot systems align with user expectations and privacy needs.

C. Previous Studies on Bibliometric Analysis

Previous studies on the bibliometric analysis of chatbots and conversational agents have primarily focused on mapping the field's growth, identifying prominent authors, and analyzing keyword trends. For instance, Io et al. [9]. conducted a bibliometric analysis that emphasized the rapid rise in chatbot research since 2015, focusing on NLP and AI advancements. However, their methodology is limited by its reliance on traditional NLP techniques, and the lack of integration of emerging deep-learning methods is now central to chatbot development. Similarly, Adomopoulou et al. [10] examined the transition of chatbots from rule-based to machine-learning-driven systems, highlighting applications across various sectors, but noting an underrepresentation of specific areas, such as health-related chatbots tailored for older adults.

Pears et al. [11] analyzed chatbots in healthcare and revealed the increasing role of conversational agents in enhancing patient interaction. However, further research on ethical considerations and patient data security is required. Collectively, these studies provide valuable insights into the evolution of chatbot research. However, they are limited by their focus on specific technologies or sectors and often overlook broader applications or recent AI advancements. Notably, there is a gap in the comprehensive frameworks that assess chatbot adoption across multiple domains. Future research could address these gaps by incorporating advanced bibliometric techniques, examining ethical implications in more depth, and exploring the diverse applications of chatbots in emerging fields, such as government services and personalized education.

Manigandan et al. [12] conducted a bibliometric analysis for mapping research landscapes across various fields. Chatbots, conversational agents, and virtual assistants were employed to identify publication trends, key research themes, and collaborative networks among authors and institutions. Recent studies have revealed a significant increase in chatbot-related research since 2018, highlighting its growing importance in business, management, and accounting. Common themes include chatbot design, customer experience enhancement, and business operations automation, although ethical concerns such as privacy and transparency remain underexplored. Despite increased interest, the field is characterized by limited collaboration among researchers and nations, underscoring the need for more inclusive partnerships. Future studies should

address underrepresented areas such as the ethical implications of chatbot use, user perceptions, and long-term organizational impacts. Expanding research into industries such as healthcare and education could also uncover the unique challenges and opportunities for chatbot adoption.

A similar study was conducted by Tawar et al. [13], who focused on the field of business management, although this research is rooted in computer science applications. Bibliometric analyses have proven valuable for systematically reviewing research trends, themes, and gaps across disciplines. In chatbot research, these analyses reveal the growing adoption and application of chatbots in areas such as customer service, marketing, and other sectors, showing their positive impact on customer engagement, trust, and business efficiency. However, significant limitations remain, including a focus on developed countries and insufficient exploration of ethical concerns such as privacy and transparency. Moreover, many studies rely on short-term data, lacking the longitudinal insights necessary to understand the evolving user behavior and technological progress. Addressing these gaps requires exploring cross-cultural contexts, developing ethical frameworks for chatbot use, and expanding applications in underexplored fields, such as healthcare and education. Future research integrating interdisciplinary theoretical approaches and longitudinal studies could deepen understanding and drive innovative developments in chatbot technologies. The summary of previous studies is presented in Table I.

TABLE I. SUMMARY OF PREVIOUS STUDIES

Author	Domain & Search Query	Objective of the Study	Total Document, Data Source & Coverage	Attributes Examined	Main Findings
Tanwar et al.[13]	The range of subject areas was restricted to “Business, Management and Accounting”, “Social Sciences”, “Psychology” and “Multidisciplinary”.	science mapping, performance analysis, and bibliographic coupling to identify significant trends and areas of research emphasis	798 articles from Scopus.	Publication Trends, Most Productive and Influential Countries, Most Prolific Authors, Most Prolific Contributing Institutions, Most Prolific Journals, Intellectual Structure of Chatbots.	Key areas of chatbot research include: applications of chatbots, behavioral and relational effects of chatbot use, and factors influencing chatbot adoption, including barriers. The United States leads in contributions, with the highest number of research articles and citations in this field.
Agarwal et al.[16]	Domain: Chatbots in Computer areas. Search Query: (TITLE-ABS-KEY (“chatbots” AND “virtual assistants”))	to examine the past research, to provide a conclusive mark, to explore the combination of keywords used in chatbots	130 articles from Scopus.	Theoretical contributions, Research implications, Managerial implications, Social implications.	The authors with maximum number of citations are Yan, Zaho, Bengio, Weizenbaum, Song, Zhou and Maedche with jointly 180 citations.
L et al. [12]	Domain: Chatbot in Business Management and Accounting	to identify key publication metrics, examine the intellectual structure, and explore the social structure of research in this field.	378 articles from scopus	number of publications and citations over time, productive authors, country productivity, h-index and intellectual structure	The keywords “Chatbot”, “conversational agent” and “virtual assistant” emerged as the most frequently employed terms in the majority of publications.
Io et al.[9]	consider the two terms “chatbot” and “conversational agent”	s to help researchers to identify research gaps for the future research agenda in chatbots	4,246 articles from wos and proQuest	Clustering keywords, co-occurrences.	The results of the analysis found a potential research opportunity in chatbot
Xia et al.[17]	Domain “AI Chatbo”	to offer bibliometric assessments of the expanding literature about AI chatbot services	571 article from scopus	the most influential work, authors, and co-cited authors on AI chatbots	based on the author’s cocitation analysis and the intellectual structure, Computer science is the most critical discipline regarding AI applications

III. METHODS

This Bibliometric Literature Review (BLR) utilized the bibliography analysis and meta-analysis approach. Bibliometric analysis is a crucial and effective method for comprehending the progression and patterns of research. Science mapping analysis is an integral part of bibliometric analysis that aims to provide a structured examination of trends periodically, studied themes, changes in fields of knowledge, and productive researchers [14], [15].

A. Search Strategy

Search Strategy: Explain how you identified the relevant literature for bibliometric analysis. This may include details such as the databases and search terms used, and any inclusion or exclusion criteria applied to the search results. The period covered in your analysis should also be described.

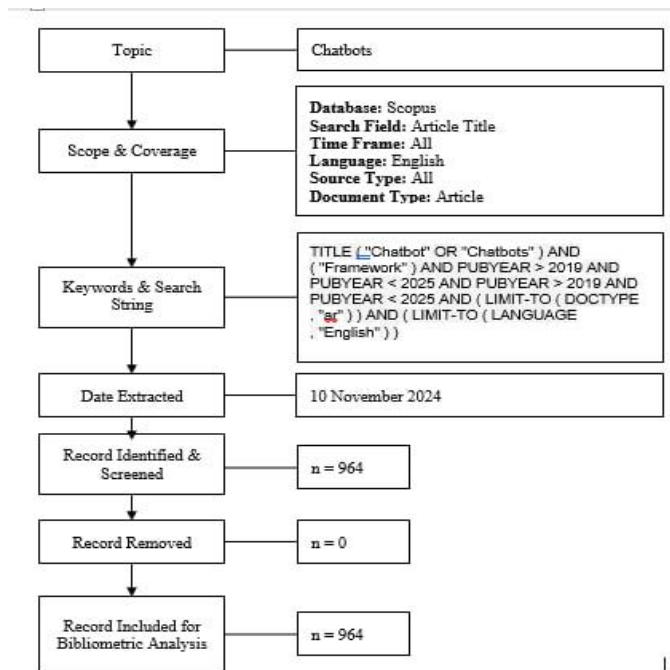


Fig. 1. Flow diagram of the search strategy.

Fig. 1 shows the stages in collecting data starting from topic and scope, keywords and search query, date extracted, record removed, record identified, and record included for bibliometric analysis.

B. Data Collection

The Search Strategy stages encompass distinct procedures: data extraction, identification, screening, record removal, and record inclusion in bibliometric analysis. We used Search Strategy stages to determine entries from the Scopus database by executing a query. The search yielded 964 documents.

C. Data Cleaning and Harmonization

To ensure the accuracy and consistency of the data imported into Biblioshiny, rigorous data cleaning was performed using the OpenRefine tool. The BibTeX file generated by Biblioshiny was exported to OpenRefine, and the dataset was verified to match the original source. Specific adjustments included standardizing

author data by converting them to lowercase data. The integrity of the dataset was validated through a series of analytical processes using OpenRefine, ensuring that it was clean and suitable for bibliometric analysis.

D. Tools

The Scopus database was used to obtain studies for a comprehensive evaluation of chatbot research. The identification process began with a well-defined search strategy that was used to query the Scopus database and retrieve relevant entries. Following data collection, the articles were analyzed using the Biblioshiny tool in R Studio, enabling a detailed bibliometric analysis and visualization of research trends.

IV. RESULTS

This section examines publications, research activities, journals, papers, and references, providing an analytical overview. Following the analysis, the findings are further explored and discussed to offer deeper insights into the subject matter.

A. Analysis of Publications

The main information from the article was entered into R bibliometric software for bibliometric analysis. From the analysis results, a significant increase in research in the field of chatbots was found from year to year. The growth of this research is depicted in the following Table II.

TABLE II. MAIN INFORMATION DATA

Main Information	Data
Publication Years	2010 - 2024
Total Publications	955
Annual Growth Rate %	60.28
Document Average Age	1.13
Average citations per doc	20.61
Keywords Plus (ID)	3061
Author's Keywords (DE)	2848
Single-authored docs	66
Co-Authors per Doc	3.91
International co-authorships	27.64

From 2010 to 2024, there were 955 publications with an impressive annual growth rate of 60.28%, indicating a rapidly expanding field. The average document age is 1.13 years, and each publication receives an average of 20.61 citations, reflecting substantial academic impact. Collaboration is prevalent, with 3.91 co-authors per document on average, and 27.64% of publications involve international co-authorships, highlighting the global nature of the research.

Fig. 2 shows Annual scientific production has grown consistently from 2020 to 2024, reflecting a substantial increase in research activity. Starting with 60 articles in 2020, the output increased to 98 articles in 2021 and 142 in 2022. This upward trend continues with 259 articles in 2023 and peaks at 396 articles in 2024, indicating an accelerating interest and investment in the field over the five-year period.

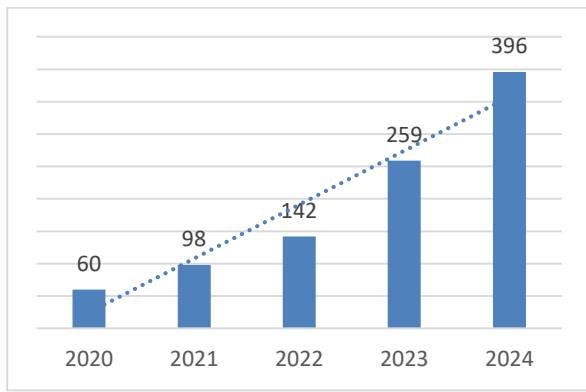


Fig. 2. Annual scientific productions.

TABLE III. AVERAGE CITATIONS PER YEAR

Year	Mean TCperArt	N	Mean TCperYear	CitableYears
2020	84.27	60.00	16.85	5
2021	49.37	98.00	12.34	4
2022	30.42	142.00	10.14	3
2023	15.67	259.00	7.84	2
2024	3.55	396.00	3.55	1

Table III provides information on the Mean Total Citations per article (MeanTCperArt), the number of articles (N of article), and the resulting Mean Total Citations per year (MeanTCperYear) for these years 2020 to 2024. In 2020, the mean total number of citations per article was 84.27, with eight articles contributing to a mean total number of citations per year of 18.85. Subsequently, there is a noticeable increase in the mean total citations per article per year in 2020, 2021, and 2022, followed by a significant decrease in 2023 and 2024. Measure an author's impact using the citation count, average citation rate, h-index, g-index, m-index, total citations, number of documents, and py_start, as shown in Table IV.

TABLE IV. AUTHORS' LOCAL IMPACT

Authors	h_index	g_index	m_index	TC	NP	PY_start
FØLSTAD A	6	9	1.200	286	9	2020
ZHU Y	6	10	2.000	235	10	2022
MOU J	5	5	1.667	305	5	2022
CHEN Q	4	5	1.333	145	5	2022
CHEN Y	4	4	1.000	201	4	2021
HOBERT S	4	5	0.800	156	5	2020
JEON J	4	7	2.000	220	7	2023
KIM H	4	5	0.800	146	5	2020
LI Y	4	5	2.000	34	7	2023
LIU Y	4	5	1.333	145	5	2022

Table IV provides an analysis of the top authors in chatbot-related research, based on bibliometric indices. FØLSTAD A and ZHU Y demonstrate the highest h-index scores (6), indicating consistent citation impact, with ZHU Y leading in g-index (10) and m-index (2.000), showcasing sustained

productivity and influence since 2022. MOU J stands out with the highest total citations (305) despite having a lower g-index (5), reflecting impactful but fewer publications. JEON J and LI Y show notable m-index values (2.000), suggesting rapid citation growth relative to their research duration, particularly with recent publications in 2023. Meanwhile, CHEN Q and LIU Y exhibit balanced productivity and citation metrics, while HOBERT S and KIM H display consistent contributions since 2020. Overall, the data revealed diverse patterns of research impact, emphasizing both sustained contributions and emerging influencers in the field. Meanwhile, the influence of countries contributing to this field is dominated by the USA and China, as shown in the picture of the country scientific production below.

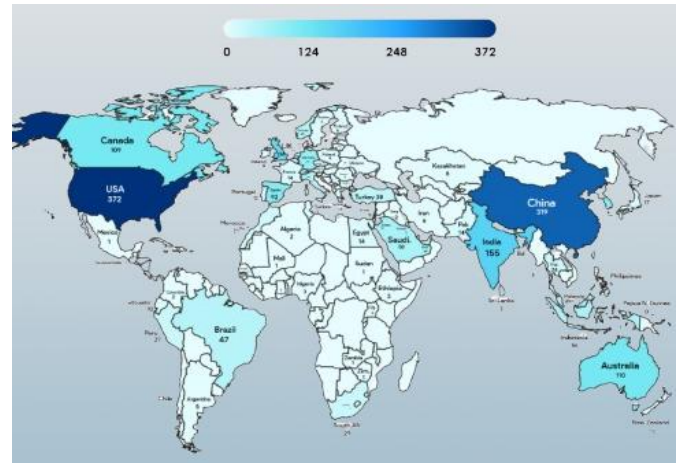


Fig. 3. Countries' scientific productions.

The Fig. 3. highlights the regional distribution of publications on chatbot research, revealing significant contributions from the USA (372), followed by China (319), and India (155), demonstrating their leadership in the field. Notable contributions also come from the UK (134), Australia (110), and Canada (109), indicating strong engagement from English-speaking countries. Asian nations such as South Korea (97), Indonesia (56), and Malaysia (49) also feature prominently, showing increasing research activity in the region. European countries, including Spain (92), Germany (77), and Italy (50), have contributed substantially, reflecting their growing interest in the domain. While regions such as the UAE (38), Brazil (47), and Turkey (38) indicate moderate contributions, countries with emerging research activities such as Morocco (21) and Thailand (24) highlight the global expansion of chatbot research. Smaller contributions from nations such as Ethiopia, Zambia, and Fiji emphasize the nascent but widespread adoption of chatbot research across diverse geographical areas.

B. Analysis of References

In bibliometric analysis, the purpose of analyzing references is to examine and scrutinize citations in academic publications, aiming to uncover insights into the intellectual structure and progression of a specific field of study. By analyzing references, bibliometrics can identify the most influential works and authors in a field, track their evolution, and identify key research trends and collaborations. This analysis can help researchers and policymakers to make informed decisions regarding the direction of future research. In this article, the analysis of

references is divided into two analyses: the most relevant word and evolution trend analysis.

Using Biblioshiny, word cloud visualization was performed to examine the keywords commonly used in articles related to chatbots. In this study, the most relevant words were identified by analyzing the keywords listed in the author's section of each article. The word cloud generated from the top 20 words shows a broad spectrum of crucial terms in the realm of chatbots, covering pivotal areas such as "artificial intelligence," "natural language processing," "conversational agents," and "chatgpt," among others. It encompasses an extensive range of concepts, including "anthropomorphism," "machine learning," and "covid-19," reflecting the multidisciplinary nature and depth of subjects related to chatbot technology and human-computer interaction. Fig 4 shows the most relevant keyword descriptions.

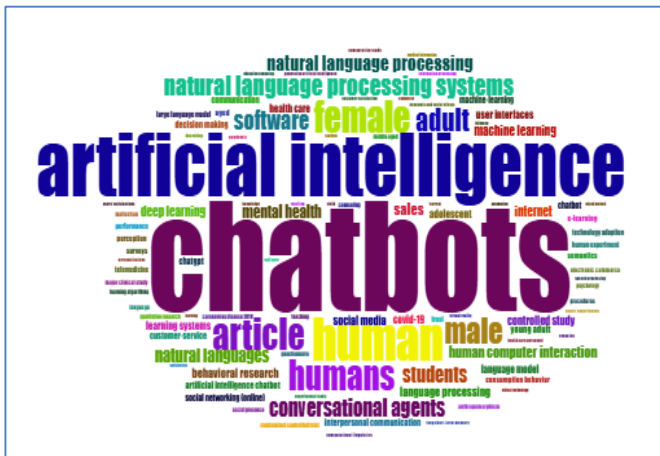


Fig. 4. Word cloud by keywords.

By analyzing the titles of articles from the authors, the words frequently used were Artificial Intelligence, AI chatbots, Mental Health, Customer Service, Intelligence Chatbots, Generative AI, Mixed Methods, and Natural Language. The frequency of occurrence in the analysis using N-Grams (Bigrams) is shown in Table V.

TABLE V. WORD FREQUENCY

Word	2020	2021	2022	2023	2024
Artificial Intelligence	2	5	10	35	68
Ai Chatbots	1	2	10	14	101
Mental Health	1	4	11	17	26
Customer Service	0	2	4	11	19
Intelligence Chatbots	2	3	6	19	36
Generative Ai	0	0	0	1	16
Mixed Methods	0	1	4	6	16
Natural Language	1	2	5	10	18

This table illustrates the dynamic evolution of key research topics in chatbot-related studies from 2020 to 2024. "Artificial Intelligence" consistently leads the discourse, showing significant growth from 2 mentions in 2020 to 68 in 2024, reflecting its foundational role in chatbot development. Similarly, "AI Chatbots" demonstrates exponential growth,

particularly in 2024, with a sharp rise to 101 mentions, highlighting their increasing adoption and application. Emerging topics like "Mental Health" and "Customer Service" reveal a steady increase, emphasizing the diversification of chatbot applications in addressing user needs. "Intelligence Chatbots" and "Natural Language" also exhibited consistent growth, underlining advancements in chatbot sophistication and linguistic capabilities. Notably, "Generative AI" which appeared only recently, shows substantial growth in 2024, indicating a shift towards innovative frameworks and methodologies in chatbot research. These trends underscore the expanding scope and interdisciplinary nature of this field.

Meanwhile, the word "framework" is in the 31st position as the most frequently appearing word, and its occurrence has increased from year to year, two times in 2020, six times in 2021, 12 times in 2022, 18 times in 2023, and 33 times in 2024. Several types of framework have been used in chatbot applications. These include RASA, Microsoft Bot Framework, Google Dialogflow, Telegram, Facebook, Twitter, Whatsapp, Wit.Ai, IBM Watson, Line, BotPress and Streamlit. Table VI lists some frameworks used for chatbot development in this study.

TABLE VI. FRAMEWORK OF DEVELOPMENT CHATBOTS

Framework	Authors
RASA	Fauzia[18], Windiatmoko[19]
Microsoft Bot	M. Cont & A. Ciupe[20], L. Zhou[21]
Google Dialogflow	S. Valtolina, Jr. [22], Villanueva G.R.[23]
Facebook	Aina, Pashev[24] & Gaftandzhieva[25]
Wit.ai	Li C. [26], Chu E [27]
Twitter	Y. M. Çetinkaya[28], E. Alothali[29]
Telegram	Y. Wahyuni[30], W. Santoso[31]
Whatsapp	Mash[32], Paschetto[33]
IBM Watson	R. J. Moore[34], C. V. M. Rocha[35]
Line	AI Rasyid [36]
Botpress	Macanu B[37] ,
Streamlit	Kiangala K [38], Kothari S [39]

Interestingly, Streamlit was used in the 2023 and 2024 studies, where it was used as a generative AI chatbot using a large language model. The analysis and interpretation of research related to mental health dominate the health sector. Meanwhile, e-learning dominates the education sector and sales dominate the business sector. The top six fields where chatbots were developed are Health, Education, University, Financial, Industry and Tourism. Thematic evolution analysis was used to analyze chatbot trends in this study. Thematic evolution analysis was performed based on a co-word network and clustering. Using these diagrams, it can be shown that chatbot research changes yearly. The existing keywords indicate the direction of research changes. Fig. 5 shows a thematic evolution diagram of the Keyword Plus field.

Thematic analysis revealed a dynamic evolution in chatbot research trends from 2020 to 2024. During 2020–2022, broad foundational themes such as "AI" and "NLP" were central,

representing the underlying technologies for chatbot development. As research advanced, a notable shift occurred towards applied themes in 2023–2024, such as "e-commerce," "human-computer interaction," and refined chatbot designs. For example, the transition from "AI-2020-2022" to "chatbot-2023-2024" underscores the growing emphasis on creating specialized systems tailored to industry-specific needs.

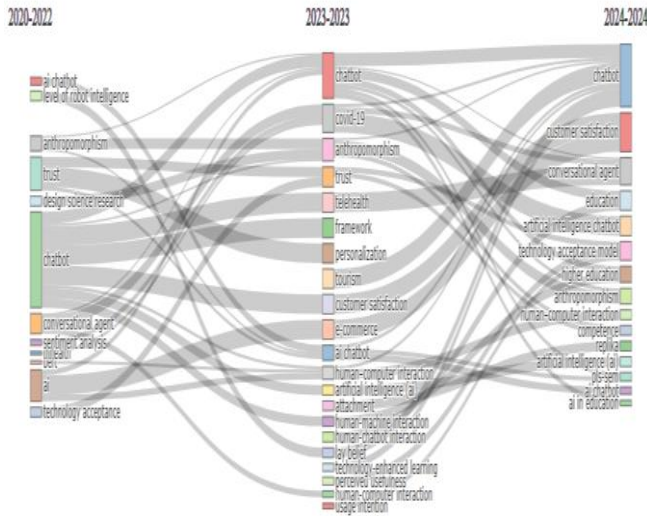


Fig. 5. Thematic evolution.

From 2023–2024, the emergence of "e-commerce" chatbots (weighted inclusion index of 1.00) highlights their importance in enhancing online retail and customer interaction. Similarly, themes like "human-computer interaction" reflect an increasing focus on usability and user satisfaction, as chatbots evolve to provide more intuitive and engaging experiences. Additionally, "AI chatbot," which persisted across both periods, illustrates continuous efforts to improve conversational models, ensuring their adaptability and effectiveness in addressing real-world challenges.

The stability indices across themes revealed sustained interest in specific areas. While "e-commerce" chatbots exhibit relatively high stability (0.11), foundational themes such as "AI" and "NLP" show lower stability, reflecting their transformation into specialized applications. By 2024, research trends will clearly demonstrate a movement from broad theoretical discussions to practical implementations, with chatbots playing critical roles in domains such as retail, customer service, and user interaction design. This evolution highlights the increasing maturity and impact of chatbot technologies in various sectors.

V. DISCUSSION

Research Question 1. The analysis of prolific authors and their contributions to chatbot research revealed significant growth and diversification in the field from 2020 to 2024. Annual scientific production has increased substantially, with article outputs rising from 60 in 2020 to 396 in 2024, demonstrating an accelerating interest and investment in chatbot studies. Citation trends, as presented in Table III, highlight declining mean total citations per year over time, suggesting that while publication volume has grown, the citation impact of

individual articles may have been diluted because of the field's rapid expansion. As detailed in Table IV, the key contributors include FØLSTAD A and ZHU Y, which lead to h-index scores, reflecting a consistent influence. At the same time, ZHU Y and JEON J exhibited high m-index values, signifying rapid citation growth in recent studies. MOU J achieved the highest total citations (305), indicating a significant impact despite fewer publications. Emerging influencers such as JEON J and LI Y demonstrate rapid citation accumulation, with research starting in 2023. Additionally, country-level data underscores the dominance of the USA and China in scientific production, positioning them as central players in advancing chatbot research. Together, these findings highlight both the maturation of established contributors and the emergence of new voices in the evolving chatbot research landscape.

Research Question 2. The analysis of chatbot research activity across different countries and regions revealed a rapidly expanding global field, with a remarkable annual growth rate of 60.28% from 2010 to 2024, culminating in 955 publications. The research is highly collaborative, with an average of 3.91 co-authors per document, and 27.64% involving international co-authorships, underscoring its global nature. The USA has 372 publications, followed by China (319), and India (155), reflecting their dominance in advancing chatbot technologies. Significant contributions from the UK (134), Australia (110), and Canada (109) highlight the strong engagement of English-speaking nations. Asian countries, such as South Korea (97), Indonesia (56), and Malaysia (49), also show robust participation, indicating the region's growing focus on chatbot research. European nations, including Spain (92), Germany (77), and Italy (50), have contributed substantially, showcasing their increasing interest in this domain. While countries such as Brazil (47), Turkey (38), and the UAE (38) reflect moderate engagement, emerging players, such as Morocco (21) and Thailand (24), highlight the global diffusion of chatbot research. Small contributions from nations such as Ethiopia, Zambia, and Fiji demonstrate the field's expansion to diverse and underrepresented regions, marking a promising trajectory for future research collaboration and development worldwide.

Research Question 3. Utilizing R Biblioshiny, the study yielded significant findings regarding the dominant keywords in articles focused on chatbots through the creation of word cloud visualizations. The analysis uncovered a wide range of essential terms that influence the development of chatbot technology based on the keywords provided by the authors. The word cloud consists of the top terms that emphasize essential topics vital to chatbots, such as "artificial intelligence," "natural language processing systems," "conversational agents," and "machine learning." This provides a complete overview of the essential principles of the field. Furthermore, the incorporation of words such as "semantics," "reinforcement learning," and "user interfaces" user interfaces emphasized the interdisciplinary character of chatbot technology and its convergence with human-computer interactions. These results are different from those found by previous researchers, where the keywords that are widely used are natural processing language, big data, learning algorithm, language learning system, and user experience [16].

Fig. 4 illustrates the visualization that effectively displays the essential characteristics of the discovered keywords, highlighting an extensive range of issues within the chatbot field. Moreover, an analysis of article titles revealed frequent terms such as “Artificial Intelligence”, “AI chatbots”, “Mental Health”, “Customer Service”, “Intelligence Chatbots”, “Generative AI”, “Mixed Methods” and “Natural Language”. The continued presence of frameworks in these titles indicates their enduring significance and prominence in conversations related to chatbots. The keyword frequency analysis of article titles revealed a notable pattern: the topic of Mental Health Chatbots was the most commonly discussed compared to other domains. This finding underscores a significant inclination towards using chatbot technology for mental health purposes, signifying the growing acknowledgment of the significance of technology in augmenting overall well-being. Additionally, it underscores the expanding apprehension over trust and acceptance in employing technology to address mental health concerns.

Research Question 4. The findings illustrate a clear evolution in chatbot research, transitioning from foundational themes to specialized applications between 2020 and 2024. In earlier years, the focus was predominantly on developing underlying technologies, such as artificial intelligence and natural language processing, which laid the groundwork for subsequent advancements. By 2023–2024, the emergence of themes like “e-commerce” and “human-computer interaction” demonstrates a shift towards practical and user-oriented applications. This transition reflects the increasing demand for chatbots in sectors that require tailored, intelligent, and efficient communication systems, particularly in e-commerce and service-oriented industries.

The persistence and refinement of themes such as “AI chatbot” also highlight the continued need for innovation in conversational models. While foundational topics showed lower stability indices owing to their evolution into specialized areas, emerging applications demonstrated higher stability, underscoring their growing importance. These findings suggest the maturation of the field, with research increasingly addressing the real-world challenges. Future studies could build on this trend by exploring the integration of advanced chatbot systems into more diverse sectors, emphasizing ethical considerations, user satisfaction, and the long-term sustainability of these technologies. In contrast, previous research observed that chatbot trends have evolved from basic rule-based systems to advanced machine-learning models [10]. However, this study found that the machine learning trend has shifted towards reinforcement learning and knowledge-based systems.

Research Question 5. In chatbot architecture, key elements include the Question Answering System, The Environment, The Front-end System, Traffic Server, and Custom Integrations. This discussion delves into the methods employed within the Question Answering System components and the framework applied to the front-end system. The utilization of AI, particularly deep-learning technology, has emerged as a prevailing trend in chatbot development. This aligns with the findings of Caldarini et al. [40], who emphasized the future of chatbot development through NLP techniques and utilization of deep learning technology. This correlation is in agreement with

the thematic evolution analysis conducted in this research, revealing a prominent theme centered on deep learning.

Various techniques have been employed in chatbot development, each of which has its own strengths and applications. Rule-based systems rely on predefined rules and patterns to generate responses [41]. Meanwhile Kandasamy et al. [42] stated that Information retrieval-based systems focus on retrieving relevant information from a large text corpus. On the other hand, Ngai et al. [43] asserted that Knowledge-based systems use structured knowledge bases or ontologies to provide answers. NLP-based systems utilize NLP techniques to understand and generate context-aware responses [44]. Machine learning-based systems leverage supervised, unsupervised, or reinforcement learning to train models capable of answering questions [45]. Thus, the choice of development technique depends on the chatbot's specific requirements and goals.

There have been substantial advancements in the broader field of NLP. NLP encompasses the interaction between computers and human language. Recent developments in deep learning, particularly with models such as transformers, for example, bidirectional encoder representations from transformers (BERT), have led to significant breakthroughs in NLP applications such as sentiment analysis, machine translation, and question-answering [46]. Furthermore, NLP research spans various domains, including question-answering systems (QAS), summarization, machine translation, speech recognition, and document classification, each contributing to the overall evolution of NLP [47].

The chatbot architecture plays a crucial role in their functionality and effectiveness. Hwang et al. [48] research provided an overview of chatbot architecture, with components such as NLU, Intents/Entitas, Message Generator, Knowledge-Based database, and Respon. This architecture serves as the foundation for the design and operation of chatbots, encompassing various components and modules that enable them to interact with users and respond.

The need for external references and context becomes evident in open-domain Question Answering (QA), where the chatbot is expected to answer a wide range of questions. Researchers, such as Chen et al. [49] have developed frameworks such as the Retriever-Reader Framework. This framework allows chatbots to retrieve relevant information from documents and generate responses, thus enabling them to answer questions that require broader contextual knowledge.

Another approach to open-domain QA is the Retriever Generator QA framework, often called Generative Question Answering. This framework combines document retrieval systems with general language models such as BERT and GPT, as demonstrated by Petroni et al. [50] demonstrated in their research. This approach aims to leverage the strengths of both retrieval and generative methods to improve chatbot performance by providing comprehensive answers. The third model framework, known as the Generative Language Model, is associated with closed-domain QA, where a model, such as T5 Framework developed by Robers et al. [51] can generate responses to questions without requiring additional information or context. This approach is particularly valuable in scenarios in

which chatbots are expected to provide precise and concise answers without external references.

In summary, chatbot architecture is essential for their functionality, and researchers have classified chatbot QA models into different categories based on their capabilities. In open-domain QA, frameworks such as the Retriever-Reader Framework and Retrieve Generative Framework have been developed to enhance chatbots' ability to provide comprehensive responses. Closed-domain QA models, such as the Generative Language Model, excel in independently generating answers, making them suitable for scenarios where external references are not required. These developments in chatbot architectures and models contribute to the continuous improvement of chatbot performance and applicability in various domains.

VI. FINDINGS

Growth and Diversification in Chatbot Research: The field of chatbot research has experienced significant growth from 2020 to 2024, with annual publication output increasing from 60 in 2020 to 396 in 2024. Despite this growth, a decline in mean total citations per article was observed, indicating the potential dilution of individual article impacts due to the field's rapid expansion. Prominent authors, such as FØLSTAD A and ZHU Y, demonstrated sustained influence. Simultaneously, emerging contributors such as JEON J and LI Y exhibited rapid citation accumulation, highlighting the field's dynamic and evolving nature.

Global Research Collaboration and Regional Participation: Chatbot research has high collaboration rates, with an average of 3.91 co-authors per document, and 27.64% of papers involve international partnerships. The USA and China led global contributions, producing the highest number of publications, while countries such as India, the UK, and Australia also made substantial contributions. Emerging participation from Southeast Asia and Africa, particularly Indonesia, Malaysia, Morocco, and Ethiopia, signifies an expanding and inclusive research landscape.

Thematic Evolution and Practical Applications: Thematic analysis revealed a transition from foundational technologies, such as AI and NLP, to specialized applications in domains like "e-commerce" and "mental health chatbots" by 2024. Mental health chatbots have emerged as a critical focus area, reflecting the growing societal interest in leveraging chatbot technology for well-being. These findings underscore a shift towards practical, user-centered chatbot applications that address diverse real-world needs.

Advancements in Chatbot Architectures: The development of chatbot architectures has progressed significantly, incorporating sophisticated frameworks such as retriever-reader systems, generative QA models, and closed-domain QA models. These innovations enhance chatbot performance in open domain and domain-specific applications. The integration of reinforcement learning, user-interface design, and deep learning technologies demonstrated the field's interdisciplinary nature and trajectory towards advanced conversational capabilities.

Emerging Trends in Keywords and Research Focus: Analyzing keyword trends revealed a focus on topics like "artificial intelligence," "natural language processing," and

"machine learning," alongside newer themes such as "semantics," "user interfaces," and "reinforcement learning." Mental health applications, customer service, and generative AI are prominent themes, highlighting the diverse and evolving priorities of the field. This progression reflects the broadening scope and deeper specialization of chatbot research.

VII. CONCLUSION

This study provides a comprehensive examination of the evolution and diversification of chatbot research from 2020 to 2024, highlighting the rapid growth and global collaboration of the field. The rising annual publications and citations have significantly shaped this dynamic domain, as evidenced by prolific contributors and emerging researchers. Notable scholars such as FØLSTAD A and ZHU Y demonstrate sustained influence, while newer contributors such as JEON J exhibit rapid citation growth, reflecting fresh perspectives and innovative approaches. Geographically, countries like the USA and China dominate research output, whereas emerging contributions from regions such as Southeast Asia and Africa underline the global expansion and inclusivity of chatbot studies.

In addition to analyzing global research activities, this study highlights emerging trends and technological advancements. Themes such as "e-commerce" and "mental health chatbots" underscore the practical applications of chatbot technology, while the evolution of architectural frameworks illustrates advancements in conversational capabilities. The findings emphasize a transition towards sophisticated and user-centered applications from foundational AI technologies to specialized Generative AI. This trajectory provides promising opportunities for future research to address societal challenges, enhance user satisfaction, and expand chatbot utility across diverse domains. These insights collectively contribute to a deeper understanding of the evolving chatbot landscape and its potential transformative impact.

This research provides several recommendations for future research related to SLR in this area. This study used references from Scopus as the primary data sources. Although the Scopus database is the largest, it does not necessarily encompass all research related to chatbots. It is hoped that future reviews can incorporate literature from the WoS and PubMed databases to expand the findings on chatbots and achieve more complete results. This study focused on chatbot articles, including natural language processing, artificial intelligence, machine learning, deep learning, and neural networks. It is challenging to apply these concepts to existing subfields. Future research should focus on specific sub-fields, such as AI. It must also focus on existing health, education, industry, social, and political sectors.

ACKNOWLEDGMENT

We extend our sincere thanks to everyone who played a role in bringing this research to fruition. In particular, we are grateful to Universiti Pendidikan Sultan Idris (UPSI) for supplying the essential resources and backing for this investigation. Our work has been significantly improved by the valuable input and direction from our peers and reviewers, for which we are deeply thankful. We also recognize the scholars and researchers whose prior work laid the groundwork for our analysis, and we

appreciate the spirit of cooperation within the academic sphere. Finally, we want to thank our loved ones for their constant support throughout the course of this research endeavor.

REFERENCES

- [1] J. Skrebeca, P. Kalniete, J. Goldbergs, L. Pitkevica, D. Tihomirova, and A. Romanovs, "Modern Development Trends of Chatbots Using Artificial Intelligence (AI)," in 2021 62nd International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS), Oct. 2021. doi: 10.1109/itms52826.2021.9615258.
- [2] S. Pandey and S. Sharma, "A comparative study of retrieval-based and generative-based chatbots using Deep Learning and Machine Learning," *Healthcare Analytics*, vol. 3, Nov. 2023, doi: 10.1016/j.health.2023.100198.
- [3] A. Bhagchandani and A. Nayak, "Deep Learning Based Chatbot Framework for Mental Health Therapy," in *Advances In Data And Information Sciences*, S. Tiwari, M. C. Trivedi, M. L. Kolhe, K. K. Mishra, and B. K. Singh, Eds., in *Lecture Notes in Networks and Systems*, vol. 318. Switzerland: Springer International Publishing AG, 2022, pp. 271–281. doi: 10.1007/978-981-16-5689-7_24.
- [4] S. H. Anwar, K. M. Abouaish, E. M. Matta, A. K. Farouq, A. A. Ahmed, and N. K. Negied, "Academic assistance chatbot-a comprehensive NLP and deep learning-based approaches," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 2, pp. 1042–1056, Feb. 2024, doi: 10.11591/ijeecs.v33.i2.pp1042-1056.
- [5] P. C. K. Hung, H. Demirkan, and S. C. Huang, "Editorial: Special Section on Services Computing Management for Artificial Intelligence and Machine Learning," Feb. 01, 2021, Institute of Electrical and Electronics Engineers Inc. doi: 10.1109/TEM.2020.3024363.
- [6] D.-M. Park, S.-S. Jeong, and Y.-S. Seo, "Systematic Review on Chatbot Techniques and Applications," *Journal of Information Processing Systems*, vol. 18, no. 1, pp. 26–47, 2022, doi: 10.3745/JIPS.04.0232.
- [7] S. Hussain, O. Ameri Sianaki, and N. Ababneh, "A Survey on Conversational Agents/Chatbots Classification and Design Techniques," in *Advances in Intelligent Systems and Computing*, Springer Verlag, 2019, pp. 946–956. doi: 10.1007/978-3-030-15035-8_93.
- [8] A. Ramachandran, "User Adoption of Chatbots," *SSRN Electronic Journal*, 2019, doi: 10.2139/ssrn.3406997.
- [9] H. N. Io and C. B. Lee, "Chatbots and conversational agents: A bibliometric analysis," *IEEE International Conference on Industrial Engineering and Engineering Management*, vol. 2017-Decem, pp. 215–219, 2018, doi: 10.1109/IEEM.2017.8289883.
- [10] E. Adamopoulou and L. Moussiades, "Chatbots: History, technology, and applications," *Machine Learning with Applications*, vol. 2, p. 100006, Dec. 2020, doi: 10.1016/j.mlwa.2020.100006.
- [11] M. Pears and S. Konstantinidis, "Bibliometric Analysis of Chatbots in Health-Trend Shifts and Advancements in Artificial Intelligence for Personalized Conversational Agents," in *Studies in Health Technology and Informatics*, IOS Press BV, Jun. 2022, pp. 494–498. doi: 10.3233/SHTI220125.
- [12] M. L and S. Alur, "Mapping the Research Landscape of Chatbots, Conversational Agents, and Virtual Assistants in Business, Management, and Accounting: A Bibliometric Review," *Qubahan Academic Journal*, 2023, doi: <https://doi.org/10.58429/qaj.v3n4a252>.
- [13] M. Tanwar and H. V. Verma, "Scientific Mapping of Chatbot Literature: A Bibliometric Analysis," *International Journal of Mathematical, Engineering and Management Sciences*, vol. 9, no. 2, pp. 323–340, Apr. 2024, doi: 10.33889/IJMEMS.2024.9.2.017.
- [14] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *J Bus Res*, vol. 133, pp. 285–296, Sep. 2021, doi: 10.1016/j.jbusres.2021.04.070.
- [15] M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," *J Informetr*, vol. 11, no. 4, pp. 959–975, Nov. 2017, doi: 10.1016/j.joi.2017.08.007.
- [16] S. Agarwal, B. Agarwal, and R. Gupta, "Chatbots and virtual assistants: a bibliometric analysis," *Library Hi Tech*, vol. 40, no. 4, pp. 1013 – 1030, 2022, doi: 10.1108/LHT-09-2021-0330.
- [17] Z. Xia, N. Li, and X. Xu, "A Bibliometric Review of Analyzing the Intellectual Structure of the Knowledge Based on AI Chatbot Application from 2005-2022," *Journal of Information Systems Engineering and Management*, vol. 8, no. 1, 2023, doi: 10.55267/iadt.07.14428.
- [18] L. Fauzia, R. B. Hadiprakoso, and Girinoto, "Implementation of Chatbot on University Website Using RASA Framework," in 2021 4Th International Seminar On Research Of Information Technology Andintelligent System (ISRITI 2021), New York, USA: IEEE, 2020. doi: 10.1109/ISRITI54043.2021.9702821.
- [19] Y. Windiatmoko, A. F. Hidayatullah, and R. Rahmadi, "Developing FB Chatbot Based on Deep Learning Using RASA Framework for University Enquiries," *IOP Publishing Ltd*, vol. 1077, 2021, doi: 10.1088/1757-899X/1077/1/012060.
- [20] M. Cont, A. Ciupe, B. Orza, I. Cohut, and G. Nitu, "Career Counseling Chatbot using Microsoft Bot Frameworks," in 2021 4Th International Seminar On Research Of Information Technology Andintelligent Systems, I. Kallel, H. M. Kammoun, A. Akkari, and L. Hsairi, Eds., in *IEEE Global Engineering Education Conference*. New York, USA: IEEE, 2022, pp. 1387–1392. doi: 10.1109/EDUCON52537.2022.9766485.
- [21] L. Zhou, J. Gao, D. Li, and H. Y. Shum, "The design and implementation of xiaoice, an empathetic social chatbot," *Computational Linguistics*, vol. 46, no. 1, pp. 53–93, 2020, doi: 10.1162/COLI_a_00368.
- [22] S. Valtolina and L. Neri, "Visual design of dialogue flows for conversational interfaces," *Behaviour and Information Technology*, vol. 40, no. 10, pp. 1008–1023, 2021, doi: 10.1080/0144929X.2021.1918249.
- [23] G. R. Villanueva and T. Palaog, "Design architecture of FAQ chatbot for higher education institution," *Journal of Advanced Research in Dynamical and Control Systems*, vol. 12, no. 1 Special, pp. 189–196, 2020, doi: 10.5373/JARDCS/V12SP1/20201062.
- [24] S. Aina, S. D. Okegbile, P. Makanju, and A. I. Oluwaranti, "An architectural framework for facebook messenger chatbot enabled home appliance control system," *International Journal of Ambient Computing and Intelligence*, vol. 10, no. 2, pp. 18–33, 2019, doi: 10.4018/IJACI.2019040102.
- [25] G. Pashev and S. Gaftandzhieva, "Facebook Integrated Chatbot for Bulgarian Language Aiding Learning Content Delivery," *TEM Journal*, vol. 10, no. 3, pp. 1011–1015, 2021, doi: 10.18421/TEM103-01.
- [26] C. Li and H. J. Yang, "Bot-X: An AI-based virtual assistant for intelligent manufacturing," *Multiagent and Grid Systems*, vol. 17, no. 1, pp. 1–14, 2021, doi: 10.3233/MGS-210340.
- [27] E. T.-H. Chu and Z.-Z. Huang, "Dbos: A dialog-based object query system for hospital nurses," *Sensors (Switzerland)*, vol. 20, no. 22, pp. 1–15, 2020, doi: 10.3390/s20226639.
- [28] Y. M. Çetinkaya, İ. H. Toroslu, and H. Davulcu, "Developing a Twitter bot that can join a discussion using state-of-the-art architectures," *Soc Netw Anal Min*, vol. 10, no. 1, 2020, doi: 10.1007/s13278-020-00665-4.
- [29] E. Alothali, K. Hayawi, and H. Alashwal, "Hybrid feature selection approach to identify optimal features of profile metadata to detect social bots in Twitter," *Soc Netw Anal Min*, vol. 11, no. 1, 2021, doi: 10.1007/s13278-021-00786-4.
- [30] Y. Wahyuni, F. Ammar, and I. Anggraeni, "Application Of Pregnant Mom's Diet Based on Raspberry Pi Using Telegram Chatbot," *Journal of Applied Engineering and Technological Science*, vol. 4, no. 1, pp. 209–214, 2022, doi: 10.37385/jaets.v4i1.989.
- [31] W. Santoso, W. Nurjannah, M. Shudhuashar, A. T. Fadilah, M. D. Junas, and D. Handayani, "The Development of Telegram Bot Api to Maximize The Dissemination Process of Islamic Knowledge in 4.0 Era," *Jurnal Teknik Informatika*, vol. 15, no. 1, pp. 52–62, Jun. 2022, doi: 10.15408/jti.v15i1.24915.
- [32] R. Mash, D. Schouw, and A. E. Fischer, "Evaluating the Implementation of the GREAT4Diabetes WhatsApp Chatbot to Educate People with Type 2 Diabetes during the COVID-19 Pandemic: Convergent Mixed Methods Study," *JMIR Diabetes*, vol. 7, no. 2, 2022, doi: 10.2196/37882.

- [33] I. V. Pasquetto, E. Jahani, S. Atreja, and M. Baum, "Social Debunking of Misinformation on WhatsApp: The Case for Strong and In-group Ties," *Proc ACM Hum Comput Interact*, vol. 6, no. CSCW1, 2022, doi: 10.1145/3512964.
- [34] R. J. Moore, S. An, and G.-J. Ren, "The IBM natural conversation framework: a new paradigm for conversational UX design," *Hum Comput Interact*, 2022, doi: 10.1080/07370024.2022.2081571.
- [35] C. V. M. Rocha et al., "A Chatbot Solution for Self-Reading Energy Consumption via Chatting Applications," *Journal of Control, Automation and Electrical Systems*, vol. 33, no. 1, pp. 229–240, 2022, doi: 10.1007/s40313-021-00818-6.
- [36] M. Udin Harun Al Rasyid, S. Sukaridhoto, M. Iskandar Dzulqornain, and A. Rifa, "Integration of IoT and chatbot for aquaculture with natural language processing," *TELKOMNIKA Telecommunication, Computing, Electronics and Control*, vol. 18, no. 2, pp. 640–648, 2020, doi: 10.12928/TELKOMNIKA.v18i1.14788.
- [37] B.-C. Mocanu et al., "ODIN IVR-Interactive Solution for Emergency Calls Handling," *Applied Sciences (Switzerland)*, vol. 12, no. 21, 2022, doi: 10.3390/app122110844.
- [38] K. S. Kiangala and Z. Wang, "An experimental hybrid customized AI and generative AI chatbot human machine interface to improve a factory troubleshooting downtime in the context of Industry 5.0," *International Journal of Advanced Manufacturing Technology*, vol. 132, no. 5–6, pp. 2715 – 2733, 2024, doi: 10.1007/s00170-024-13492-0.
- [39] S. Kothari, P. Bagane, M. Mishra, S. Kulshrestha, Y. Asrani, and V. Maheswari, "CropGuard: Empowering Agriculture with AI driven Plant Disease Detection Chatbot," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 12s, pp. 530 – 537, 2024, [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85185654968&partnerID=40&md5=5f3945e30295d202c7f21f44d68ab569>
- [40] G. Caldarini, S. Jaf, and K. McGarry, "A Literature Survey of Recent Advances in Chatbots," *Information (Switzerland)*, vol. 13, no. 1, 2022, doi: 10.3390/info13010041.
- [41] M. Akour, S. Abufardeh, K. Magel, and Q. Al-Radaideh, "QArabPro: A Rule Based Question Answering System for Reading Comprehension Tests in Arabic," *Am J Appl Sci*, vol. 8, no. 6, pp. 652–661, 2011.
- [42] S. Kandasamy and A. K. Cherukuri, "Information retrieval for question answering system using knowledge based query reconstruction by adapted lesk and Latent Semantic Analysis," *International Journal of Computer Science and Applications*, Technomathematics Research Foundation, vol. 14, no. 2, pp. 31–46, 2017, [Online]. Available: <https://www.researchgate.net/publication/326173786>
- [43] E. W. T. Ngai, M. C. M. Lee, M. Luo, P. S. L. Chan, and T. Liang, "An intelligent knowledge-based chatbot for customer service," *Electron Commer Res Appl*, vol. 50, 2021, doi: 10.1016/j.elerap.2021.101098.
- [44] A. N. Mathew, V. Rohini, and J. Paulose, "NLP-based personal learning assistant for school education," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 5, pp. 4522–4530, 2021, doi: 10.11591/ijece.v11i5.pp4522-4530.
- [45] I.-C. Hsu and J.-D. Yu, "A medical Chatbot using machine learning and natural language understanding," *Multimed Tools Appl*, vol. 81, no. 17, pp. 23777–23799, 2022, doi: 10.1007/s11042-022-12820-4.
- [46] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." [Online]. Available: <https://github.com/tensorflow/tensor2tensor>
- [47] N. Capuano, S. Caballé, J. Conesa, and A. Greco, "Attention-based hierarchical recurrent neural networks for MOOC forum posts analysis," *J Ambient Intell Humaniz Comput*, vol. 12, no. 11, pp. 9977–9989, 2021, doi: 10.1007/s12652-020-02747-9.
- [48] M. H. Hwang, J. Shin, H. Seo, J. S. Im, and H. Cho, "KoRASA: Pipeline Optimization for Open-Source Korean Natural Language Understanding Framework Based on Deep Learning," *Mobile Information Systems*, vol. 2021, 2021, doi: 10.1155/2021/9987462.
- [49] D. Chen, A. Fisch, J. Weston, and A. Bordes, "Reading Wikipedia to Answer Open-Domain Questions," Mar. 2017, [Online]. Available: <http://arxiv.org/abs/1704.00051>
- [50] F. Petroni et al., "How Context Affects Language Models' Factual Predictions," May 2020, [Online]. Available: <http://arxiv.org/abs/2005.04611>
- [51] A. Roberts, C. Raffel, and N. Shazeer, "How Much Knowledge Can You Pack Into the Parameters of a Language Model?," Feb. 2020, [Online]. Available: <http://arxiv.org/abs/2002.08910>

Application of Collaborative Filtering Optimization Algorithm Based on Semantic Relationships in Interior Design

Kai Zhao¹, Lei Wang^{2*}

College of Fine Arts and Art Design, Henan Vocational University of Science and Technology, Zhoukou, 466000, China¹
School of Residential Environment and Design, Shijiazhuang Institute of Technology, Shijiazhuang, 050228, China²

Abstract—Due to the diversity of interior design, it is difficult for users to mine target data, so personalized recommendation systems for users are particularly important. Therefore, an optimized collaborative filtering recommendation system is proposed. Firstly, a random walk recommendation model based on category combination space is constructed, abandoning the traditional flat relationship connection and using Hasse diagram to achieve one-to-one mapping between items and types. The semantic relationship and distance are defined. Finally, a basic recommendation framework for random walks is established based on data such as jump behavior. Next, the potential semantic relationships between entities are explored, and a lightweight knowledge graph is proposed to define the social and explicit relationships between entities. Finally, the short-term features of the project are obtained using deep collaborative filtering technology, and a deep collaborative filtering temporal model based on semantic relationships is constructed. In subsequent validation, these experiments confirmed that under the vector dimension of 10, the average HR@K and NDCG@K were 6.9% and 12.9% higher than the other models. Therefore, the collaborative filtering recommendation model based on semantic relationships proposed in the study is reliable.

Keywords—*Semantic relationships; category combination space; random walks; collaborative filtering; temporal recommendations*

I. INTRODUCTION

The Internet has brought great influence to human society, which is not only a simple query tool, but also a method to tap the potential interests of users. This change is to adapt to the gradual increase in the amount of big data currently available, making it difficult for users to choose their preferences among the vast amount of information. At this time, the algorithm can build a recommendation model for users using previous preferences, and recommend relevant preference data and potential interest preference data for users [1]. Collaborative Filtering (CF) is the most commonly used recommendation algorithm at present. The key to this technology is searching for neighboring users and calculating similarity. However, the cold start and high computational dimension limit the improvement of recommendation performance. More researches focus on the introduced information. Project type is an important additional data, which can alleviate the sparse data. However, the limitation of its flattened organization is also one of the problems to be solved [2]. As a semantic network, knowledge mapping can construct semantic relations to connect entities,

which effectively solves the neglected explicit and implicit interactions between information in the traditional model [3]. In recent years, recommendation systems have made significant progress, from early CF algorithms to advanced technologies such as deep learning and graph neural networks. The accuracy and efficiency of recommendation systems continue to improve. However, traditional recommendation models often rely on flat relational connections and fail to fully explore the complex semantic relationships between items and types, resulting in insufficient recommendation accuracy and semantic understanding. Most studies often lack temporal considerations when dealing with dynamically changing user needs, making it difficult to adapt to real-time changes in user interests. Therefore, a deep CF recommendation system based on semantic relationship is proposed. It is expected to make significant contributions to the development of recommendation systems and provide more efficient and accurate tools for solving information overload problems. The innovation of this study lies in proposing a random walk recommendation model based on type combination space, which abandons the traditional flat relationship connection and uses Hasse diagram to achieve one-to-one mapping between items and types, providing a new perspective and idea for recommendation systems.

The research includes six sections. Firstly, the research status of recommendation system is introduced. Secondly, the designed CF recommendation system based on semantic relationship is described in Section III. Section IV verifies the recommendation model. Results and discussion is presented in Section V and finally, the paper is concluded in Section VI.

II. RELATED WORKS

In various current online platforms, personalized recommendation systems have gradually become a mainstream and necessary part. Gou L et al. believed that the marketing model in the communication field shifted towards socialization. Therefore, a recommendation model based on social analysis was proposed. Initially, a two-layer communication network for users and platforms was constructed through historical information, followed by a network of similar users and platforms using CF. Finally, a bipartite graph weighting method was used to achieve project recommendation, and the feasibility of the model was experimentally verified [4]. Wu et al. believed that knowledge maps had a good auxiliary effect on recommending a large amount of data content. Therefore, a

context aware algorithm based on Knowledge Graph (KG) was proposed, while gaining the advantages of both propagation and path-based technologies. The user preferences were represented through rules, and an automatic rule mining model was used in entity interaction. Finally, the performance of the model was further optimized through the local features in the neighborhood [5]. Yan et al. believed that the data density was the key factor affecting the CF recommendation performance. Therefore, a neighboring clustering method based on granular computing technology was introduced, and a CF optimization scheme based on coverage rough granular computing was proposed. The basic framework was built on user project scoring data, and local rough granularity sets were established through user preference thresholds, solving the data sparsity. The accuracy improvement in simulation experiments verified the reliability of the optimization model [6]. Miao et al. believed that the subjectivity of social networks has driven the improvement of their platforms. A recommendation algorithm based on user profiles was proposed and the user profile application model on short video platforms was described. These experiments confirmed that in four different platforms, the user satisfaction rate with recommended content was greater than 75% [7].

Liang et al. believed that the recommendation balance and trust of educational resources were generally poor. In response to this phenomenon, a trust relationship recommendation model was proposed. Firstly, support vector machines were used to classify educational resources and remove duplicate and useless data. Based on the remaining data, resource features were extracted. Finally, Kalman filtering was introduced to denoise and reduce the dimensionality of the features, constructing the recommendation model. These experimental data confirmed

that the equilibrium degree reached 96 [8]. Qi et al. proposed to apply graph neural networks to recommendation algorithms, using user historical data and second-order social data to learn user project feature representation. Multiple graph attention network modules were utilized to construct the model. Finally, simulation experiments were conducted to verify the effectiveness of the model [9]. Wang et al. believed that intelligent English has solved a large number of educational challenges. The Internet of Things serves as the technical foundation in this field and can effectively collect and manage information. An attention mechanism module was introduced to propose an educational resource recommendation algorithm, and a deep collaborative recommendation model was constructed. The effectiveness of the method was verified through experiments [10]. The summary of the research on recommendation models mentioned above is shown in Table I.

In summary, although recommendation models have received widespread attention from researchers, current recommendation models still have problems such as not delving into the complex semantic relationships between items and types, and requiring a large amount of data and computing resources. Therefore, this study introduces Hasse diagram, lightweight KG, and deep CF techniques to comprehensively consider the semantic relationships between items and types, as well as the social and explicit implicit relationships between entities. A semantic-based optimized deep CF algorithm is proposed to improve the accuracy and robustness of the recommendation system. This study has significant advantages over previous research in terms of the depth of recommendation models, semantic relationship mining, temporal considerations, and generalization ability.

TABLE I. SUMMARY TABLE

Researcher	Method	Insufficient
Gou L	Recommendation Model Based on Social Analysis	Not fully considering the semantic relationship and distance between projects and types
Wu C	Context Aware Algorithm Based on Knowledge Graph	Not delving deeply into the social and explicit implicit relationships between entities
Yan HC	Collaborative Filtering Optimization Based on Covering Rough Particle Computation	Not considering the complex semantic relationship between projects and types
Miao R	Recommendation Algorithm Based on User Profile	Lack of in-depth exploration of the complex semantic relationships between projects and types
Liang X	Trust Relationship Recommendation Model	Requires a large amount of data and computing resources
Qi W	Recommendation Algorithm Based on Graph Neural Network	Strong dependence on secondary social data
Wang F	Deep Collaboration Recommendation Model	Failed to fully explore the complex relationship between users and projects

III. APPLICATION OF CF OPTIMIZATION ALGORITHM INTEGRATING SEMANTIC RELATIONSHIPS IN INTERIOR DESIGN

The interior design data are relatively rich, making it difficult for users to accurately search for more types of preferences and to mine potential preferences. Therefore, it is necessary to optimize it to achieve more accurate user recommendations. The recommendation algorithm based on project types is a common personalized strategy. Traditional model building methods based on type similarity and preference types often overlook the relationship structure between types, which hinders the recommendation accuracy

[11-12]. This study proposes a random walk based recommendation model based on Category Combination Space (CCS), and introduces a temporal model based on semantic relationships to achieve final recommendations through deep CF.

A. A Random Walk Recommendation Algorithm Defined by Semantic Relationship

CCS is a collection of types contained in multiple projects, which Hasse diagrams to map projects and types one by one, including various relational structures such as up and down, same layer, and skip. Eq. (1) represents each element in space.

$$\begin{cases} U = \{u_1, u_2, \dots, u_m\} \\ I = \{i_1, i_2, \dots, i_m\} \\ C = \{c_1, c_2, \dots, c_m\} \end{cases} \quad (1)$$

In Eq. (1), $U/I/C$ represent a collection of users, projects, and types, respectively. The relationship between projects and types is one-to-many. CCS considers all types corresponding to each project as a whole and obtains a one-to-one correspondence diagram, as shown in Fig. 1.

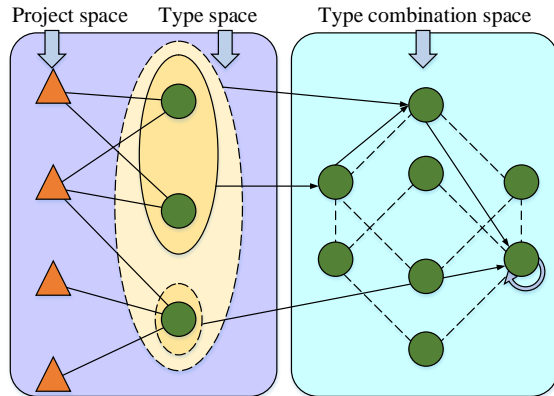


Fig. 1. Category combination space transformation structure.

CCS is essentially a partially ordered set of project types, belonging to an efficient Hasse graph. In actual operation, only some elements in the space are applied, and their project set is a subset of the partially ordered set. CCS achieves user preference mining through the jump sequence formed by user browsing behavior on nodes and the semantic relationship between jump nodes. Among them, the semantic relationships of browsing behavior include four categories, corresponding to the reduction, expansion, stability, and jump of user interests [13]. The qualitative description of relationships lays a solid foundation for mining dynamic preferences, while semantic distance further describes the preference change. This can be analyzed from two aspects: the true distance on the Hasse diagram and the changes in the number of basic and common elements [14]. Eq. (2) is the minimum distance between Hasse graph types.

$$d_L(\tilde{c}_i, \tilde{c}_j) = 2|\tilde{c}_i \cup \tilde{c}_j| - |\tilde{c}_i| - |\tilde{c}_j| \quad (2)$$

In Eq. (2), d_L is the link distance. This distance value represents the hierarchical span between types. If the span is small, the corresponding interests are stable. $(\tilde{c}_i, \tilde{c}_j)$ represents different types of combinations. Eq. (3) is the preference distance.

$$d_p(\tilde{c}_i / \tilde{c}_j) = \begin{cases} 0 & \tilde{c}_i = \tilde{c}_j \\ \frac{1}{|C|} & \tilde{c}_i \succ \tilde{c}_j \\ |\tilde{c}_j - \tilde{c}_i| & \tilde{c}_i \prec \tilde{c}_j \\ |C| + d_L(\tilde{c}_i / \tilde{c}_j) & \text{other} \end{cases} \quad (3)$$

The preference distance represents the changes in type combinations. There are four situations: the combination remains unchanged, the combination becomes larger, the combination becomes smaller, and the combination is located in different projects [15]. The last one has the highest preference distance and the most significant change in interest. User browsing graph $G = (\tilde{C}, E, A^E)$ is introduced to collect semantic relationships, distance sizes, and temporal features between types. \tilde{C}, E, A^E represent the vertex, edge, and edge feature matrix of the graph, respectively. PageRank based on random walk was introduced to construct an interest preference model. The core idea of this algorithm is to evaluate the importance of a page based on its quantity and quality of links in Fig. 2.

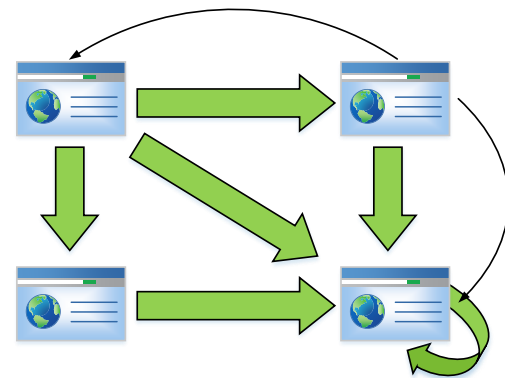


Fig. 2. Evaluation chart of the number of webpage links.

The number of times a webpage is linked is proportional to its importance, and its corresponding PageRank will also be improved. This transition probability design is the key to this method. The edge features of user browsing graphs contain a large amount of data such as node semantic relationships and preference distances [16]. The transfer matrix needs to transfer these data into probability, as shown in Eq. (4).

$$p_{ij}(\omega) = \frac{\omega^T a_{ij}^E}{\sum_j \omega^T a_{ij}^E} = \frac{\sum_k \omega_k a_{ijk}^E}{\sum_j \sum_k \omega_k a_{ijk}^E} \quad (4)$$

In Eq. (4), $p_{ij}(\omega)$ represents each element in the transition matrix. a_{ij}^E represents the edge feature vector. ω represents the weight of each edge. k represents the number of neighbors. The preference vector is the key to personalized random walk, as shown in Eq. (5).

$$q_u(i) = \frac{b_u(i)}{\sum_{k=1} |C| b_u(k)} \quad (5)$$

In Eq. (5), b_u represents the browsing vector of user u . The objective function can be obtained through the iterative process, as shown in Eq. (6).

$$\min_{\omega, \pi} (\omega^{(s)}, \pi^{(s)}) = \left\| \pi^{(s+1)} - \pi^{(s)} \right\|^2 \quad (6)$$

In Eq. (6), $\pi = (\pi_1, \pi_2, \dots, \pi_n)^T$ represents the scoring vector of the type combination. $\pi^{(s+1)}$ represents the iteration of a random walk. $\|\pi^{(s+1)} - \pi^{(s)}\|^2$ represents the loss function, and $\|\pi^{(s)}\| = 1$. The study introduces gradient descent method to minimize the objective function, while using cosine similarity to compare user similarity. Fig. 3 shows the overall CCS-based random walk algorithm.

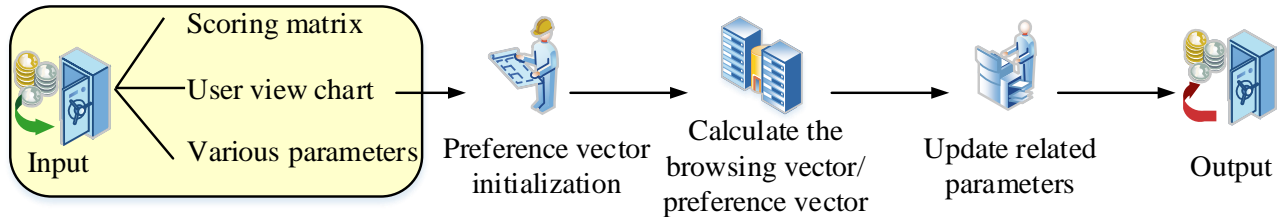


Fig. 3. Random walk algorithm model based on type combination space.

B. Construction of Semantic Relationship Temporal Recommendation Model Based on Deep CF

KG based on semantic relationships can extend project data to mine deep relationships between project types. However, this also means that a large number of databases are required as support. Due to the limitations of mature and publicly available databases and the large volume of training features, the feasibility of actual operation is low. Moreover, using only semantic relationships as predictive information cannot mine the interactions between user items [17]. Therefore, a timing lightweight KG is proposed in Fig. 4.

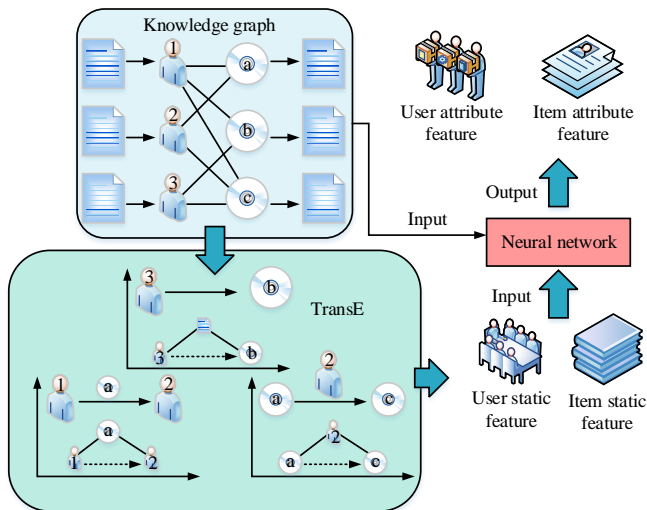


Fig. 4. Lightweight knowledge graph structure based on time series recommendation.

The model represents the explicit and implicit interactions between entities, users, and projects in KG by defining semantic relationships. The user rating of the project will create an edge between the two, and the attribute features of the two are linked to the corresponding entities [18]. The static attribute refers to the semantic feature representation vector of the two, which is obtained by associating the entity neighborhood with

According to Fig. 3, this model initializes the user preference vector through information such as rating matrix, user browsing graph, and weight parameters. Personalized preferences for each user are calculated based on their browsing data. Subsequently, the transfer matrix, objective function, and weight values are updated. When the scoring vector matches the iterative function, the final scoring prediction matrix is output.

the entity's first and second order neighbors. The attribute triplets of both are defined, including rating behavior, user relationships, project relationships, and attribute characteristics. Eq. (7) is the learning optimization probability of the rating behavior triplet.

$$P(u, r, v) = \sum_{(u, r, v^+) \in KG} \sum_{(u, r, v^-) \in KG^-} \sigma(g(u, r, v^+) - g(u, r, v^-)) \quad (7)$$

In Eq. (7), $\sigma(x) = 1/(1 + \exp(x))$ represents the sigmoid function. $g(\square)$ represents the energy function. The optimization probability $P(u_i, v, u_j)$ for user relationship triplets and project relationship triplets $P(v_i, u, v_j)$ are the same as Eq. (7). Eq. (8) is the energy function.

$$g(u, r, v) = \|u + r - v\|_{L_1/L_2} + b_1 \quad (8)$$

In Eq. (8), b_1 represents the bias constant. The attribute relationship between users and projects is essentially a bandit problem. Based on the representation vector, the neural network is selected for training. The optimization probabilities of user and project attribute triplets are expressed as $P(u, a, e) / P(v, a, e)$, and their definitions are similar. Eq. (9) is the calculation of the former.

$$P(u, a, e) = \sum_{(u, a, e^+) \in KG} \sum_{(u, a, e^-) \in KG^-} \sigma(h(u, a, e^+) - h(u, a, e^-)) \quad (9)$$

In Eq. (9), $h(\square)$ represents the classification function expressed by in Eq. (10).

$$h(u, a, e) = \|f(uW_a + b_a) - e_{ae}\|_{L_1/L_2} + b_2 \quad (10)$$

In Eq. (10), b_a, b_2 represent the learning parameter and bias constant, respectively. $f(\square)$ represents a nonlinear transformation function. e_{ae} represents the vector of attribute a . Due to the changing interests of users, popular elements can affect recommendation results, while static features do not change over time [19]. Therefore, short-term features of entities

are introduced for improvement. Model features can include static and attribute features based on KG semantic relationships, as well as short-term preference features generated by short-term browsing. The latter is implemented through Long Short-Term Memory (LSTM), which extracts popular items from the current browsing set. Taking the static and attribute features of the project as input, learning effectively reduces the computational pressure of the model in the attention mechanism. Finally, the recommendation is achieved through the long-term and short-term feature connections between entities in a multi-layer perception, as displayed in Fig. 5.

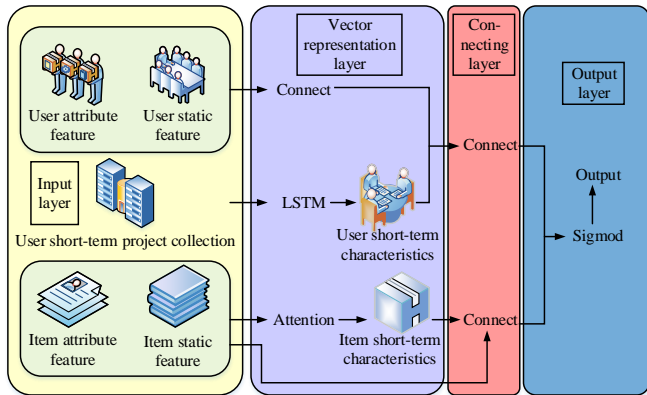


Fig. 5. LSTM recommendation model based on deep collaborative filtering.

The popularity of projects varies. A high level of popularity indicates that the project has a significant impact and requires learning through attention mechanisms. On the basis of maintaining sequential information, the attention mechanism extracts element relationships. The nearly 1-hour browsing sequence is used as input to match the entire project. The specific input data is the attributes and static features of the project. The weighted sum of each sequence is output, and Eq. (11) is the weight matrix.

$$T_i^t = z^T \tanh(W_c c_i + W_y y_i) \quad (11)$$

In Eq. (11), z^T, W_c represent vectors and matrices, respectively. W_y represents the learning parameter. c_i represents the training program at time t . y_i represents the i -th input item. The weight matrix also needs to be normalized in softmax to obtain S_i^t . The final short-term feature V_s needs to be output after iteration in Eq. (12).

$$\begin{cases} S_i^t = \text{soft max}(T_i^t) \\ V_s^t = \sum S_i^t y_i \end{cases} \quad (12)$$

The key to achieving the final recommendation lies in the model-based CF recommendation algorithm. It filters similar preferences in the neighborhood and recommends them to the target user through rating interaction between entities [20]. A multi-layer perception can perform deep CF recommendation, which combines to obtain diverse features in Eq. (13).

$$\begin{cases} q_1 = \phi_1(U_{s,a,r}, V_{s,a,r}) = \begin{bmatrix} U_{s,a,r} \\ V_{s,a,r} \end{bmatrix} \\ \phi_2(q_1) = a_2(w_2^T q_1 + b_2) \\ \dots \\ \phi_l(q_{l-1}) = a_l(w_l^T q_{l-1} + b_l) \\ \hat{y}_{uv} = \sigma(h^T \phi_l(q_{l-1})) \end{cases} \quad (13)$$

In Eq. (13), $U_{s,a,r}$ represents the combination of user dynamic preference U_s , attribute U_a , and static feature U_r . $V_{s,a,r}$ represents the union of corresponding characteristics of the project. w_x, b_x, a_x represent the weight matrix, bias vector, and ReLU activation function of the x -th layer perception, respectively. \hat{y}_{uv} represents the probability of interaction between entities. The model treats y_{uv} as a label. When there is already an association between the user and the project, the value is 1. On the contrary, it is 0. The probability range of entity interaction after training is [0, 1]. Eq. (14) is the final training objective function.

$$p(y, y^- | \Theta_f) = \prod_{(u,v) \in y} \hat{y}_{uv} \prod_{(u,v) \in y^-} (1 - \hat{y}_{uv}) \quad (14)$$

In Eq. (14), $p(y, y^- | \Theta_f)$ represents the objective function obtained through the probability function. Finally, a negative logarithmic likelihood loss function is obtained to minimize the optimization results in Eq. (15).

$$L = - \sum_{(u,v) \in y} \log \hat{y}_{uv} - \sum_{(u,v) \in y^-} \log(1 - \hat{y}_{uv}) \quad (15)$$

In Eq. (15), y^- represents negative instances, which are randomly generated by non-interacting items in the iteration and control the sampling probability of positive instances.

IV. PERFORMANCE TESTING OF CF OPTIMIZATION ALGORITHM MODEL BASED ON SEMANTIC RELATIONSHIPS

To verify the comprehensive performance of the recommendation model, simulation experiments are conducted, including two parts. The first step is to verify the feasibility of the CCS-based random walk basic framework. Subsequently, further validation is conducted on deep CF model based on semantic relationship to understand its excellent recommendation performance.

A. Performance Verification of the Basic Framework of a CCS Based-Random Walk

The study first validated the Optimized Random Walk (ORW) recommendation framework. Table II shows the experimental environment and parameters.

The experiment selected User-based CF (UCF), User-based CF by Genre Correlation (UBGC), Category Hierarchy Latent Factor Model (CHLF), and Genre to Classification model (GENC) for comparison with ORW. Firstly, the proximity

number of UCF was analyzed using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) indicators in Fig. 6.

Fig. 6 shows the error impact of proximity number on UCF, respectively. In both MAE and RMSE, UCF tended to gradually decrease with the potential number, and there was a certain rebound phenomenon in the later stage. When the proximity number was 45, the MAE reached a minimum of

0.615. When the proximity number was 65, the RMSE reached a minimum of 0.785. Therefore, in subsequent experiments, the median value of 55 was chosen as proximity value. In further comparative experiments on various algorithms, precision, recall rate, F1 index, and Area Under Curve (AUC) index are selected for analysis. Conditions with 10 and 20 recommended lists belonging to the test set (i.e. different vector dimensions K) were selected for comparison, as shown in Table III.

TABLE II. EXPERIMENTAL ENVIRONMENT AND PARAMETER SETTINGS

Name		Parameter/Settings
Datasets		MovieLens: 1M
Data set total	User	6040
	Item	3952
	Review	1000209
Data sparsity rate		95.8%
Item type quantity		18
Number of combination types		301(Basic type: 6)
Verification method		5 fold cross verification
Training set: Test set		8:2

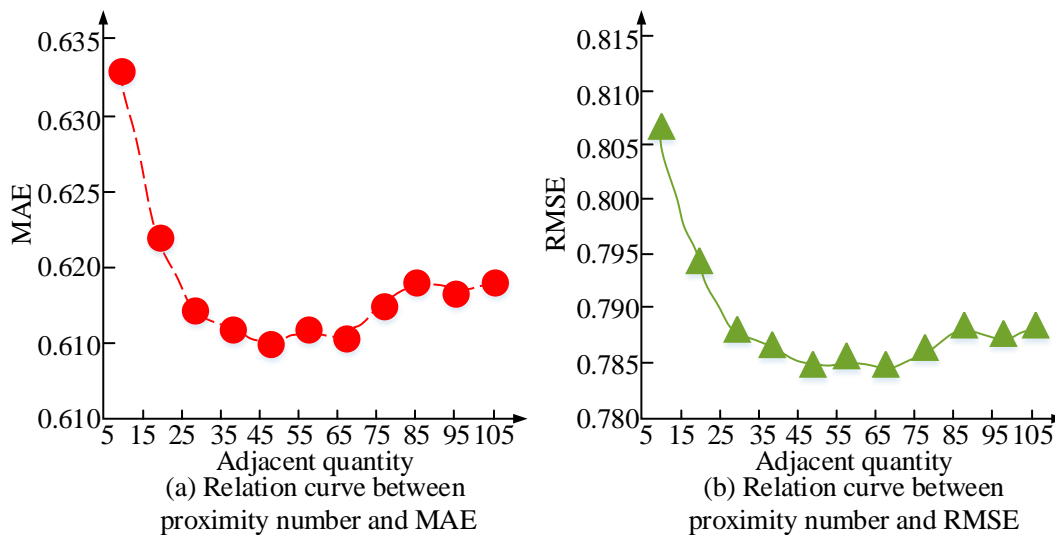


Fig. 6. Relation curve between proximity number and MAE/RMSE in UCF model.

According to Table III, the number of recommended lists that belong to the test set affects the performance of each model. For precision, all models reached their optimal values at Top@25. The difference between ORW and CHLF was relatively small, only 2.6% lower than the latter. Compared to UCF, UBG, and GENC, its precision has increased by 61.3%, 61.3%, and 77.3%, respectively. For the recall rate, all models reached their optimal value at Top@35. ORW performed the best, outperforming the other models by 57%, 59%, 6%, and 72.5%, respectively. F1 demonstrates a comprehensive performance of precision and recall. The performance difference between ORW and CHLF was minimal and almost negligible. Compared to UCF, UBG, and GENC, ORW was more than 55% higher. AUC represents the probability that the classifier outputs positive samples higher than negative samples,

which is a high value indicating good classification performance of the model. When it is not less than 0.5, it indicates that the classification effect is better than that of a random classifier. It also reached its optimal value at Top@35. The recommendation results of each model were superior to the random classifier, and the difference between the models was not more than 11%, indicating that these models were relatively excellent in recommendation performance. UCF, UBG, and GENC perform relatively poorly, with UCF performing the worst because the latter two use project-based associations, while the former uses user-based associations. ORW and CHLF are better because they both organize project types and obtain richer structural information. Fig. 7 shows the MAE and RMSE of each model.

From Fig. 7, CHLF performed the best, with MAE and

RMSE of 0.2 and 0.29, respectively. ORW was second only to CHLF, but its error was still much greater than CHLF, with MAE and RMSE of 0.6 and 0.79, respectively, an increase of 66.6% and 63.3%. In the comparison between ORW and other models, its MAE and RMSE were 25.9%/25.3% lower than UCF, 24.1%/45.6% lower than UBGC, and 31.6%/24.0% lower than GENC. CHLF outperforms all other models due to the used hierarchical type structures to construct entity-based implicit semantic models. Although the proposed model has CCS, it only performs a relatively simple similarity calculation on type numbering, which further demonstrates the necessity of introducing deep CF recommendations in the future.

TABLE III. COMPARISON OF THE COMPREHENSIVE PERFORMANCE OF EACH RECOMMENDED ALGORITHM

Index		ORW	UCF	UBGG	CHLF	GENC
Precision	Top@15	0.183	0.069	0.063	0.182	0.030
	Top@25	0.150	0.058	0.058	0.154	0.034
	Top@35	0.130	0.051	0.055	0.135	0.035
Recall	Top@15	0.075	0.029	0.024	0.065	0.012
	Top@25	0.117	0.049	0.044	0.108	0.026
	Top@35	0.149	0.064	0.061	0.140	0.041
F1	Top@15	0.107	0.041	0.035	0.095	0.017
	Top@25	0.132	0.053	0.050	0.127	0.029
	Top@35	0.139	0.057	0.058	0.138	0.037
AUC	Top@15	0.799	0.654	0.652	0.784	0.615
	Top@25	0.810	0.707	0.713	0.805	0.693
	Top@35	0.820	0.733	0.745	0.818	0.750

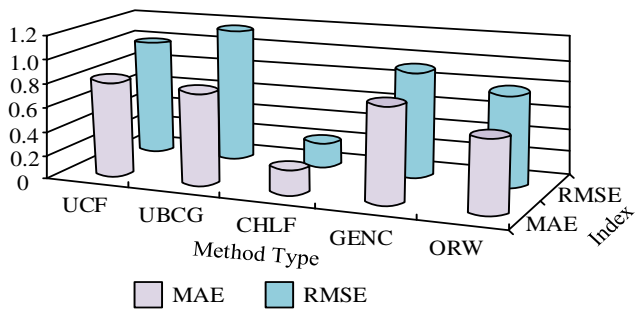


Fig. 7. Comparison of MAE/RMSE indicators of each model.

B. Performance Comparison Analysis of Semantic Relationship Temporal Recommendation Model Based on Deep CF

The study conducted performance validation analysis on the final model using a dataset. k in the triplet encoding Transe is 100. b_1, b_2 are 7 and -2, respectively. L_1 represents a regular term. The model operation framework and runtime platform are Keras and Python, respectively. To select the optimal Batch Size (S) and Learning Rate (R), this study used whether the test item was on the recommendation list and its position in the list as evaluation indicators, represented by HR and NDCG, respectively. The high value of both indicates excellent model performance. Table IV shows the experimental

results.

TABLE IV. HR/NDCG CHANGES UNDER DIFFERENT R/S SETTINGS

Index		HR@10	NDCG@10
R($\times 10^{-3}$)	0.1	0.681	0.405
	0.5	0.690	0.438
	1.0	0.700	0.451
	1.5	0.697	0.449
S	128	0.657	0.415
	256	0.701	0.452
	384	0.689	0.436
	512	0.674	0.428

In Table IV, when R increased, both HR and NDCG showed an initial increase followed by a slow decrease, with turning points of 0.001. Compared to the initial R of 0.0001, HR and NDCG were increased by 2.7% and 10.2%, respectively. Afterwards, HR and NDCG showed a gradually decreasing trend, but the decrease was smaller compared to the increase. In the testing of S, the change was the same as that of R, with a turning point of 256. Compared to the initial 128 batches, HR and NDCG were increased by 6.3% and 8.2%, respectively. As the S and R continue to increase, the proportion of test items in the recommendation list of the model changes relatively weakly, but the change in the position of the item in the list is relatively significant. When R=0.001 and S=256, the proportion of items in the recommendation list is higher, and their positions in the list are higher. Therefore, these parameters are selected as model settings for the experiment. CoupledCF, a Wide & Deep model based on logistic regression and feedforward deep neural networks, and a multi-layer perception NCF are compared with the Semantic Relationship Timing Recommendation model based on deep CF (SRT-DCF). SRT-DCF includes three types: user, item, and NCF. Fig. 8 shows the recommended results for TOP@10.

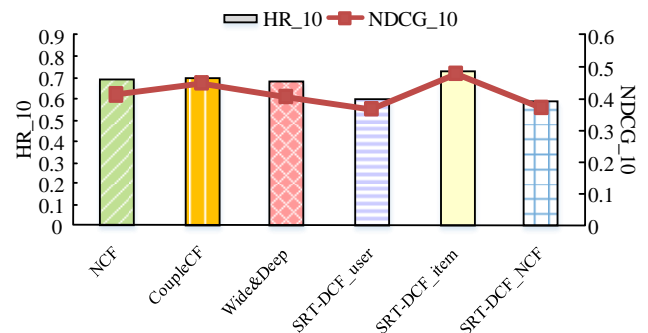


Fig. 8. Comparative analysis of HR_10/NDCG_10 performance of each model.

In Fig. 8, the difference among SRT-DCF_user, SRT-DCF_item, and SRT-DCF_NCF lies in the input of LSTM and the input of attention module. The inputs of SRT-DCF_user were the static and attribute features of recent projects learned from knowledge, as well as the onr-hot coding projects of the past hour. The inputs of SRT-DCF_item were onr-hot encoding item, as well as the static and attribute features of recent items

learned from knowledge. The inputs of SRT-DCF_NCF were static and attribute features of recent projects learned from knowledge. The combined performance of CoupledCF and SRT-DCF_item was the best, while the performance of SRT-DCF_user and SRT-DCF_NCF was the worst. The average values of HR₁₀ and NDCG₁₀ were 0.583 and 0.3705,

respectively. Compared to NCF and Wide&Deep models, the SRT-DCF_item had an average increase of 5.55% in HR₁₀ and 14.6% in NDCG₁₀. In summary, the research should select SRT-DCF_item as the final model. To further understand the impact of vector dimensions on various models, a comparative analysis was conducted in Fig. 9.

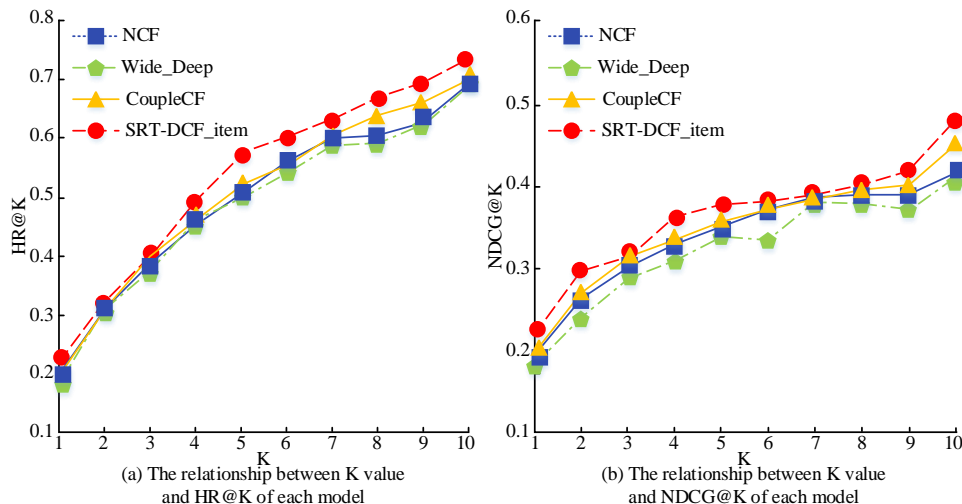


Fig. 9. Performance comparison of different models in different vector dimensions.

According to Fig. 9 (a), the HR@K change for each model was basically consistent, with the deviation node being K=5. At this time, the HR@K of the SRT-DCF_item model reached 0.583, which was 1.9%, 12.5%, and 1.8% higher than NCF, Wide&Deep, and CoupledCF models, respectively. When K was 10, the HR@K of other models was distributed around 0.67, while the HR@K of SRT-DCF_item reached 0.72, with an average increase of 6.9%. According to Fig. 9 (b), the NDCG@K variation was more tortuous. When K was 10, the values of each model were 0.493, 0.482, 0.409, and 0.397, respectively. Therefore, the average NDCG@K of SRT-DCF_item was 12.9% higher than that of other models. In summary, the proposed model has the best overall performance.

V. RESULTS AND DISCUSSION

To enhance the personalized recommendation function for users, a CF model based on semantic relationships is proposed. Firstly, a CCS based on random walk basic recommendation framework was constructed, and the semantic relationships between entities were defined. Then, the final deep CF temporal recommendation model was further constructed using semantic relationships. The recommendation algorithm proposed in study [7] mainly focuses on extracting and analyzing user features, and lacks in mining complex semantic relationships between items and types. This study achieved a one-to-one mapping between projects and types by constructing the Hasse diagram and the lightweight KG, and delved into the potential semantic relationships between entities, surpassing the recommendation algorithm proposed in study [7] in terms of semantic understanding ability. The trust relationship recommendation model proposed in study [8] has some

innovations in data preprocessing and trust relationship modeling, but it is slightly lacking in the depth and temporal considerations of the recommendation model. This study not only constructs a deep CF temporal model based on semantic relationships, but also fully considers the dynamic changes in user needs, thus outperforming the recommendation model proposed in study [8] in terms of recommendation accuracy and real-time performance.

The study conducted experimental verification according to the above two parts. In the experiments on ORW, compared to UCF, UBG, and GENC, the precision of the proposed method was increased by 61.3%, 61.3%, and 77.3%, respectively, but 2.6% lower than CHLF. The recall rate was 57%, 59%, 6%, and 72.5% higher than other models, respectively, but the average MAE and RMSE were higher than CHLF by 66.6% and 64.85%, indicating the importance of subsequent optimization. In the final model validation, R and S were first analyzed. When R=0.001 and S=256, the HR@K and NDCG@K of this model reached the maximum mean of 0.7 and 0.45. The experiment selected vector dimension 10 and analyzed each model. Compared to NCF and Wide & Deep, the SRT-DCF_item had an average increase of 5.55% in HR₁₀ and 14.6% in NDCG₁₀. In the validation of the impact on vector dimensions, HR@K and NDCG@K were on average 6.9% and 12.9% higher than other models.

VI. CONCLUSION

A semantic-based CF recommendation system was proposed to address the personalized user preferences in interior design, and its recommendation performance was tested. In summary, the proposed CF recommendation algorithm based on semantic relationships has excellent performance. However, this study still remains at a static level in constructing KG,

failing to fully capture the dynamic growth and evolution of information in actual situations. Therefore, in future research, real-time data stream processing technology should be further introduced to capture and process new information. Meanwhile, efficient dynamic update algorithms should be developed to update entities and relationships in the KG in real time.

REFERENCES

- [1] Wu Q, Cheng X, Sun E, et al. Collaborative filtering algorithm based on optimized clustering and fusion of user attribute features//The International Society for Applied Computing (ISAC), Tokyo University of Science, Japan and Cisco Networking Academy. Proceedings of 2021 4th International Conference on Data Science and Information Technology (DSIT 2021). ACM, 2021, 1(1):136-140.
- [2] Wang M, Li Q. A Multi-Agent Based Model for User Interest Mining on Sina Weibo. China Communications (English), 2022, 19(2):225-234.
- [3] Li M, Li Y, Xu YC, et al. Explanatory Q&A recommendation algorithm in community question answering. Data Technologies and Applications, 2020, 54(4):437-459.
- [4] Gou L, Zhou L, Xiao Y. Design and Implementation of Collaborative Filtering Recommendation Algorithm for Multi-layer Networks. Proceedings of the International Computer Frontiers Conference, 2021, 000(001): 32-50.
- [5] Wu C, Liu S, Zeng Z, et al. Knowledge graph-based multi-context-aware recommendation algorithm. Information Sciences, 2022, 595(1):179-194.
- [6] Yan HC, Wang ZR, Niu JY, et al. Application of covering rough granular computing model in collaborative filtering recommendation algorithm optimization. Advanced Engineering Informatics, 2022, 51(1):101485-101495.
- [7] Miao R, Li B. A user-portraits-based recommendation algorithm for traditional short video industry and security management of user privacy in social networks. Technological Forecasting and Social Change, 2022, 185(3): 122103-122103.
- [8] Liang X, Yin J. Recommendation Algorithm for Equilibrium of Teaching Resources in Physical Education Network Based on Trust Relationship. Journal of Internet Technology, 2022, 23(1): 133-141.
- [9] Qi W, Yu J, Liang Q, et al. Design of Graph Neural Network Social Recommendation Algorithm Based on Coupling Influence. International journal of pattern recognition and artificial intelligence, 2022, 36(14): 122103-122111.
- [10] Wang F. IoT for smart English education: AI-based personalised learning resource recommendation algorithm. International Journal of Computer Applications in Technology, 2023, 71(3): 200-207.
- [11] Zhang L, Du Y. Resilience of space information network based on combination of complex networks and hypergraphs. Computer communications, 2022, 195(11): 124-136.
- [12] Zheng L, Zhao T, Han H, et al. Personalized Tag Recommendation Based on Convolution Feature and Weighted Random Walk. International Journal of Computational Intelligence Systems, 2020, 13(1):24-35.
- [13] Taheri M, Farnaghi M, Alimohammadi A., et al. Point-of-interest recommendation using extended random walk with restart on geographical-temporal hybrid tripartite graph. Spatial Science, 2021, 68(1):71-89.
- [14] Yang Q, Wang H, Bian M, et al. Incorporating Reverse Search for Friend Recommendation using Random Walk. The international arab journal of information technology, 2020, 17(3): 291-298.
- [15] Manju G, Abhinaya P, Hemalatha MR, et al. Cold Start Problem Alleviation in a Research Paper Recommendation System Using the Random Walk Approach on a Heterogeneous User-Paper Graph. International Journal of Intelligent Information Technologies (IJIT), 2020, 16(2): 24-48.
- [16] Liu Y, Ma H, Jiang Y, et al. Learning to Recommend via Random Walk with Profile of Loan and Lender in P2P Lending. Expert Systems with Applications, 2021, 174(1): 114763-114776.
- [17] Masood F, Masood J, Zahir H, et al. Novel approach to evaluate classification algorithms and feature selection filter algorithms using medical data. Journal of Computational and Cognitive Engineering, 2023, 2(1): 57-67.
- [18] Fang Y, Luo B, Zhao T, et al. ST-SIGMA: Spatio-temporal semantics and interaction graph aggregation for multi-agent perception and trajectory forecasting. CAAI Transactions on Intelligence Technology, 2022, 7(4):744-757.
- [19] Li G, Liu H, Li G, et al. LSTM-based argument recommendation for non-API methods. Science in China: Information Science (English), 2020, 63(9): 8-30.
- [20] Dib B, Kalloubi F, Nfaoui EH. A Boulaalam. Incorporating LDA with LSTM for followee recommendation on Twitter network. International Journal of Web Information Systems, 2021, 17(3): 250-260.

Hybrid Clustering Framework for Scalable and Robust Query Analysis: Integrating Mini-Batch K-Means with DBSCAN

Hybrid Model for Complex Data Clustering

Sridevi K N^{1*}, Dr. Rajanna M²

Assistant Professor and Research Scholar, Department of Information Science and Engineering,
Vemana Institute of Technology, Bengaluru, Karnataka, India¹

Professor, Department of Information Science and Engineering, Vemana Institute of Technology, Bengaluru, Karnataka, India²

Abstract—Query clustering is a significant task in information retrieval. Research gaps still exist due to high-dimensional datasets, noise detection, and cluster interpretability. Solving these challenges will support large language models with faster and more efficient responses. This study aims to develop a hybrid clustering approach combining Mini-Batch K-means (MBK) and Density-Based Spatial Clustering of Application with Noise (DBSCAN) to cluster large-scale query datasets for information retrieval. The proposed method utilizes a preprocessing technique for data cleaning, extracts meaningful features, and scales all the features from the query dataset. The proposed hybrid clustering framework utilizes preprocessed data for clustering. The clustering algorithms MBK provide fast, scalable clustering, and DBSCAN delivers a precise, density-based refinement to efficiently process large-scale datasets while enhancing cluster boundaries to handle outliers. The proposed hybrid clustering framework effectively performs query analysis in information retrieval with a Silhouette score of 72.14 % and adjusted rand index of 78.23%. Thus, the hybrid clustering approach provides a robust and scalable solution for query analyzing tasks.

Keywords—Hybrid clustering; information retrieval; mini-batch k-means; query analysis

I. INTRODUCTION

Query analysis in a clustering framework is vital for understanding user intent and identifying behavioral patterns in information retrieval (IR) systems. Clustering is a group of objects that groups related objects into the same cluster and unrelated objects into diverse clusters. It is an analytical technique for grouping unlabeled data to extract meaningful information [1, 2]. People pursue information by asking queries on search engines. If the search engine generates unsatisfactory information, the users create another query by redeveloping the previous queries [3]. Query analysis in the clustering framework is discovering commonly asked questions and current popular topics on a search engine [4]. The clustering faces challenges in computational complexity, cluster refinement, high-data dimensionality, convergence speed, scalability, and evaluation measures [5].

Clustering is extensively utilized in pattern recognition, data mining, and query analysis, and different types of clustering algorithms have been proposed recently. The Spider

Optimization Algorithm with the Sequenced User Search Pattern Query Optimization (SOA-SUSPQO) improves the efficacy of the clustering query analysis and IR [6]. The Sampling-based Density Peaks Clustering (SDPC) algorithm minimizes the distance calculations in large datasets [7]. The hierarchical clustering, k-means, and Gaussian Mixture Models (DMM) on the Domain Name System (DNS) query dataset analyze and recognize the illicit activities in the domain names [8]. The centralized clustering procedure Low-Energy Dynamic Clustering improves the energy efficacy in query-based networks that target the cluster queries in a centralized way and supports the separation of clusters that match query targets [9]. The Machine Learning (ML) approach of K-means clustering classifies the data into K groups of similar examples for information retrieval [10]. However, traditional clustering algorithms face scalability, noise detection, and cluster interpretability challenges. The proposed model solves this issue by introducing hybrid clustering algorithms MBK and DBSCAN. The motivation for the proposed work is due to raising a few research questions.

- What are the limitations of existing clustering algorithms in handling large-scale, high-dimensional, and noisy query datasets?
- How can a hybrid approach combining Mini-Batch K-means and DBSCAN improve clustering scalability, accuracy, and outlier handling?
- What are the specific performance improvements in terms of clustering quality?

The major contributions of the proposed work are summarized as follows:

- The proposed methodology develops a hybrid clustering model combining Mini-Batch K-means (MBK) for fast, scalable initial clustering and Density Based Spatial Clustering of Application with Noise (DBSCAN) for precise refinement, addressing the limitations of each algorithm.
- In the proposed methodology, we design a two-phase clustering pipeline that leverages MBK's computational efficiency to reduce processing time for large datasets.

This is followed by DBSCAN for localized, detailed analysis.

- Address a significant gap in query clustering by proposing a hybrid approach that is both efficient and effective. This will set a benchmark for future clustering techniques that deal with dynamic, noisy data in real-world environments.

The remaining section of the paper is organized as follows: Section II analyses the existing clustering algorithms, Section III describes the proposed methodology, Section IV analyses the experimental results with an ablation study, and Section V concludes the paper.

II. LITERATURE SURVEY

This section reviews the performance of the existing clustering algorithms. Ates and Yaslam et al. [11] proposed a Graph-SeTES method that combines feature extraction and a decision network utilizing distance metrics and networks. The graph-based search task extraction addresses the challenges in search query logs. It improves the accuracy of short and misspelt queries, incomplete datasets, and limited labelled datasets. However, the model needed further improvements in quality embeddings. Shaik et al. [12] proposed graph-based and Machine Learning (ML) methods to improve the classification and clustering of incident ticket management systems using Resource Description Framework (RDF). The model faced limitations in scalability when utilizing a large dataset. Victor et al. [13] suggested integrated ML and PL/SQL tools to improve the database query performance. The model combined the Multi-Layer Perceptron (MLP) with k-means clustering to enhance the efficiency of the database response time.

Gong et al. [14] established a Query-Driven Clustering (QDC) protocol that leverages 5G infrastructure to improve energy efficiency and increase the network lifetime. However, the QDC algorithm required higher time complexity. To overcome this, Jia et al. [15] suggested a large-scale clustering model based on the Nystrom approximation to reduce the clustering complexity and enhance the clustering quality. Bashir et al. [16] proposed a proxy-terms-based query analysis by transforming true search queries into proxy queries that utilized the IR system, which preserves users' privacy. However, the model provided a query obfuscation only for the sensitive information contained in queries. Rehman et al. [17] suggested a Deep Learning (DL) based query response system to map farmers' queries to similar clusters. The system utilized a threshold-based clustering approach to group similar queries and enhanced the model's efficiency. The system faced difficulties with queries that had unclear meanings queries and included keywords in the dataset. Huang et al. [18] suggested a ML technique, K-means clustering, to identify the anomalous activities in the financial sector. However, the K-means clustering algorithms faced challenges in large datasets. Hartman et al. [19] developed a clustering algorithm for peptides to enhance the analysis of large spectrometry-based data that reduced the peptidomics dimensionality. The model needed further improvements in larger datasets. Zubair et al. [20] developed a model for improving the traditional K-means clustering algorithm, finding the optimal initial centroids to decrease the iterations and execution time.

Thus, the existing clustering works faced limitations in unclear and keyword queries, scalability, noise detection, large and high-dimensional datasets, and cluster interpretability. We proposed a hybrid clustering framework using the MBK and DBSCAN clustering algorithms to overcome these limitations.

III. PROPOSED METHODOLOGY

The framework of the proposed methodology is represented in Fig. 1.

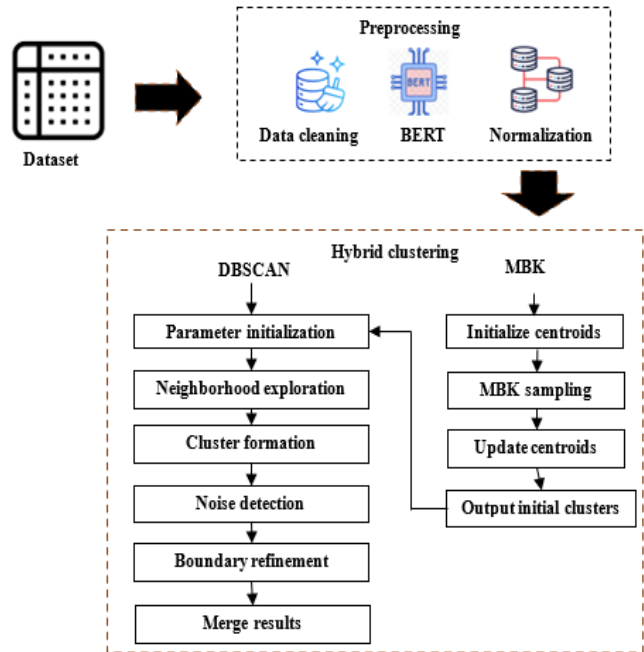


Fig. 1. Proposed hybrid clustering framework.

The proposed methodology utilizes an AOL User Session Collection 500K dataset to propose a scalable and robust query analysis framework. The query data preprocess using data cleaning, Bidirectional Encoder Representations from Transformers (BERT), and min-max normalization techniques. The data cleaning process removes the null values in the dataset, BERT extracts the meaningful features from the query data, and the min-max normalization technique scales all features in the dataset to ensure consistency across different feature dimensions. The proposed hybrid clustering algorithms MBK and DBSCAN evaluate the quality of the final clusters. The MBK clustering algorithm uses the pre-processed dataset to generate the initial clusters efficiently. The DBSCAN clustering algorithm processes each initial cluster independently to enhance cluster boundaries and detect dense regions and outliers. We utilize the hybrid clustering algorithms MBK for fast, scalable clustering and DBSCAN for precise, density-based refinement to efficiently process large datasets while enhancing cluster boundaries to handle outliers.

A. Dataset Description

The data were taken from the AOL User Session Collection 500K dataset [21]. The dataset covers 20 million web queries collected from 6,50,000 users over three months. A sequentially arranged unnamed user ID arranges the data. It provides real

query log data that is based on real users. The dataset includes four columns: AnonID, Query, QueryTime, ItemRank, and ClickURL. The AnonID contains the unnamed user ID number, the Query column contains the details of the user-issued queries, QueryTime represents the submitted time of the query, ItemRank denotes if the user clicks on a search result, the rank of the clicked item is listed, and the ClickURL listed the domain portion of the URL clicked. A query was NOT followed by the user clicking on a result item.

B. Preprocessing

Preprocessing enhances the quality of the query dataset. This research utilizes data cleaning, BERT, and min-max scaling normalization preprocessing techniques to improve the proposed query dataset.

1) *Data cleaning*: The data cleaning process removes the null values, such as queries with missing fields and keywords in the dataset. This process standardizes the format of the user's anonymous ID.

2) *Feature engineering*: BERT [22] is a language model that extracts meaningful query data features. BERT pretrains the query texts in the dataset. It breaks a sequence of tokens into characters, sentences, and words. The BERT language model pre-trained the queries in the dataset if the tokenizer did not identify the pretraining word, it split into sub words (eg: "Query"= "Que" and "ry") until the tokenizer found the sub word.

3) *Min-max normalization*: This research utilizes a min-max normalization [23] to scale all features in the dataset to a common range to ensure consistency across different feature dimensions. Min-max normalization performs linear transformations of the input data to generate a balance of value comparison between data after and before the process. The formulation of min-max normalization is expressed in equation (1).

$$Z_n = \frac{Z - \min(Z)}{\max(Z) - \min(Z)} \quad (1)$$

Where $\min(Z)$, and $\max(Z)$ denotes minimum and maximum values in the dataset, Z is the old value, and Z_n represents the new value from the normalized results.

C. Proposed Hybrid Clustering Algorithms

In this research, we propose a hybrid clustering approach that combines MBK and DBSCAN clustering algorithms that handle large-scale, high-dimensional datasets efficiently. The integration of the DBSCAN algorithm's ability to detect outliers with the MBK algorithm's computation speed to improve noise identification in large query datasets. In the first phase, the dataset is first clustered using the MBK algorithm. MBK uses mini-batches of data to update the centroids sequentially and ensure a good approximation of the clustering result. This helps speed up the process. The query dataset is first divided into a predetermined number of clusters by MBK. The overall computing load can be decreased by using this quick computation, particularly when working with big, high-dimensional query datasets. The DBSCAN algorithm is used in the second phase to refine the clusters after MBK has created the

initial clusters. DBSCAN is a density-based clustering technique that can effectively manage outliers and detect clusters of different sizes and shapes by concentrating on the local density of points. The clustering process is made better overall by DBSCAN's capacity to identify and separate noise or outliers, which results in more meaningful and cohesive clusters. Additionally, it improves cluster boundary definitions that may have been ambiguous in the MBK step. Although MBK is quick, DBSCAN addresses MBK's limitations when handling noise and irregular cluster forms by fine-tuning the cluster boundaries to ensure higher accuracy in the clustering results. A hybrid framework that is accurate and computationally efficient is developed by combining the speed of MBK with the density-based refining of DBSCAN.

1) *Mini-batch K-means clustering algorithm*: The MBK algorithm is unsupervised learning, an improved version of the K-means clustering algorithm. It resolves clustering techniques in mixed and large datasets. This research utilizes the MBK algorithm [24] to cluster the large-scale query dataset to enhance the scalability and optimize the clustering output. The MBK requires mini-batches as input, which are random subsets of the whole dataset. The MBK has a faster computation time than the k-mean algorithm. This clustering algorithm finds the set F of cluster centers $p \in R^S$ with $|F| = k$, to minimize over a set YD of examples $yd \in R^S$ the below objective function.

$$\min \sum_{yd \in YD} \|g(F, yd) - yd\| \quad (2)$$

In equation (2), $g(F, yd)$ returns the Euclidean distance of the adjacent cluster center $c \in F$ to yd . The problem is NP-hard, that gradient descent approach converges to the local optimum when seeded with an original set of k examples are drawn randomly from YD . The algorithm of MBK clustering is represented in pseudocode 1.

As shown in Fig. 2 MBK algorithm splits the dataset into smaller units known as mini-batches. MBK efficiently handles large datasets due to its fast computations by iteratively processing the mini-batches. In the MBK, centroids are placed as an initial marker; data points in each mini-batch are made reachable with the neighbouring centroid to locate their cluster. In the centre of the respective clusters, the centroids alter the changing distribution of data points. The final centroid generates a picture of the dataset's original clusters. Every data item fits into a certain cluster, which serves as an objective perspective for understanding, analyzing, and interpreting.

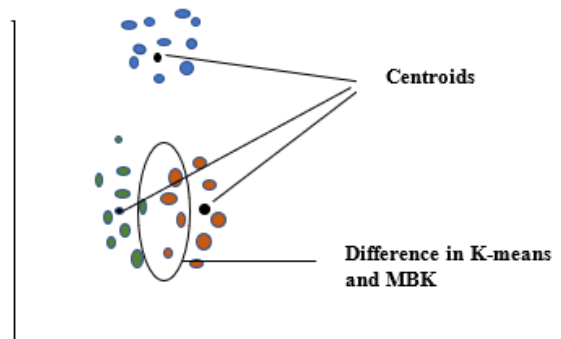


Fig. 2. Illustration diagram of the MBK clustering algorithm.

Pseudocode 1: Algorithm of MBK clustering phase

Objective: Initialize centroids for clustering

Input: $k, s \rightarrow$ mini-batch size m , iterations $i, YD \rightarrow$ data set.

Output: Set of clusters.

```

1: Initialize  $c \in F$  with  $yd$  select randomly from  $YD$ 
2:  $z \leftarrow 0$ 
3: for  $l = 1$  to  $i$  do
4:    $S \leftarrow m$  examples select randomly from  $YD$ 
5:   for  $yd \in S$  do
6:      $d[yd] \leftarrow g(F, yd)$ 
7:   end for
8:   for  $yd \in S$  do
9:      $c \leftarrow d[yd]$ 
10:     $z[c] \leftarrow z[c] + 1$ 
11:     $\eta \leftarrow 1/z[c]$ 
12:     $c \leftarrow (1 - \eta)c + \eta yd$ 
13:   end for
14: end for

```

2) *DBSCAN Clustering algorithm:* DBSCAN refers to a density-based clustering algorithm that efficiently processes the high-dimensional data and effectively distinguishes the noises in the clusters. This research utilizes the DBSCAN [25] clustering algorithm to cluster the high-density areas of target point data into clusters. It splits the data points into core, border, and noise points, as shown in Fig. 3, respectively, to the neighbourhood density points. This algorithm has neighbourhood ϵ and MinPts for density threshold parameters. The process of the DBSCAN clustering algorithm is denoted in pseudocode 2.

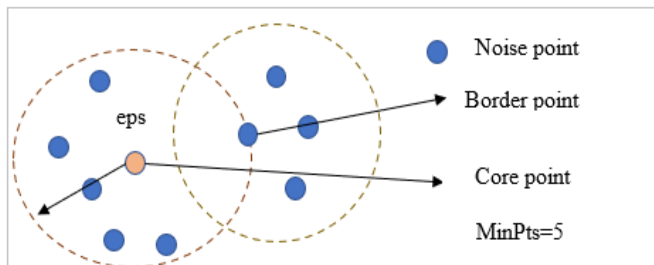


Fig. 3. Illustration diagram of DBSCAN clustering algorithm.

Pseudocode 2: Algorithm of DBSCAN clustering phase

Input: Initial MBK clustering with ϵ , MinPts

Output: Set of clusters

```

1: Randomly select a point  $P$ 
2: Regain all points from density reachable from  $P$ ,  $\epsilon$ , and MinPts
3: If  $P$  is a core point cluster formed
4: If  $P$  is a core point, no points are density reachable from  $P$  and DBSCAN visits the next point
5: Continue the process
6: All points are processed

```

IV. RESULTS

This section analyses the process and experimental results of the proposed hybrid clustering framework. The AOL User Session Collection 500K data preprocess using the data cleaning,

BERT, and min-max normalization techniques. The performance of the proposed model is evaluated using the metrics of the Silhouette score, Adjusted Rand Index (ARI), and Davies-Bouldin index, and a comparative assessment analyzes the effectiveness of the proposed clustering framework. The proposed hybrid clustering framework was executed in the Python 3.10 platform on a computer with Windows 10 Pro, Intel(R) Xeon(R) CPU E5-1650 v3 @ 3.50 GHz.

A. Performance Analysis of the Proposed Model

The proposed hybrid clustering framework is evaluated using the performance metrics of the Silhouette score, ARI, and Davies-Bouldin Index. Fig. 4 illustrates the proposed hybrid clustering framework performance. The silhouette score of 72.14 % estimates the quality of clustering algorithms, the ARI of 78.23 % calculates the similarity between the two partitions of a dataset and the Davies-Bouldin Index of 86.79 % estimate the quality of the clustering models.

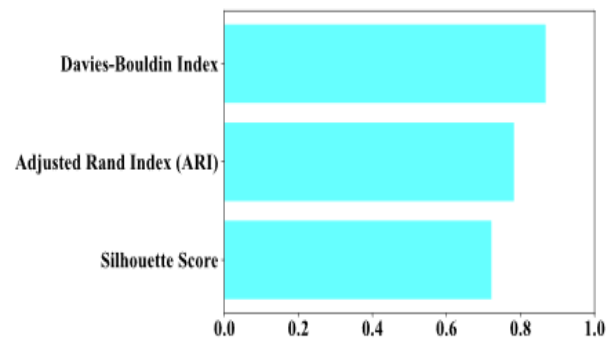


Fig. 4. Performance of the proposed hybrid clustering model.

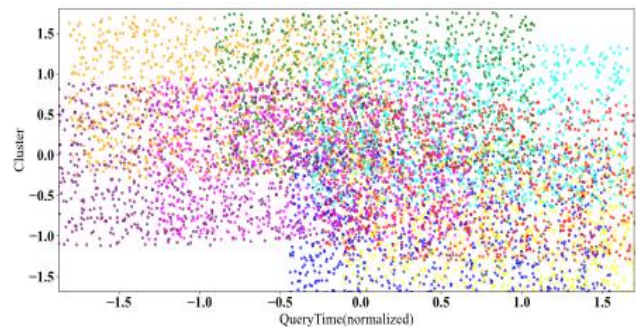


Fig. 5. Performance of the MBK initial clustering.

Fig. 5 illustrates the initial clusters generated by Mini-Batch K-means. The x-axis denotes the query time, and the y-axis represents the clusters in the query data. This represents the different clusters in distinct colors with centroids. In this clustering phase, the MBK includes various noises and outliers.

Fig. 6 demonstrates the clustering with centroids, which are represented in a group of different colours. In this graph, the x-axis signifies the query time, and the y-axis denotes the clusters in the query data. The outliers in the MBK clustering are denoted by red dots.

Fig. 7 illustrates the performance of DBSCAN clustering. It highlights DBSCAN's effect on refining clusters and detecting outliers. This graph shows how the proposed hybrid clustering algorithm efficiently reduces outliers in large-scale datasets.

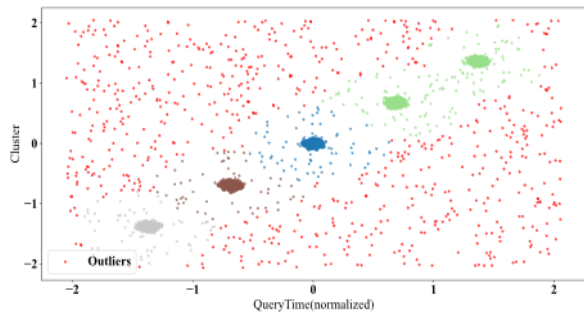


Fig. 6. Clustering with noise outliers.

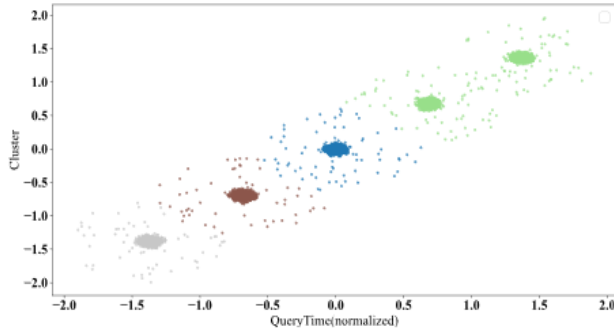


Fig. 7. Performance of the DBSCAN clustering.

Fig. 8 illustrates the performance of the queries in the dataset based on the execution time. This graph highlights the efficiency of the proposed hybrid approach. It shows that when the number of clusters increases at the same time, the execution time of the cluster also gradually increases. It is used to analyze the progression of clusters in the query data.

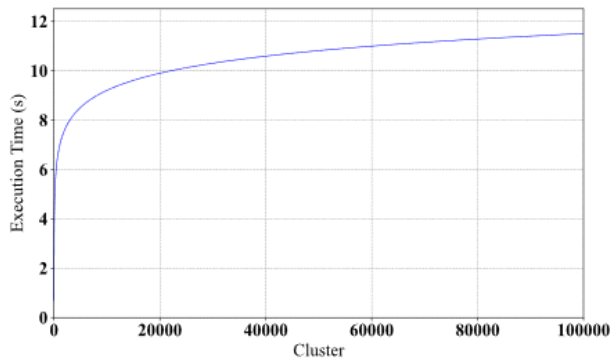


Fig. 8. Performance of the clusters based on the execution time.

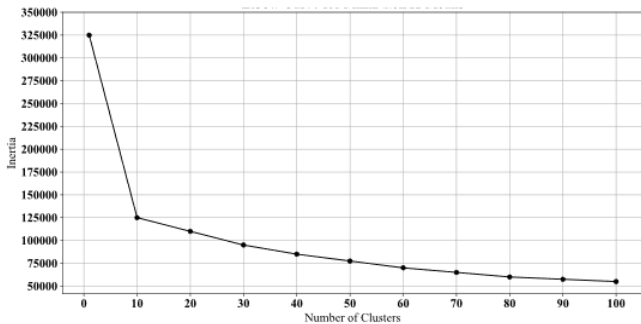


Fig. 9. Elbow curve of MBK clustering algorithm.

Fig. 9 represents the elbow curve, which shows the optimal number of clusters (k) by plotting the Within-Cluster Sum of Squares (WCSS) against diverse values of k . The visualization of the plot and the location of the elbow joint indicate the most suitable number of clusters. The MBK clustering algorithm was applied to partition the dataset into distinct clusters. The obtained clusters are denoted in the above figure.

B. Ablation Study

The ablation study evaluates the model's performance with the MBK, DBSCAN, and hybrid clustering algorithms. It highlights the importance of the proposed hybrid approach, which significantly improves the clustering quality. The MBK clustering algorithm ensures the noise detection abilities of DBSCAN and reduces the execution time. Proposed model performance with ablation study is listed in Table I.

TABLE I. PROPOSED MODEL PERFORMANCE WITH ABLATION STUDY

Methods	Silhouette score (%)	ARI (%)	Noise (%)	Execution time (s)
MBK	0.65	0.72	4.5	1.2
DBSCAN	0.72	0.75	7.2	3.5
Hybrid clustering algorithms (mini-batch & DBSCAN)	0.7214	0.7823	6.0	2.3

C. Comparative Analysis of the Proposed Model with the Existing Approaches

Table II compares the proposed hybrid clustering algorithms with the existing K-means+GMM, GMM+DBSCAN, and KisanQRS approaches.

TABLE II. COMPARATIVE ANALYSIS OF THE PROPOSED VS EXISTING CLUSTERING APPROACHES

Model	Silhouette score (%)	Davies-Bouldin Index (%)
Shaik et al., 2024, K-means+GMM [12]	0.6569	0.4614
Shaik et al., 2024, GMM+DBSCAN [12]	0.6569	0.4614
Rehman et al., 2023, KisanQRS [17]	0.6218	0.4998
Proposed	0.7214	0.8679

V. DISCUSSION

The proposed model develops a hybrid clustering approach that overcomes the limitations of traditional clustering algorithms and achieves a balance between speed, accuracy, and noise handling. As represented in Table I, the MBK clustering algorithm attains a low silhouette score of 0.65%, and ARI of 0.72% with an execution time of 1.2 s. The DBSCAN clustering algorithm alone obtains a silhouette score of 0.72%, and ARI of 0.75% with an execution time of 3.5 s. This shows the proposed hybrid clustering framework performs better with a silhouette score of 0.7214%, and ARI of 0.7823% with an execution time of 2.3 s. The MBK clustering algorithm improves the framework's scalability by processing the query dataset into smaller batches. This algorithm processes the AOL User Session Collection large-scale data effectively but has limitations in

handling noised data, and its centroids approach is complex to outliers. As shown in Table II, the existing approach achieves a Silhouette score of 0.6569 %, 0.6569 %, and 0.6218 %, while the proposed model attains a better Silhouette score of 0.7214 %, which shows that the proposed model attains a better clustering performance. The DBSCAN clustering algorithm refines the clusters and identifies the outliers effectively. This algorithm's density-based approach effectively detects the noisy points in the clusters and improves the overall clustering quality. The Silhouette score, ARI, and Davies-Bouldin index performance metrics evaluate the clustering quality. The ablation study highlights the hybrid approach significantly improves the clustering quality. It proves the noise-handling capabilities of DBSCAN while reducing the execution time through the MBK clustering algorithm. This research demonstrates the significance of hybrid clustering algorithms for scalable and robust query clustering. Training was not required for the line query when $k = m$ because the number of clusters and data lines was equal, allowing for instantaneous data retrieval. The efficiency in the group query, $k = n$, where n is the number of clusters chosen, is determined by the training level: time is proportionately direct to k . After combining a few randomly chosen centroids, the data were obtained in the case of the whole query. The model's friction and fatigue ultimately serve as a representation of training consumption: friction for the k number and fatigue for the epoch amount specified in the parameters. Table III lists a few additional models that have been employed in various studies to determine the ideal k value.

TABLE III. QUALITY COMPARISON WITH OTHER SIMILAR MODELS

Ref	Model	Data	Epochs	K
[26]	Kmeans FE	50	N/A	N/A
[27]	Unsupervised Kmeans	400	11	k
[28]	Kmeans spherical	REST	N/A	k=6
Proposed	MBK and DBSCAN	100000	10	k=5

VI. CONCLUSION

In this research, we addressed the problem of clustering large-scale, high-dimensional query datasets. We proposed a hybrid clustering algorithm that provides scalable and robust noise-handling query intent detection in IR systems. The proposed methodology utilizes data cleaning, BERT, and min-max normalization techniques to extract meaningful features from the dataset. The proposed hybrid methodology begins with the MBK, which generates initial clusters by producing mini-batch sampling to handle large datasets. The DBSCAN refines the cluster boundaries and detects outliers in each cluster. The proposed hybrid algorithms overcome traditional clustering algorithms' limitations, enhancing their performance and interpretability. The experimental results demonstrate that the hybrid approach achieved superior clustering performance with a Silhouette Score of 72.14% and an ARI of 78.23%, making the model more suitable for developing LLMs. There are a number of directions for further investigation, even though the proposed approach offers notable advancements. First, advanced deep learning-based clustering methods such as deep embedded clustering may improve performance even further by taking advantage of latent feature representations. Second, real-time

clustering requirements in streaming data environments could be met by expanding proposed work to accommodate dynamic datasets that change over time. These possible paths provide chances to support advanced systems, improve clustering techniques, and create more effective models.

REFERENCES

- [1] D. Cheng, Y. Li, S. Xia, G. Wang, J. Huang and S. Zhang, "A Fast Granular-Ball-Based Density Peaks Clustering Algorithm for Large-Scale Data," IEEE Transactions on Neural Networks and Learning Systems, vol. 35, no. 12, pp. 17202-17215, 2024.
- [2] Oyewole, Gbeminiyi John, and George Alex Thopil. "Data clustering: application and trends." Artificial Intelligence Review 56, no. 7 (2023): 6439-6475.
- [3] Xiong, Haoyi, Jiang Bian, Yuchen Li, Xuhong Li, Mengnan Du, Shuaiqiang Wang, Dawei Yin, and Sumi Helal. "When search engine services meet large language models: visions and challenges." IEEE Transactions on Services Computing (2024).
- [4] Dhanaraj, Rajesh Kumar, Vinothsaravanan Ramakrishnan, M. Poongodi, Lalitha Krishnasamy, Mounir Hamdi, Ketan Kotecha, and V. Vijayakumar. "Random forest bagging and x-means clustered antipattern detection from SQL query log for accessing secure mobile data." Wireless communications and mobile computing 2021, no. 1 (2021): 2730246.
- [5] Pitafi, Shahneela, Toni Anwar, and Zubair Sharif. "A taxonomy of machine learning clustering algorithms, challenges, and future realms." Applied sciences 13, no. 6 (2023): 3529.
- [6] Surya, S., and P. Sumitra. "Efficient query clustering and information retrieval using Sequenced User Search Pattern Query Optimization." Multimedia Tools and Applications (2024): 1-23.
- [7] Ding, Shifei, Chao Li, Xiao Xu, Ling Ding, Jian Zhang, Lili Guo, and Tianhao Shi. "A sampling-based density peaks clustering algorithm for large-scale data." Pattern Recognition 136 (2023): 109238.
- [8] Khaoula, Radi, Moughit Imane, and Moughit Mohamed. "Improving Cyber Defense with DNS Query Clustering Analysis." In 2024 11th International Conference on Wireless Networks and Mobile Communications (WINCOM), pp. 1-6. IEEE, 2024.
- [9] Gong, Yadong, and Guoming Lai. "Low-energy clustering protocol for query-based wireless sensor networks." IEEE Sensors Journal 22, no. 9 (2022): 9135-9145.
- [10] Purohit, Karan, Satvik Vats, Rishabh Saklani, Vinay Kukreja, Vikrant Sharma, and Satya Prakash Yadav. "Improvement in K-Means Clustering for Information Retrieval." In 2023 4th International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 1239-1245. IEEE, 2023.
- [11] Ates, Nurullah, and Yusuf Yaslan. "Graph-SetES: A graph based search task extraction using Siamese network." Information Sciences 665 (2024): 120346.
- [12] Shaik, Mohammed Ali, N. Sai Anu Deep, G. Srinath Reddy, B. Srujana Reddy, M. Spandana, and B. Reethika. "Graph Based Ticket Classification and Clustering Query Recommendations through Machine Learning." Library Progress International 44, no. 3 (2024): 25828-25837.
- [13] Silva-Blancas, Victor Hugo, Hugo Jiménez-Hernández, Ana Marcela Herrera-Navarro, José M. Álvarez-Alvarado, Diana Margarita Córdova-Esparza, and Juvenal Rodríguez-Reséndiz. "A Clustering and PL/SQL-Based Method for Assessing MLP-Kmeans Modeling." Computers 13, no. 6 (2024): 149.
- [14] Gong, Yadong, Junbo Wang, and Guoming Lai. "Energy-efficient Query-Driven Clustering protocol for WSNs on 5G infrastructure." Energy Reports 8 (2022): 11446-11455.
- [15] Jia, Hongjie, Qize Ren, Longxia Huang, Qirong Mao, Liangjun Wang, and Heping Song. "Large-scale non-negative subspace clustering based on nystrom approximation." Information Sciences 638 (2023): 118981.
- [16] Bashir, Shariq, Daphne Teck Ching Lai, and Owais Ahmed Malik. "Proxy-terms based query obfuscation technique for private web search." IEEE Access 10 (2022): 17845-17863.
- [17] Rehman, Mohammad Zia Ur, Devraj Raghuvanshi, and Nagendra Kumar. "KisanQRS: A deep learning-based automated query-response system for

- agricultural decision-making." *Computers and Electronics in Agriculture* 213 (2023): 108180.
- [18] Huang, Zengyi, Haotian Zheng, Chen Li, and Chang Che. "Application of machine learning-based k-means clustering for financial fraud detection." *Academic Journal of Science and Technology* 10, no. 1 (2024): 33-39.
- [19] Hartman, Erik, Fredrik Forsberg, Sven Kjellström, Jitka Petrlova, Congyu Luo, Aaron Scott, Manoj Puthia, Johan Malmström, and Artur Schmidtchen. "Peptide clustering enhances large-scale analyses and reveals proteolytic signatures in mass spectrometry data." *Nature Communications* 15, no. 1 (2024): 7128.
- [20] Zubair, Md, MD Asif Iqbal, Avijeet Shil, M. J. M. Chowdhury, Mohammad Ali Moni, and Iqbal H. Sarker. "An improved K-means clustering algorithm towards an efficient data-driven modeling." *Annals of Data Science* 11, no. 5 (2024): 1525-1544.
- [21] <https://www.kaggle.com/datasets/dineshydv/aol-user-session-collection-500k>
- [22] Panoutsopoulos, Hercules, Borja Espejo-Garcia, Stephan Raaijmakers, Xu Wang, Spyros Fountas, and Christopher Brewster. "Investigating the effect of different fine-tuning configuration scenarios on agricultural term extraction using BERT." *Computers and Electronics in Agriculture* 225 (2024): 109268.
- [23] Jlassi, Oussama, and Philippe C. Dixon. "The effect of time normalization and biomechanical signal processing techniques of ground reaction force curves on deep-learning model performance." *Journal of Biomechanics* 168 (2024): 112116.
- [24] Purba, Andrew Castello, and Teny Handhayani. "Comparison of K-Means, Affinity Clustering, and Mini Batch K-Means Algorithms for Market Segmentation Analysis." *Komputa: Jurnal Ilmiah Komputer dan Informatika* 13, no. 1 (2024): 54-63.
- [25] Han, Jianfeng, Xuefei Guo, Runcheng Jiao, Yun Nan, Honglei Yang, Xuan Ni, Danning Zhao et al. "An automatic method for delimiting deformation area in insar based on hns-w-dbscan clustering algorithm." *Remote Sensing* 15, no. 17 (2023): 4287.
- [26] Benaimeche, M.A., Yvonne, J., Bary, B. and He, Q.C., 2022. A k-means clustering machine learning-based multiscale method for anelastic heterogeneous structures with internal variables. *International Journal for Numerical Methods in Engineering*, 123(9), pp.2012-2041.
- [27] Sinaga, K.P. and Yang, M.S., 2020. Unsupervised K-means clustering algorithm. *IEEE access*, 8, pp.80716-80727.
- [28] George, S., Seles, J.K.S., Brindha, D., Jebaseeli, T.J. and Vemulapalli, L., 2023. Geopositional Data Analysis Using Clustering Techniques to Assist Occupants in a Specific City. *Engineering Proceedings*, 59(1), p.8.

Modified Moth-Flame Optimization Algorithm for Service Composition in Cloud Computing Environments

Yeling YANG*, Miao SONG

College of Computer and Internet of Things, Chongqing Institute of Engineering, Chongqing, 400056, China

Abstract—Cloud computing service composition integrates services, distributed and diverse by nature, into an integrated entity that can meet a user's requirement with better effectiveness. However, some obstacles regarding high latency and suboptimal Quality of Service (QoS) still exist in a dynamic multi-cloud environment. This study addresses the limitations of traditional optimization algorithms in service composition, specifically the premature convergence and lack of population diversity in the Moth-Flame Optimization (MFO) algorithm. We propose the modified MFO algorithm with a new mechanism called Stagnation Finding and Replacement (SFR) to enhance the diversity of the population. It finds the static solutions based on a distance metric from globally optimal representative solutions and replaces them. MFO-SFR drastically improved all QoS metrics, such as response time, delay, and service stability. Empirical evaluations prove that MFO-SFR outperforms the baseline methods of multi-cloud service composition. It provides a computationally efficient and adaptive solution to cloud service composition problems, ensuring better resource utilization and higher user satisfaction in dynamic multi-cloud environments.

Keywords—Cloud computing; quality of service; service composition; edge cloud; moth-flame optimization

I. INTRODUCTION

Due to the growing demand for high-performance computing resources, the computing infrastructure has been transformed over the last several years [1]. Several new computational environments, ranging from cluster to grid and cloud computing models, have been created due to technological innovation [2]. As an architectural model, cloud computing provides users with shared computing capabilities available on-demand, with minimal Cloud Service Provider (CSP) interaction [3]. The main goal of this infrastructure is to consolidate geographically distributed resources to achieve greater efficiency, reliability, and performance [4]. Cloud computing facilitates the sharing of services and offers a diverse range of services that can be accessed from any location worldwide [5].

Cloud deployment models can generally be classified into four categories: public, private, hybrid, and community [6]. In public cloud deployments, multiple organizations subscribe to and utilize the exact cloud resources through a shared infrastructure model [7]. This method encourages cost-effectiveness as businesses only bear expenses for their particular resource usage. Private cloud deployments offer dedicated infrastructure environments for a single organization [8]. This model emphasizes heightened security and control as

it houses sensitive applications and data within the company's private cloud environment.

Hybrid cloud setups incorporate aspects of both public and private cloud designs. Companies can benefit from this method by having the flexibility to strategically place data and applications according to their sensitivity and processing needs [9]. Sensitive data that requires high security can be stored in the private cloud, while the public cloud can be utilized for cost-efficient and scalable computing operations. Community cloud deployments aim at a particular community of users with common interests or objectives [10]. These designs offer a shared infrastructure setting for numerous organizations in the community and may encourage cooperation and efficient use of resources.

Cloud computing services are broken down into three major classes: Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [11]. PaaS offers businesses and developers a robust platform for deploying and hosting software [12]. IaaS allows companies to monitor and control their network, storage, and servers using cloud computing [13]. SaaS involves the provision of software or applications as a service. External providers manage these programs [14]. Users can run applications and software through their web browsers without installing them on their devices.

As the said cloud service model evolves and expands worldwide, it can improve how services are delivered and controlled, allowing the CSP to respond to the different needs of the Cloud Service User (CSU). Service Level Agreements (SLAs) are essential in this situation as they define the desired level of service quality between the CSP and CSU. An SLA is a legally enforceable contract or formally negotiated agreement establishing the understanding and objectives between the CSP and the CSU. The document describes the specific terms and circumstances that govern the provision of services by the CSP.

Cloud computing relies on ensuring accessibility and efficient allocation of all necessary services [15]. There are two main challenges to overcome: first, it is difficult to anticipate the full range of potential service demands, particularly for software services. To solve this problem, complex services must be broken down into more straightforward, discrete, and essential components offered by different providers. Second, selecting the best combination of individual services from multiple providers with varying QoS attributes is an NP-hard optimization problem. Both challenges can be addressed through service composition. To guarantee user satisfaction,

this approach includes service selection from a diverse pool, adherence to composition constraints, identification of crucial QoS indicators, and accommodating the dynamism of services and network conditions.

The dynamic nature of cloud computing environments necessitates effective service composition strategies [16]. While heuristics, metaheuristics, and machine learning algorithms have been employed to address this challenge, each presents distinct advantages and limitations [17]. Heuristic approaches, often limited to single-objective optimization, may struggle with multi-objective problems [18]. Machine learning techniques, such as Deep Q Network (DQN), ADEC, and DQTS, have shown promise in solving multi-objective service composition problems [19]. However, their reliance on extensive training data can be prohibitive, particularly in complex scenarios. Metaheuristic algorithms, including evolutionary and swarm-based methods, offer a versatile and scalable approach to multi-objective optimization [20].

This paper proposes a novel service composition method for cloud computing environments enhancing the Moth-Flame Optimization (MFO) algorithm. By integrating a Stagnation Finding and Replacing (SFR) mechanism, the MFO algorithm addresses the common challenge of stagnation during optimization processes. This innovative approach dynamically detects and replaces stagnant solutions, effectively rejuvenating the search process and preventing the algorithm from converging prematurely on suboptimal solutions. Briefly, this research contributes to the following areas.

- We propose a novel service composition strategy designed explicitly for multiple-cloud environments. This strategy capitalizes on the distributed characteristics of service elements across multiple clouds to improve service quality.
- We introduce the MFO-SFR algorithm, a significant advancement over the traditional MFO algorithm. The MFO-SFR algorithm demonstrates demonstrably improved performance and diversification capabilities.
- A key innovation of our approach is the SFR strategy. This strategy dynamically detects and replaces stagnant solutions within the optimization process, leading to an overall improvement in performance.
- We incorporate an archive mechanism to enhance solution diversity further and ensure a more comprehensive search space exploration. This mechanism integrates both representative and globally optimal solutions encountered during the search process.

The rest of the paper is organized as follows. Section II presents related work on service composition and optimization techniques, identifying the gaps the current study intends to fill. Section III includes the simulation setup, results, and analysis to illustrate the efficiency of the proposed approach for improving key QoS metrics. Lastly, Section IV concludes the study with an overview of key results and contributions and suggests possible directions for further investigation.

II. RELATED WORK

Cloud computing environments demand real-time execution for quality-of-service conscious service composition. This entails maintaining coordination between achieving the best service configurations and ensuring efficient execution times for service composition. Prior research thoroughly examined combinatorial optimization methods to identify optimal service compositions within a specified time constraint. Nevertheless, the continuous expansion of cloud services results in a proportional increase in the problem's search space size. Consequently, these conventional methods are less effective at efficiently combining services within acceptable time limits.

As outlined in Table I, Karimi, et al. [21] suggested utilizing a genetic algorithm-based method to attain global optimization while complying with SLAs. Their methodology involves using service clustering to decrease the complexity of the search space and using association rule mining to improve service composition efficiency based on historical service consumption data. Experimental evaluations show that the proposed strategy is more efficient than comparable efforts.

TABLE I. OPTIMIZATION TECHNIQUES FOR SERVICE COMPOSITION IN CLOUD COMPUTING ENVIRONMENTS

Reference	Methodology	Key features	Limitations
[21]	Genetic algorithm with service clustering and association rule mining	Decreases complexity of search space; improves efficiency with historical data	Potential scalability issues with growing service datasets
[22]	The hybrid of the artificial bee colony and genetic algorithm	Two-stage optimization: GA for fitness, ABC for selection	High computational complexity
[23]	Capuchin search algorithm	Inspired by capuchin monkeys' social foraging behavior, focusing on both global and local optimization	It may require fine-tuning for different cloud environments
[24]	Honeybee mating optimization with trust-based clustering	Incorporates honeybee reproductive behavior; tackles trust issues	Underperforms in computational time for large-scale problems
[25]	Combining Aquila optimizer and particle swarm optimization	Hybrid approach; adaptive transition strategy	A complex implementation may be resource-intensive
[26]	Ant colony optimization with multi-pheromone mechanism and GA-inspired mutation	Addresses ACO's local optima issue; balanced exploration and exploitation	Potential risk of premature convergence without proper parameter tuning

Sefati and Halunga [22] used the Artificial Bee Colony and Genetic Algorithm (ABCGA) to generate optimal service compositions. This approach utilizes a two-stage optimization process. During the initial phase, a Genetic Algorithm (GA) determines potential services that satisfy particular fitness

requirements. Once the fitness function evaluation produces encouraging outcomes, these potential services are introduced to the Artificial Bee Colony (ABC) algorithm during the second step. The ABC algorithm enhances the service selection process by determining the service most closely matches individual user requirements. The effectiveness of the suggested ABCGA approach was assessed through experimentation utilizing the CloudSim simulator.

To tackle the task of enhancing service composition for multiple Quality of Service (QoS) metrics in cloud environments, Wang [23] introduced a new approach that utilizes the Capuchin Search Algorithm (CapSA). This algorithm mimics capuchin monkey social foraging patterns and exhibits its efficacy in addressing global and local optimization challenges. CapSA is chosen due to its inherent simplicity, reduced processing complexity, and well-rounded approach to exploration and exploitation. By conceptualizing service composition as an optimization problem, the proposed methodology seeks to reduce energy consumption and costs. According to empirical evaluations, the CapSA-based strategy substantially outperforms existing methods for achieving faster convergence and producing superior service compositions.

Zanbouri and Jafari Navimipour [24] investigated how Honeybee Mating Optimization (HMO) can address service composition in cloud computing environments. They focus on the connections between worker bees and the queen bee when choosing a new queen, utilizing knowledge from honeybee reproductive behavior. The optimization algorithm incorporates these biological inspirations to enhance the QoS. In addition, a trust-based clustering technique is used to tackle trust-related concerns specifically. The simulation results obtained from a C# implementation indicate that the suggested method outperforms existing algorithms, including GA, Particle Swarm Optimization (PSO), and the discrete best-guided ABC algorithm, for small-scale service composition problems. The enhancement results from the clustering method diminish the scope of the search and thus enhance the speed of response while also allowing for the choice of more dependable services. However, extensive simulations demonstrate that the computational time performance of the suggested method underperforms the average results of earlier studies.

Liu [25] developed a novel hybrid optimization technique known as the Integrated Aquila Optimizer (IAO), combining the functions of the PSO algorithm and Aquila Optimizer (AO). Hybridization addresses the inherent limitations of individual algorithms, such as their vulnerability to getting stuck in local optima and their limited ability to generate diverse solutions. The proposed IAO algorithm includes an innovative transition strategy for these difficulties. This method allows the AO and PSO algorithms to adjust their search operators flexibly. By employing this method, possible solutions are consistently improved. Utilizing both the AO and PSO algorithms can be a strategic move when each method reaches a standstill or when the range of possibilities decreases. This adaptive behavior improves the effectiveness and efficiency of the IAO approach. The proposed solution was thoroughly evaluated by testing in the CloudSim simulation environment. The numerical data indicate that the IAO technique successfully enhances

dependability, availability, and cost optimization within cloud computing.

Bei, et al. [26] discussed the composition of services in multiple cloud scenarios. They proposed an Ant Colony Optimization (ACO) algorithm to optimize QoS parameters, incorporating a multi-pheromone mechanism. This technique seeks to surpass conventional ACO constraints, which may become trapped in local optima. They incorporated a mutation operation influenced by the GA to improve the algorithm's exploration ability and avoid premature convergence. This hybridization approach promotes a more equitable and effective process of exploring and exploiting, resulting in the discovery of service compositions with superior QoS metrics, such as decreased latency and enhanced response times. Proposed method

A. Problem Definition and System Architecture

Cloud computing has made significant advancements in the past decade. Global infrastructure and market expansion have given rise to several cloud computing forms, including central and edge clouds. Central clouds are frequently utilized for extensive data analysis and deep learning training because of their robust processing and storage capacities. On the other hand, edge clouds are essential for collecting data, controlling processes in real-time, perceiving information intelligently, and making quick decisions at the outermost part of the network.

In contrast to centralized cloud infrastructures, edge computing provides users access to robust computational resources while mitigating the delay challenges inherent in remote data center interactions. This dramatically minimizes the data transmitted on the leading network and guarantees quick response times for upcoming services requiring minimal delays. As a result, the widespread use of these services in edge clouds is anticipated to grow prevalent.

This paper explores the architecture of cloud-edge devices, where service elements are mainly placed on a centralized cloud. Docker and other containerization technologies facilitate seamless and efficient migration to the cloud when consumers need a particular service component. This methodology enables the combination of services and the virtualization of resources (such as storage and computation) to meet users' requirements, as shown in Fig. 1. Docker containers are gaining popularity in cloud computing, as evidenced by their use in constructing genuine cloud environments for research purposes. Cloud services are highly advantageous in a dynamic cloud environment due to their effectiveness and ease.

The current service landscape is experiencing a significant change towards autonomous and loosely connected service designs, commonly called microservice architectures. Although service components can be spread out throughout different edge clouds, there is still a need to investigate and understand the current approaches for combining services in multiple-cloud setups. On the other hand, a multiple-cloud setup enables consumers to select from a range of services that perform better than single CSPs with limited computing capacity. In addition, multiple-cloud deployments provide built-in redundancy, which helps prevent equipment failures and improve the system's overall stability.

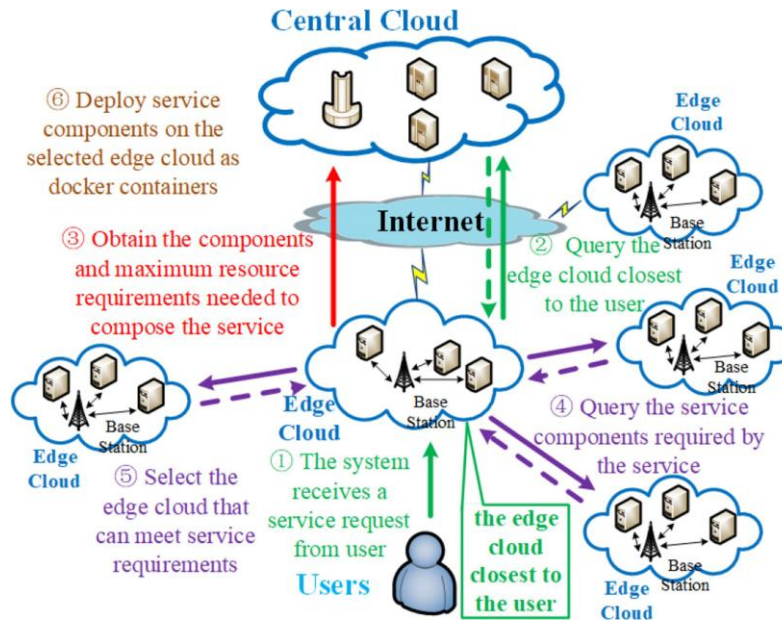


Fig. 1. Architecture of cloud-edge-devices integration.

This paper introduces a multi-cloud service composition architecture, depicted in Fig. 1, comprising M consumers, N edge clouds, and a central cloud. The central cloud is a repository for comprehensive service-related information and hosts a global network controller. Service components are distributed across an edge cloud infrastructure. User service requests are initially directed to the nearest edge cloud for preliminary processing. Each edge cloud maintains a local database containing information about neighboring edge clouds and available service components.

Table II is a crucial system component, providing essential network topology information. The index n_i varies from 0 to $N-1$, where N is the total number of edge clouds in the network. This ensures that all possible connections between edge clouds are considered. $path_i$ represents the optimal route linking the current edge cloud and another edge cloud in terms of the fewest hops. This information is crucial for routing data efficiently. hop_i quantifies the number of network hops between the current edge cloud and any other edge cloud within the network. A lower hop count generally indicates a more efficient communication path.

TABLE II. NETWORK TOPOLOGY INFORMATION FOR EDGE CLOUDS

Edge cloud count	Path	Hop
n_1	$path_1$	hop_1
n_2	$path_2$	hop_2
...
n_i	$path_i$	hop_3

Table III serves as a critical repository for service component metadata within the edge cloud environment. Service element names identify the service items available in the edge cloud and its neighbors. QoS attributes provide essential performance metrics for each service element, such as delay and reliability. The QoS attribute for the j^{th} parameter of

the i^{th} service element is represented as η_{ij} . This standardized notation facilitates data manipulation and analysis.

TABLE III. SERVICE ELEMENT DATABASE

Service element	Edge cloud count	QoS
$element_1$	n_1	η_1
$element_2$	n_2	η_2
...
$element_i$	n_i	η_3

Upon receipt of a service request, the proximate edge cloud initiates communication with a central controller to procure optimal computational resources, storage capacity, and network bandwidth. The service composition process proceeds in situ if the local edge cloud possesses sufficient residual capacity to fulfill the service's maximal requirements. Conversely, if resource constraints are encountered, the edge cloud embarks on a search for an adjacent edge cloud with minimal network hops. This iterative exploration continues until an edge cloud with ample resources to accommodate the service composition is identified.

Eq. (1) quantifies the resource demands (R_l) of service l . The set L encapsulates the services scheduled for orchestration on edge cloud i , while C_i represents the aggregate resource capacity of edge cloud i . These parameters constitute critical determinants in edge cloud service provisioning.

$$\sum_{l \in L} R_l \leq C_i, i \in N \quad (1)$$

Containerization virtualization has played a significant role in microservice adoption. Cloud computing can utilize containerization to flexibly install, migrate, or scale virtual machines under changing service demands. Containerization benefits conventional virtual machines by using the host's operating system kernel. This strategy minimizes the

administrative burden of delivering resources as needed and promotes optimal resource utilization. Containerized microservices typically involve the simultaneous creation of lightweight components within containers, which are then provisioned and scaled based on their requirements.

Expanding on these ideas, this method enables the quick and flexible deployment of service components by utilizing containerization technologies such as Docker in a multiple-cloud setting. This allows for deploying all necessary service components for composition onto respective edge clouds. A notification mechanism is built to guarantee real-time accuracy of records held in edge clouds and service components. Whenever one edge cloud stops serving customers or undergoes modifications to its deployed service components, it sends these updates to all connected edge clouds. Utilizing this broadcast technique allows for the timely updating of databases on other edge clouds, ensuring data consistency throughout the system.

B. QoS model

Services typically comprise k distinct groups, each containing abstract service component definitions with specific order requirements. Users seek a combination of services that fulfill user-specified requirements and QoS constraints to complete their desired operations during service composition.

The service composition process is divided into K steps according to the user's requirements. Every individual step, S_i , is linked to a particular service set. The algorithm chooses service components from each set S_i to fulfill the user's operation. The selection procedure yields numerous possible routes from the initial service component set (S_1) to the final set (S_k). The ideal combination of services is attained by determining the pathway that produces the most advantageous service combination.

When choosing a service, both the functional and non-functional aspects are considered. Functional attributes pertain to the explicit purpose and content offered by a service, whereas non-functional attributes encompass the overall quality of the service, as evaluated using QoS measurements.

Services are evaluated on the essential aspects of QoS, as described by internationally recognized standards organizations. QoS, as specified by these standards, includes non-functional features such as throughput, availability, response time, and dependability.

Ensuring high-quality service while combining multiple services is essential for distinguishing between the various components of the service. This optimization method assesses the QoS attributes of the constructed service. QoS parameters can be divided into two main categories: dynamic attributes, which include response speed, dependability, and availability, and fixed attributes, which include security, accuracy, and robustness.

This study focuses on throughput, reliability, delay, and response time. Throughput indicates the maximum rate at which data can be processed or transmitted successfully. Availability refers to the likelihood that service components are operational and ready for use in a particular environment. Delay

refers to the time it takes for data packets to travel between a server hosting a service component and a client.

Response time represents the time the service provider takes to respond to a user's service request. Table IV presents QoS attribute formulas for composed services. Calculations rely on j (number of service components chosen from service set i) and k (total number of service sets).

TABLE IV. QOS ATTRIBUTE FORMULAS FOR COMPOSED SERVICES

QoS parameters	Expression
Delay	$\sum_{i=1}^k L(\eta_{ij})$
Throughput	$\sum_{i=1}^k L(\eta_{ij})$
Availability	$\sum_{i=1}^k A(\eta_{ij})$
Response time	$\sum_{i=1}^k R(\eta_{ij})$

It is crucial to optimize various QoS parameters during service composition. However, it is equally essential to guarantee service stability and other relevant metrics. This study introduces a novel concept of QoS parameter stability, defined by the absolute value of each parameter across service elements. Eq. (2) represents the stability calculation for QoS parameter j within the service.

$$Sta_j = \sum_{i=1}^{k-1} \|\eta_{(i+1)(j+1)} - \eta_{ij}\| \quad (2)$$

Services with minimal cumulative absolute differences between their QoS parameters (QoS_i) are considered more stable. This approach mitigates significant fluctuations in QoS. Additionally, to prevent data size variations across service sets from skewing the final results, this paper incorporates a data normalization step for the QoS information associated with the service components. Following normalization, higher parameter values correspond to superior performance. Consequently, all subsequent references to QoS metrics (response time, availability, throughput, and delay) within this work will pertain to their normalized values.

This paper proposes a methodology that considers all four QoS criteria to determine the most effective technique for composing consumer services. This technique guarantees the optimization of these crucial parameters. The following section will explore an improved service composition technique based on the modified MFO algorithm. This approach has been specifically developed to boost the optimization of QoS.

C. Enhanced MFO algorithm

The MFO algorithm mimics the behavior of moths in nature. The unique navigational strategies of moths have generated considerable interest among researchers studying metaheuristics. Moths are nocturnal creatures that rely on lunar illumination for navigation Shehab, et al. [27]. Moth flight patterns can be mathematically modeled using the transverse orientation mechanism (Fig. 2). This strategy approximates a straight-line trajectory by maintaining a constant angular relationship with the moon. When faced with artificial light sources, moths divert from this path. When the moth is close to

the light source, it initiates a helical flight pattern that guides it towards the flame. Each moth symbolizes a potential solution, and every position is represented as a matrix of decision variables, as shown below.

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,n-1} & x_{1,n} \\ x_{2,1} & \ddots & \dots & \dots & x_{2,n} \\ \vdots & \dots & \ddots & \dots & \vdots \\ x_{N-1,1} & \dots & \dots & \ddots & x_{N-1,n} \\ x_{N,1} & x_{N,2} & \dots & x_{N,n-1} & x_{N,n} \end{bmatrix} \quad (3)$$

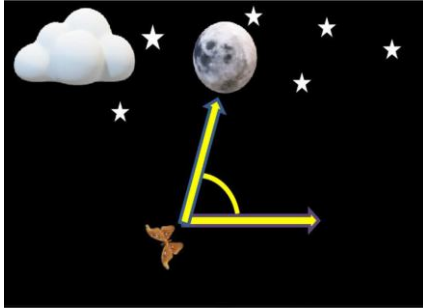


Fig. 2. Moth flight patterns model using transverse orientation mechanism.

In Eq. (3), N stands for the population size, equal to the total number of moths in the swarm. Also, n indicates the problem dimension, which measures how many variables are involved in the optimization process. The fitness of a particular moth is determined as follows.

$$Fit[X] = \begin{bmatrix} Fit[X_1] \\ Fit[X_2] \\ \vdots \\ Fit[X_n] \end{bmatrix} \quad (4)$$

Eq. (5) shows the flame matrix. Since all moths fly around a flame, the size must match the moth matrix previously defined.

$$FM = \begin{bmatrix} FM_1 \\ FM_2 \\ \vdots \\ FM_N \end{bmatrix} = \begin{bmatrix} Fm_{1,1} & Fm_{1,2} & \dots & Fm_{1,n-1} & Fm_{1,n} \\ Fm_{2,1} & \ddots & \dots & \dots & Fm_{2,n} \\ \vdots & \dots & \ddots & \dots & \vdots \\ Fm_{N-1,1} & \dots & \dots & \ddots & Fm_{N-1,n} \\ Fm_{N,1} & Fm_{N,2} & \dots & Fm_{N,n-1} & Fm_{N,n} \end{bmatrix} \quad (5)$$

Eq. (6) determines the corresponding fitness of the flame matrix.

$$Fit[FM] = \begin{bmatrix} Fit[FM_1] \\ Fit[FM_2] \\ \vdots \\ Fit[FM_n] \end{bmatrix} \quad (6)$$

The MFO algorithm relies heavily on two primary components: flames and moths. Moths fly through flames to achieve desired results. As shown in the equation below, the logarithmic spiral function is used to model the spiral movement of the moth.

$$X_i^{K+1} = \begin{cases} \delta_i \cdot e^{bt} \cdot \cos(2\pi t) + Fm_i(k) & i \leq N.FM \\ \delta_i \cdot e^{bt} \cdot \cos(2\pi t) + Fm_{N.FM}(k) & i \geq N.FM \end{cases} \quad (7)$$

δ_i represents the Euclidean distance between a moth's current position (X_i^K) and its corresponding flame (Fm_i). This value indicates the moth's proximity to a possible optimal solution. Spiral flight patterns of moths are determined by b and t , a uniformly distributed random number between -1 and 1. Moths and flames are attracted to each other based on these parameters, as shown in Fig. 3. The moth's trajectory towards the flame is depicted in Fig. 4. Throughout the optimization process, t gradually decreases toward a balance between exploitation (focusing on promising areas) and exploration (searching the entire search area). The mathematical representation of t is presented below, and Fig. 5 depicts the moth's next position.

$$r = -1 + Current_{iter} \left(\frac{-1}{Max_{iter}} \right) \quad (8)$$

$$t = (r - 1) \times k + 1 \quad (9)$$

The optimization process depends on three variables: Max_{iter} , k , and r . Max_{iter} specifies the maximum number of iterations, k indicates a uniformly distributed random number between 0 and 1, and r singularity ensures convergence. The value of r is linearly reduced throughout the optimization to balance exploration (searching the entire search space) and exploitation (focusing on promising regions).

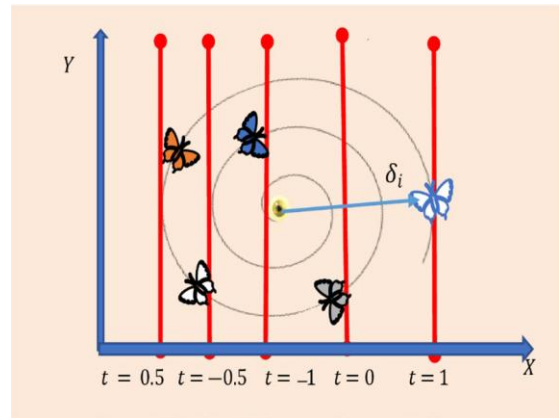


Fig. 3. Attraction mechanism between moths and flames.

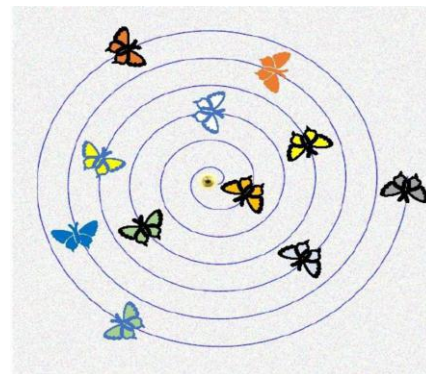


Fig. 4. Moth's spiral trajectory toward the flame.

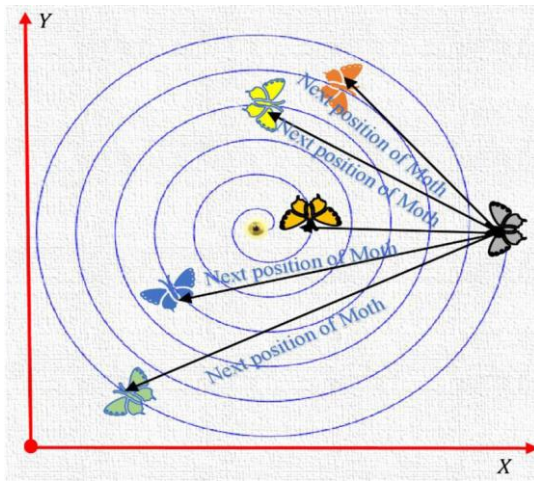


Fig. 5. Decreasing parameter t for balancing exploration and exploitation in optimization.

During the optimization process, the moths with the highest fitness values continually move towards the most promising solutions, indicated by the flames. This phenomenon can be explained by the mechanism in which the number of flames (represented as $N.FM$ in Equation 10) gradually reduces with each cycle. This decrease in the number of flames efficiently focuses the search effort on the most favorable areas of the search space.

$$N.FM =$$

$$\text{round} \left(N.FM_{\text{Last iter}} - \text{Current}_{\text{iter}} \frac{(N.FM_{\text{Last iter}} - 1)}{\text{Max}_{\text{iter}}} \right) \quad (10)$$

A population of moths is represented by the matrix $X(t) = \{X_{1D}(t), \dots, X_{iD}(t), \dots, X_{ND}(t)\}$ in a D -dimensional search space at iteration t . Each element $X_{iD}(t)$ represents the position of the i^{th} moth within the problem space. The initial positions of all moths are generated randomly using a uniform distribution during the first iteration ($t = 1$). In subsequent iterations ($t \geq 2$), the SFR mechanism is used to update moth positions based on Eq. (11).

$$X_i(t+1) = \begin{cases} D_i^\alpha(t) \times e^{b\tau} \times \cos(2\pi t) + F_j(t) & \text{if } i \leq R(t) \\ D_i^\beta(t) \times e^{b\tau} \times \cos(2\pi t) + F_R(t) & \text{else} \end{cases} \quad (11)$$

SFR is characterized by its core components, represented by Eq. (12) and Eq. (13). The constant b determines the shape of

the logarithmic spiral employed by the moth, and τ indicates a random number uniformly distributed between -1 and 1 . $F_j(t)$ and $F_R(t)$ represent the positions of the j^{th} and the R^{th} flame, respectively. The parameter r is calculated using Eq. (8).

$$D_i^\alpha(t) = |F_j(t) - M_i(t)| \quad (12)$$

$$D_i^\beta(t) = \begin{cases} |F_j(t) - X_i(t)| & \varphi_i > 0 \\ \text{Select a random position form Arc} & \varphi_i = 0 \end{cases} \quad (13)$$

Eq. (14) calculates the mean distance, denoted by φ_i , for each moth. This distance is computed based on the individual dimensions (X_{iq}) of the i^{th} moth and the corresponding dimensions (F_{jq}) of its associated flame (j). The index j for each moth is determined using Eq. (15). This equation involves sorting the results obtained from Eq. (14) in descending order to identify the most "distant" moths and subsequently utilizing these indices as flame indexes within Eq. (13).

$$\{\varphi_1, \dots, \varphi_i, \dots, \varphi_N\} \leftarrow \varphi_i = \frac{1}{D} \times \sum_{q=1}^D |F_{jq}(t) - X_{iq}(t)| \quad (14)$$

$$\{\varphi_1, \dots, \varphi_j, \dots, \varphi_N\} \leftarrow \text{Sort}(\varphi_1, \dots, \varphi_i, \dots, \varphi_N) \quad (15)$$

The archive construction process serves a dual purpose: enhancing population diversity and accelerating convergence towards promising regions within the search space. This is achieved by storing representative flames and the best solutions encountered during optimization. The archive, denoted by Arc , is represented by the matrix $M = \{M_1, \dots, M_i, \dots, M_K\}$, where K signifies the predefined archive size. Each element $M_i = [m_{i1}, m_{i2}, \dots, m_{iD}]$ represents a vector position within the archive memory.

The construction of the archive involves two key steps: generating Representative Flame (RF) and archiving entries. The first step leverages the dual population ($dual_{Pop}$) and dual fitness values ($dual_{Fit}$) created based on the flame construction process outlined in Fig. 6. Eq. (16) calculates the RF position, representing the average of all flame positions. Here, C denotes the total number of moths considered, and F_{id} represents the d^{th} dimension of the i^{th} flame. Two new entries are added to the archive memory M : the global best flame position and the calculated RF position. If the archive reaches its total capacity (K), a random replacement strategy is implemented, replacing two existing entries with new entries.

$$RF_d(t) = \frac{1}{C} \sum_{i=1}^C F_{id}(t) \quad (16)$$

Input: X : the positions of moths, Fit : the fitness values of moths, F : the position of the flame, and OF : the fitness values of flames.

Flame construction in the first iteration when $t = 1$.

1. Sort the vector Fit in ascending order and extract the sorted index in $\{j_1, j_2, \dots, j_N\}$.
2. Construct the flame matrix $F(t) = \{F_1 \leftarrow X_{j1}, F_2 \leftarrow X_{j2}, \dots, F_N \leftarrow X_{jN}\}$.

Flame construction for the rest iteration when $t > 1$.

1. Construct matrix $dual_{Pop}$ by combining matrices $F(t)$ and $X(t-1)$.
2. Construct vector $dual_{Fit}$ by combining vectors $OF(t)$ and $Fit(t-1)$.
3. Sort the vector $dual_{Fit}$ in ascending order and extract the sorted index in $\{j_1, j_2, \dots, j_{2N}\}$.
4. Construct the flame matrix $F(t) = \{F_1 \leftarrow X_{j1}, F_2 \leftarrow X_{j2}, \dots, F_N \leftarrow X_{jN}\}$.

Fig. 6. Construction of representative flame and archiving entries.

III. SIMULATION AND RESULTS

A series of tests were performed on a Windows 8.1 computer powered by an Intel Core i5-460M processor at 2.53 GHz and 16 GB of RAM. This study employed a system model comprising 32 edge clouds and a primary cloud, as illustrated in Fig. 1. The Quality of Service for Web Services (QWS) dataset contained 2507 services, each characterized by nine QoS features: description, delay, best practices, consistency, reliability, throughput, availability, and response time. For this research, delay, throughput, availability, and response time were the primary QoS parameters, with their respective ranges and units detailed in Table V. A comparative evaluation was conducted to measure the proposed algorithm against traditional MFO, PSO, and WOA algorithms, evaluating fitness, stability, delay, and response time.

TABLE V. QOS PARAMETERS AND RANGES

Parameters	Dimensions	Unit
Delay	0.1-4500	ms
Throughput	0.1-50	Mbps
Availability	5-100	%
Response time	30-5000	ms

Fig. 7 and Fig. 8 compare the proposed algorithm and its counterparts regarding fitness and stability, respectively. To conduct this analysis, 100 to 1000 service instances were chosen at random extracted from the QWS dataset, with inclusion criteria limited to services comprising at least five components. Fig. 7 demonstrates the better fitness performance of the developed algorithm compared to its competitors. While fitness is a crucial metric, the ultimate objective is to maximize service QoS and maintain stability.

Fig. 9 and Fig. 10 illustrate the comparative performance of the algorithms in terms of delay and response time. All QoS parameters were normalized to mitigate the influence of varying parameter scales. Consequently, higher values indicate improved optimization outcomes. The results in Fig. 9 clearly reveal the superiority of the proposed algorithm in minimizing delay. Similarly, Fig. 10 reveals a significant advantage of the proposed algorithm in reducing response time compared to other methods.

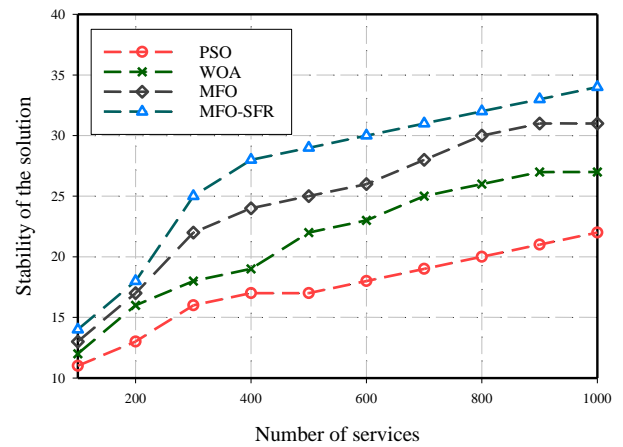


Fig. 8. Stability comparison.

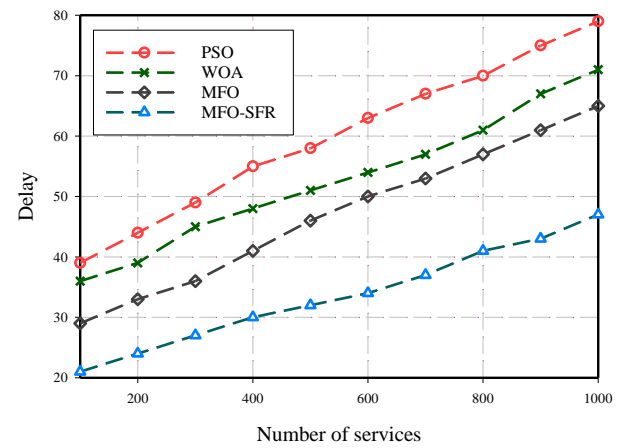


Fig. 9. Delay comparison.

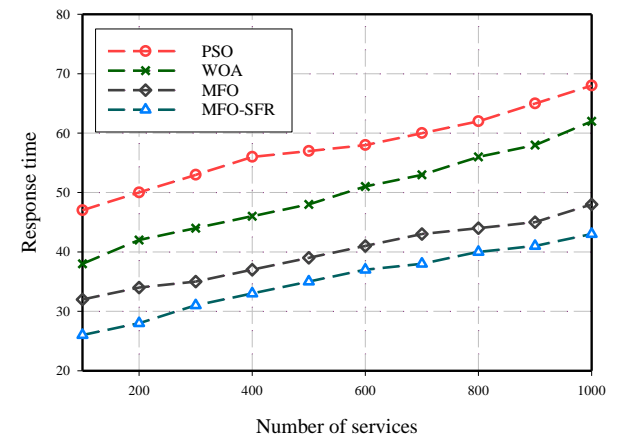


Fig. 10. Response time comparison.

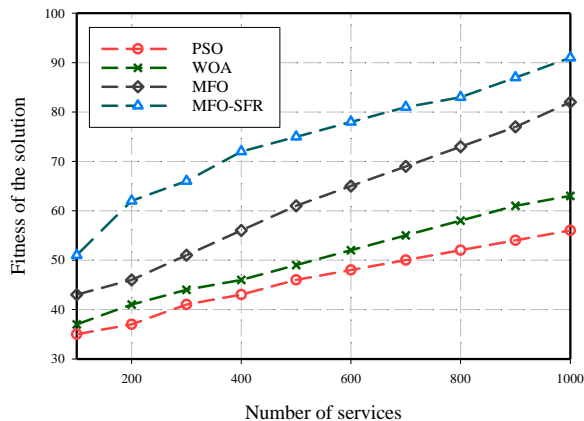


Fig. 7. Fitness comparison.

The experimental results distinctly demonstrate the superiority of the proposed algorithm over traditional optimization methods. From the fitness performance, it is obvious that the proposed algorithm constantly finds higher-quality solutions. This is mainly because the SFR mechanism can dynamically detect and replace the stagnant solution, making the search process more diverse and effective. Enhanced fitness will naturally provide better QoS outcomes,

crucial for service composition in multi-cloud environments. Moreover, the proposed algorithm outperforms others in service stability, which shows its robustness and adaptability to dynamic cloud scenarios.

The proposed MFO-SFR algorithm guarantees great efficacy when the delay and response time parameters are analyzed. A minimum delay shows how it may optimize the critical time-sensitive aspect of cloud service delivery, guaranteeing users' satisfaction with the service. On the other hand, the response times that could be retrieved by using this algorithm also promise to ensure that it may further enhance efficiency in the performance of any service. Therefore, the findings support the stated objectives of the study on QoS parameters in cloud environments and further indicate the practical relevance of the algorithm. Compared to traditional approaches, the proposed method yields better results in scalability and efficiency in solving complex service composition challenges in multi-cloud ecosystems.

IV. CONCLUSION

The swift advancement of cloud computing has led to the widespread growth of a wide range of cloud-based services. However, guaranteeing awareness of QoS during the construction of services is a substantial difficulty in cloud systems. Many individual services cannot handle complex requests and different needs that arise in real-world situations. Often, a solitary service may not be enough to fulfill the particular needs of consumers, therefore requiring the amalgamation of many services to attain the needed functionality. Due to its intrinsic NP-hard difficulty, service composition has been widely studied using various metaheuristic algorithms. This paper proposed an improved MFO algorithm with the SFR mechanism for the optimization of service composition in multi-cloud computing environments. Our approach overcomes the deficiency of early convergence in the traditional MFO by maintaining the diversity in the population with the aid of the SFR mechanism. It ensures that static solutions are identified and replaced with promising ones, which enhances the whole optimization process. The empirical results showed that our approach enhances significantly the QoS metrics such as stability of service, response time, and delay. We evaluated an algorithm using a realistic system model and the QWS dataset, considering the main QoS parameters. The comparative analysis confirmed the superior fitness and stability of our algorithm.

While the results are encouraging, many directions are open to future research: First, the proposed algorithm can be extended by allowing multi-objective optimization scenarios in which many QoS attributes can be optimized simultaneously. Second, exploring the integration of machine learning techniques and metaheuristics may lead to advanced adaptability and efficiency in service composition strategies. Third, using the MFO-SFR algorithm in other domains, like IoT-enabled edge computing, hybrid cloud environments, or real-time service orchestration, is a promising avenue for further exploration. Finally, real-world cloud deployments of the proposed approach can shed light on many practical feasibility and scalability issues.

ACKNOWLEDGMENT

This work was supported by project of Chongqing Natural Science Foundation (No. CSTB2022NSCQ-MSX1298) and project of Research on Science and Technology of Chongqing Municipal Education Commission (No. KJZD-K202101901).

REFERENCES

- [1] S. S. Gill et al., "Modern computing: Vision and challenges," *Telematics and Informatics Reports*, p. 100116, 2024.
- [2] V. Hayyolalam, B. Pourghebleh, A. A. P. Kazem, and A. Ghaffari, "Exploring the state-of-the-art service composition approaches in cloud manufacturing systems to enhance upcoming techniques," *The International Journal of Advanced Manufacturing Technology*, vol. 105, no. 1-4, pp. 471-498, 2019.
- [3] V. Hayyolalam, B. Pourghebleh, M. R. Chehrehzad, and A. A. Pourhaji Kazem, "Single - objective service composition methods in cloud manufacturing systems: Recent techniques, classification, and future trends," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 5, p. e6698, 2022.
- [4] J. Zou, K. Wang, K. Zhang, and M. Kassim, "Perspective of virtual machine consolidation in cloud computing: a systematic survey," *Telecommunication Systems*, pp. 1-29, 2024.
- [5] H. Wu, "Black widow optimization algorithm for efficient task assignment in cloud computing," *Journal of Engineering and Applied Science*, vol. 71, no. 1, p. 139, 2024.
- [6] J. Alonso et al., "Understanding the challenges and novel architectural models of multi-cloud native applications—a systematic literature review," *Journal of Cloud Computing*, vol. 12, no. 1, p. 6, 2023.
- [7] D. Tohanean and S.-G. Toma, "The Impact of Cloud Systems on Enhancing Organizational Performance through Innovative Business Models in the Digitalization Era," in *Proceedings of the International Conference on Business Excellence*, 2024, vol. 18, no. 1: Sciendo, pp. 3568-3577.
- [8] J. DesLauriers, J. Kovacs, T. Kiss, A. Stork, S. P. Serna, and A. Ullah, "Automated generation of deployment descriptors for managing microservices-based applications in the cloud to edge continuum," *Future Generation Computer Systems*, vol. 166, p. 107628, 2025.
- [9] J. Lei, Q. Wu, and J. Xu, "Privacy and security-aware workflow scheduling in a hybrid cloud," *Future Generation Computer Systems*, vol. 131, pp. 269-278, 2022.
- [10] J. L. Schaefer et al., "A framework for diagnosis and management of development and implementation of cloud-based energy communities-Energy cloud communities," *Energy*, vol. 276, p. 127420, 2023.
- [11] F. K. Parast, C. Sindhav, S. Nikam, H. I. Yekta, K. B. Kent, and S. Hakak, "Cloud computing security: A survey of service-based models," *Computers & Security*, vol. 114, p. 102580, 2022.
- [12] M. Barakat, R. A. Saeed, and S. Edam, "A Comparative Study on Cloud and Edge Computing: A Survey on Current Research Activities and Applications," in *2023 IEEE 3rd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA)*, 2023: IEEE, pp. 679-684.
- [13] A. K. Samha, "Strategies for efficient resource management in federated cloud environments supporting Infrastructure as a Service (IaaS)," *Journal of Engineering Research*, vol. 12, no. 2, pp. 101-114, 2024.
- [14] S. Aleem, R. Batoool, S. Alkobaisi, F. Ahmed, and A. Khattak, "SaaS Application Maturity Assessment Model," *IEEE Access*, 2024.
- [15] H. U. Khan, F. Ali, and S. Nazir, "Systematic analysis of software development in cloud computing perceptions," *Journal of Software: Evolution and Process*, vol. 36, no. 2, p. e2485, 2024.
- [16] C. Li, M. Song, M. Zhang, and Y. Luo, "Effective replica management for improving reliability and availability in edge-cloud computing environment," *Journal of Parallel and Distributed Computing*, vol. 143, pp. 107-128, 2020.
- [17] A. Azadi and M. Momayez, "Review on Constitutive Model for Simulation of Weak Rock Mass," *Geotechnics*, vol. 4, no. 3, pp. 872-892, 2024, doi: <https://doi.org/10.3390/geotechnics4030045>.

- [18] M. A. Nezafat Tabalvandani, M. Hosseini Shirvani, and H. Motameni, "Reliability-aware web service composition with cost minimization perspective: a multi-objective particle swarm optimization model in multi-cloud scenarios," *Soft Computing*, vol. 28, no. 6, pp. 5173-5196, 2024.
- [19] W. Ma and H. Xu, "Skyline-enhanced deep reinforcement learning approach for energy-efficient and QoS-guaranteed multi-cloud service composition," *Applied Sciences*, vol. 13, no. 11, p. 6826, 2023.
- [20] B. Pourghebleh, A. A. Anvigh, A. R. Ramtin, and B. Mohammadi, "The importance of nature-inspired meta-heuristic algorithms for solving virtual machine consolidation problem in cloud environments," *Cluster Computing*, pp. 1-24, 2021.
- [21] M. B. Karimi, A. Isazadeh, and A. M. Rahmani, "QoS-aware service composition in cloud computing using data mining techniques and genetic algorithm," *The Journal of Supercomputing*, vol. 73, pp. 1387-1415, 2017.
- [22] S. S. Sefati and S. Halunga, "A hybrid service selection and composition for cloud computing using the adaptive penalty function in genetic and artificial bee colony algorithm," *Sensors*, vol. 22, no. 13, p. 4873, 2022.
- [23] M. Wang, "A new QoS-aware service composition technique in cloud computing using capuchin search algorithm," *Journal of Intelligent & Fuzzy Systems*, no. Preprint, pp. 1-12, 2023.
- [24] K. Zambouri and N. Jafari Navimipour, "A cloud service composition method using a trust - based clustering algorithm and honeybee mating optimization algorithm," *International Journal of Communication Systems*, vol. 33, no. 5, p. e4259, 2020.
- [25] X. Liu, "Hybrid Integrated Aquila Optimizer for Efficient Service Composition with Quality of Service Guarantees in Cloud Computing," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 10, 2023.
- [26] L. Bei, L. Wenlin, S. Xin, and X. Xibin, "An improved ACO based service composition algorithm in multi-cloud networks," *Journal of Cloud Computing*, vol. 13, no. 1, p. 17, 2024.
- [27] M. Shehab, L. Abualigah, H. Al Hamad, H. Alabool, M. Alshinwan, and A. M. Khasawneh, "Moth-flame optimization algorithm: variants and applications," *Neural Computing and Applications*, vol. 32, no. 14, pp. 9859-9884, 2020.

Enhanced Task Scheduling Algorithm Using Harris Hawks Optimization Algorithm for Cloud Computing

Fang WANG

Computer School, Hubei University of Education, Wuhan 430205, China

Abstract—Amongst the most transformational technologies nowadays, cloud computing can provide resources such as CPU, memory, and storage over secure internet connections. Due to its flexibility and resource availability with guaranteed QoS, cloud computing allows comprehensive business and research adoptions. Despite the rapid development, resource management remains one of the significant challenges, especially handling task scheduling efficiently in this environment. Task scheduling strategically assigns tasks to available resources so that Quality of Service (QoS) metrics are effectively related to response time and throughput. This paper proposes an Enhanced Harris Hawks Optimization (EHHO) algorithm for scheduling cloud tasks to mitigate the common limitations found in existing algorithms. EHHO integrates a dynamic random walk strategy, enhancing exploration capabilities to avoid premature convergence and significantly improving scalability and resource allocation efficiency. Simulation outcomes reveal that EHHO minimizes makespan by up to 75%, memory usage by up to 60%, execution time by up to 39%, and cost by up to 66% compared to state-of-the-art algorithms. These benefits demonstrate that EHHO can optimize resource allocation while being highly scalable and reliable. Consistent performance over various stacks such as Kafka, Spark, Flink, and Storm further evidences the superiority of EHHO in handling complex scheduling challenges in dynamic cloud computing environments.

Keywords—Cloud computing; optimization; task scheduling; Harris Hawks Optimization; resource allocation; quality of service

I. INTRODUCTION

The Internet of Things (IoT) symbolizes change whereby many devices, from simple sensors and actuators to various everyday objects, are connected via the Internet to communicate and share information [1]. Objects like sensors and actuators form communication grids in healthcare, manufacturing, and smart cities. However, with the growth of IoT applications, volumes of generated data are huge and require immense processing and colossal storage. More importantly, real-time analytics implies vast demand [2]. Thus, cloud computing has become the backbone of IoT systems for extendable resources and rugged data management capabilities beyond IoT devices [3, 4].

Cloud computing enables customers to use the services with the help of the Internet on a pay-per-usage basis [5]. Cloud services include within their ambit a broad range of services, namely Software as a Service (SaaS), Platform as a Service (PaaS), Communication as a Service (CaaS), Data storage as a Service (DaaS), and Infrastructure as a Service (IaaS) [6]. These services allow cloud providers to provide utility-based resources whose usage supports diversified needs for IoT.

The physical servers and switches in the backbone layer of cloud computing are operated and scaled by the cloud service provider effectively as per user requirements [7]. It efficiently allocates hardware resources at a blistering pace. In terms of software, the supervisor runs these hardware resources, a hypervisor, middleware, etc. [8]. The operating system implements hardware functionalities and develops user and application communication [9]. It allows the hypervisor to create Virtual Machines (VMs) on cloud servers with specified hardware configurations and software stacks [10]. This, in turn, enables a further increase in service availability because virtualization facilitates easy service migration even during hardware failures. It is also accompanied by a tremendous rise in hardware utilization compared to the non-virtualized environment [11]. Recent advancements in reinforcement learning applications, particularly in mobile robotics, have showcased its ability to enhance decision-making and resource management in dynamic environments. Similarly, reinforcement learning's adaptive capabilities, as demonstrated in SLAM tasks, highlight its potential for optimizing virtualization and scalability in cloud computing infrastructures [12].

The middleware arranges the running and interaction of tasks on cloud servers transparently. The three fundamental types of software infrastructure are PaaS, SaaS, and IaaS [13]. IaaS allows users to create multiple VMs on servers as needed, enhancing computational resource utilization. SaaS permits users to store and access unlimited amounts of data in a minute on remotely located servers. PaaS provides secure, reliable communication services and an application development platform accessible via APIs. Lastly, the application tier enables the user to use applications stored in the cloud through the Internet, allowing quick and easy access without installation or updates locally.

Effective task scheduling is vital for managing resource allocation, execution time, and QoS in cloud-supported IoT environments. Scheduling can be classified into two types: static and dynamic approaches. In static scheduling, tasks are assigned to available machines based on a predefined strategy, whereas in dynamic scheduling, instantaneous conditions are considered to adjust resource allocations. Real-time scheduling techniques ensure priority tasks with tight timing constraints.

Task scheduling in cloud computing is complex and an NP-hard problem, directly influencing the system performance regarding resource utilization, response time, and energy consumption. In this regard, we propose an Enhanced Harris Hawks Optimization (EHHO) algorithm to optimize task scheduling problems in cloud environments. EHHO features a

novel dynamic random walk to reinforce exploration and avoid premature convergence issues, enhancing scalability, resource allocation, and energy management.

By integrating these enhancements, EHHO provides a high-performance solution for complex scheduling requirements in cloud-supported IoT systems, contributing to more efficient, reliable, and cost-effective service delivery. The following sections detail EHHO's methodology, implementation, and performance advantages over existing algorithms, underscoring its potential as a leading approach to resource management in cloud computing.

The remainder of this paper is arranged as follows: Section 2 discusses related research and highlights gaps in existing scheduling approaches. Section 3 presents the problem statement and explains the challenges of cloud task scheduling. Section 4 presents the proposed algorithm. Section 5 summarizes the experimental results and performance analysis. Section 6 discusses the practical implications and challenges. Finally, Section 7 concludes the paper and suggests future directions.

II. RELATED WORK

Shukri, et al. [14] formulated an Enhanced Multi-Verse Optimizer (EMVO) to optimize task scheduling in cloud computing contexts. The developed algorithm incorporates a new mechanism to reserve the most optimal solution from each iteration and inject it back into the population after a predefined interval to leverage better exploration and exploitation capabilities. The proposed approach minimizes task execution time and considers factors like task length, cost, and power consumption. The combination of local and global search and the core components of MVO has caused EMVO to overcome the weaknesses inherent in traditional task scheduling algorithms. Comparisons with the original Particle Swarm Optimization (PSO) and MVO have revealed the efficiency of the proposed EMVO in decreasing the makespan while improving resource utilization.

Natesan and Chokkalingam [15] developed a new Mean Grey Wolf Optimization (MGWO) algorithm to solve cloud computing scheduling issues. The study aims to optimize energy consumption. MGWO performance was evaluated using the Cloudsim toolkit under baseline workload conditions. From the simulation results, it could be revealed that MGWO substantially outperforms competing algorithms in optimizing these crucial performance metrics.

Mapetu, et al. [16] proposed a new binary PSO algorithm to cope with cloud computing load and task scheduling issues. The suggested technique embraces a formula that minimizes the overall difference in execution time between different VMs while keeping some optimization criteria. A dedicated particle position updating strategy was adopted for enhanced load balancing. The numerical evidence verifies that the algorithm performs better than the previous meta-heuristic and heuristic approaches for optimizing load balancing and task scheduling.

Liu [17] developed an effective task scheduling approach using an adaptive Ant Colony Optimization (ACO) algorithm in cloud computing contexts. Pheromone adaptation is introduced into the procedure to accelerate convergence; thus,

prematurity can be reduced. In the cloud environment, a multi-objective optimization function, which minimizes cost and time for task execution, reduces load imbalance and maximizes resource utilization, is implemented by optimized ACO. It has been proved by comparison analysis that, compared with traditional ACO, the proposed approach can always guarantee better performance in solution quality, convergence speed, and overall system efficiency, especially in handling large-scale task scheduling challenges.

Zhou, et al. [18] presented a hybrid task scheduling method based on an improved Genetic Algorithm (GA) combined with a greedy algorithm. This algorithm was designed to converge on optimal solutions in lower iteration numbers of the search process than compared approaches. It aimed at response time, completion time, and QoS performance metrics. Experimental results demonstrate that hybrid GA performs much better than existing algorithms in task-scheduling optimization than existing algorithms.

Abualigah and Diabat [19] proposed, incorporating the combination of Ant Lion Optimization (ALO) adapted to the concept of Differential Evolution (DE) to address many-objective task-scheduling issues in cloud computing settings. Elite-based DE enhanced ALO's exploitation and exploration capability, saving it from premature convergence. The effectiveness of the suggested algorithm has been tested on modeled and real-world datasets using the Cloudsim simulation environment. Additionally, experiments proved that the hybrid ALO method outperformed the other optimization algorithms regarding continuous convergence rate, especially for large-scale scheduling problems.

Panda, et al. [20] introduced a new multi-paired task scheduling algorithm for cloud computing by utilizing the Hungarian algorithm. With this approach, the logic efficiently resolves unbalanced workloads based on the pairing strategy for task scheduling. The algorithm outperformed the Hungarian Algorithm with Converse Lease Time, the Hungarian Algorithm with Lease Time, and First-Come-First-Served baselines in large-scale simulations.

Tamilarasu and Singaravel [21] proposed an Improved Coati Optimization Algorithm (ICOA) for critical challenges in cloud computing, namely lengthy scheduling times, excessive costs, and unbalanced VM loads. A task distribution and scheduling scheme involving VMs, time, and cost, was developed. A dual-objective fitness function is employed to optimize resource utilization and makespan. In the ICOA, an exploitation strategy has been incorporated to prevent the solution from converging prematurely and, hence, to enhance local search capabilities. Simulation results demonstrated the superiority of the ICOA over conventional metaheuristic task scheduling algorithms at improving makespan, success rate, turnaround efficiency, and overall system availability.

Abualigah, et al. [22] offered an enhanced hybrid optimization algorithm for cloud task scheduling, which combines Jaya algorithm strengths with Synergical Swarm Optimization (SSO) and a Levy flight. This new approach efficiently balances exploration and exploitation to accelerate and prevent premature convergence. In integrating the best of Jaya and SSO, this algorithm uses their complementary

analytics capabilities to drive an optimal assignment of tasks and allocate resources. The experimental investigation against existing methodologies confirmed the algorithm's superior scalability, convergence speed, solution quality, and performance.

Behera and Sobhanayak [23] proposed a hybrid meta-heuristic approach using GA and Gravitational Search Algorithm (GSA) for multi-target optimization of task scheduling in cloud computing. The authors addressed the NP-hard challenge of efficiently managing an exponentially growing search space while enhancing system performance. The proposed approach leveraged strengths from GA and GSA in improving the Quality of Service (QoS) measures: energy consumption, resource utilization, and makespan. As tested under CloudSim with standard, real-time, and artificial workloads, it improved degree of imbalance by about 12%, resource utilization by 9%, response time by 7%, and energy consumption by 6%.

Khademi Dehnavi, et al. [24] proposed a hybrid GA for efficient and dependable task scheduling across heterogeneous cloud computing environments. The method models an NP-hard optimization problem in scheduling to minimize costs, time, and failures. HGA introduces two novel mutation and crossover operators in global search. It also implements a localized "Walking around" step to improve solutions. Simulation runs on twelve scenarios revealed significant cost reductions compared to state-of-the-art techniques: a 14.1%

reduction in makespan, an 18.7% monetary cost reduction, and a 42.3% decrease in failure cost.

Gong, et al. [25] introduced the Enhanced Marine Predator Algorithm (EMPA) for the task scheduling challenges in the cloud computing environment. This approach incorporates the operators of the Whale Optimization Algorithm (WOA) operators, nonlinear inertia weight coefficients, and Golden Sine strategies to minimize makespan while optimizing resource utilization. Simulation runs using synthetic and GoCJ datasets showed that EMPA outperformed GWO, SCA, PSO, and WOA in makespan, resource utilization, and degree of imbalance, positioning EMPA as a very effective scheduling solution in cloud environments.

Pabitha, et al. [26] suggested a new scheduling algorithm, the Chameleon and Remora Search Optimization Algorithm (CRSOA), to tackle the task-scheduling issues arising in cloud environments due to uncertain user demands. In this proposed technique, the Chameleon Search Algorithm (CSA) is combined with the Remora Search Optimization Algorithm (RSOA) to deliver an efficient resource utilization approach that takes into consideration parameters like MIPS and network bandwidth to ensure load balancing while imposing minimal scheduling cost and, at the same time, reduced makespan. The experimental results show that the makespan reduction achieved by CRSOA is 18.9%, cost reduction is 22.1%, and the improvement in load balancing is 20.5% against baseline metaheuristic algorithms.

TABLE I. AN OVERVIEW OF RECENT TASK SCHEDULING ALGORITHMS FOR CLOUD COMPUTING

References	Algorithm	Pros	Cons
[14]	Enhanced multi-verse optimizer	Efficiently reduces makespan and improves resource utilization through enhanced exploration and exploitation mechanisms.	Limited focus on real-time dynamic scheduling challenges.
[15]	Mean grey wolf optimization	Optimizes energy consumption and makespan effectively under baseline workloads.	Does not account for task heterogeneity or scalability in large datasets.
[16]	Binary particle swarm optimization	Superior load balancing with tailored particle updating strategies reduces execution time variance.	Lacks emphasis on cost-efficiency and energy consumption.
[17]	Adaptive ant colony optimization	Enhanced convergence rate and solution quality; minimized cost, execution time, and load imbalance.	Limited applicability for large-scale dynamic task scheduling.
[18]	Hybrid genetic algorithm with greedy	Faster convergence with improved QoS metrics such as response time and completion time.	Focuses primarily on search process efficiency, with limited exploration capabilities.
[19]	Ant lion optimization with differential evolution	Enhanced convergence rates for many-objective problems; robust against premature convergence.	Complexity increases the computational costs for large-scale scheduling tasks.
[20]	Multi-paired task scheduling	Effectively handles unbalanced workloads; superior in minimizing layover times.	Limited applicability to multi-objective or heterogeneous scheduling scenarios.
[21]	Improved coati optimization algorithm	Dual-objective optimization improves makespan and system availability and prevents premature convergence.	No explicit consideration of energy efficiency metrics.
[22]	Jaya with synergistic swarm optimization	Balances exploration and exploitation; achieves high solution quality and convergence speed.	Performance under real-time or uncertain environments is not evaluated.
[23]	Genetic algorithm and gravitational search algorithm	Improves energy consumption, makespan, and resource utilization; suitable for QoS optimization.	Focus on standard workloads with limited scalability for heterogeneous tasks.
[24]	Hybrid genetic algorithm	Cost-efficient with significant reductions in makespan, monetary cost, and failure cost.	Limited application to real-time dynamic and heterogeneous scheduling problems.
[25]	Enhanced marine predator algorithm	Superior makespan reduction, resource utilization, and imbalance handling.	Lack of scalability for highly complex or real-time cloud scheduling tasks.
[26]	Chameleon and remora search optimization algorithm	Effectively minimizes scheduling costs and makespan under uncertain user demands.	High computational complexity for large-scale environments.
[27]	Horse herd-squirrel search algorithm	Demonstrates significant advantages in cost, energy, and makespan reduction.	Limited evaluation under multi-cloud or distributed cloud environments.

Parthasaradi, et al. [27] proposed a hybrid meta-heuristic for cloud computing task scheduling; the Horse Herd–Squirrel Search Algorithm (HO–SSA). This protocol combined SSA and the Horse Herd Optimization Algorithm (HOA) to increase cost efficiency, energy utilization, and scheduling performance. Furthermore, the proposed HO–SSA showed significant superiority and reduced up to 22.2% regarding tasks' cost scheduling costs, 9.68% regarding energy consumption, and makespan compared with SSA, HOA, and TSA.

As summarized in Table I, recent task-scheduling algorithms excel at optimizing specific metrics such as makespan, resource utilization, or cost. However, deficiencies remain in addressing scalability, real-time scheduling, and energy efficiency in heterogeneous and dynamic cloud environments. Although algorithms like EMPA have proven efficient in resource utilization, and huge cost reductions have been achieved in HO-SSA, few have provided a balanced approach to large-scale and multi-cloud comprehensive optimization problems for energy efficiency, load balancing,

and QoS. Furthermore, most of those methods lack real-time adaptability to uncertain user demands. Under such gaps, the proposed algorithm operates under an integrated dynamic exploration and exploitation strategy, aiming for optimal resource allocation scalability while enhancing performance in various cloud environments.

III. PROBLEM STATEMENT

Scheduling tasks in cloud computing environments is critical for effective and efficient execution, whereby resources are assigned according to user requests. Multiple layered architectures have been developed in cloud computing to offer these utility-based services. Fig. 1 depicts this kind of layered architecture. Each layer addresses specific functionalities, from data storage and processing to application development and communication support, and enables IoT applications to operate efficiently without major investments in local infrastructure. Table II provides a list of abbreviations and symbols used throughout the paper.

TABLE II. SYMBOLS AND DEFINITIONS

Symbol	Description	Symbol	Description
T	Cloud tasks	c_{ij}	Binary variable indicating task i assigned to virtual machine j
V	Cloud virtual machines	x_{ij}	Association between a virtual machine and a task
n	Total number of tasks	LB	Lower bound of the solution space
m	Total number of virtual machines	CT	Convergence time
VR	Collection of virtual machine resources	J	Randomization factor
$MIPS$	Millions of instructions per second capability of a CPU	$X_m(i)$	Updated position after applying random walk strategy
CU_j	Compute units capacity of the j^{th} virtual machine	$O(x)$	Objective function for optimization x
L_i	Task duration for the i^{th} task	$x(t+1)$	Position of a hawk in the next iteration
ET_{ij}	Execution time for the i^{th} task on the j^{th} virtual machine	$x_{random}(t)$	Random position of a hawk
BT_j	Busy period of the j^{th} virtual machine	$x_{rabbit}(t)$	Position of the prey
E	Rabbit's escaping energy	$x_{mean}(t)$	Average position of the hawk population
t	Current iteration number	UB	Upper bound of the solution space
Max_iter	Maximum number of iterations	c	Random walk deviation control constant
$\Delta x(t)$	Difference between prey and hawk positions	$rand$	Random number between 0 and 1

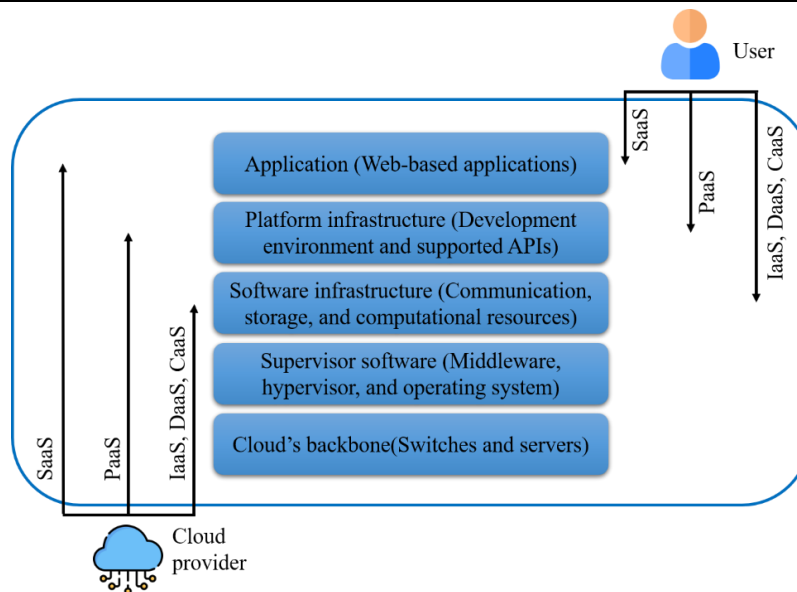


Fig. 1. Multi-layer design of cloud computing.

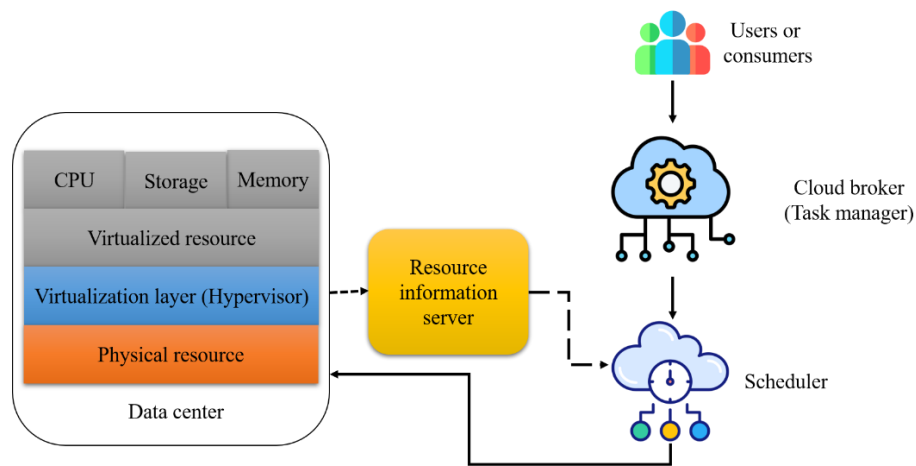


Fig. 2. Proposed framework for task scheduling.

The quality of the service will be directly affected by the following scheduling algorithm, affecting parameters such as execution time and operational costs. The existing frameworks comprise a cloud broker and Resource Information Servers (RIS) to ensure optimum scheduling and provide runtime information about resource availability and VM capabilities. While the above systems consolidate data from physical and virtualized infrastructures, significant research gaps exist in scheduling. A schematic of this framework is provided in Fig. 2.

Most traditional approaches fail to balance exploration and exploitation and thus lead to premature convergence or suboptimal resource allocation in a dynamic, real-time environment. In addition, most methodologies cannot adapt to heterogeneous workloads or consider uncertain factors such as fluctuating resource demand and VM performance. Other multi-objective optimizations, like minimal makespan, energy consumption, cost, and maximal resource utilization, have also been inadequately performed by most current strategies.

These gaps highlight the need for advanced algorithms capable of dynamic decision-making, enhanced exploration of solution spaces, and robust handling of diverse workloads. The EHHO algorithm addresses these challenges by integrating dynamic random walk strategies and stochastic adjustments to produce superior task scheduling performance, ensuring scalability and efficiency in complex cloud environments.

Cloud data centers feature an extensive range of actual machines containing functional VMs. The VMs function as the underlying infrastructure for the execution of user tasks. Tasks assigned to a particular VM are based on the task's requirements. Two alternative concepts of scheduling are common inside the cloud environment. The initial step involves identifying and allocating servers specifically intended for supporting VMs. The specific scheduling variant significantly enhances data center productivity, reduces power usage, and optimizes resource utilization. The impact of such a phenomenon is notably significant on cloud service vendors' operational activities.

On the other hand, the second classification of scheduling is concerned with assigning VMs for task execution. It is common

practice to divide large tasks into separate components and assign each one to a separate VM for execution under the virtualization setup. In the present scenario, the choice of VMs is contingent upon users' particular service requirements and the current condition of VMs. The implications of this specific scheduling method substantially affect the duration of job completion and the financial expenditure related to task execution. The scheduling paradigm exhibits a notable resonance among users, particularly concerning service quality and budgetary factors.

The responsibility for coordinating user tasks onto VMs based on user requirements and QoS factors usually falls on the data center broker and the cloud information service in a cloud computing environment. They are crucial in ensuring that user tasks are allocated to suitable VMs that meet the desired performance, resource availability, and other criteria. This coordination helps optimize resource utilization and deliver efficient cloud services. Users prefer minimizing costs associated with service expenses, whereas cloud providers strive to reduce energy consumption while maintaining optimal server performance and capacity utilization. These issues arise from the direct impact of these elements on the time of task performance.

As the duration of a task lengthens, there is a corresponding rise in cost expenditures and energy use. Therefore, the primary focus of this scholarly inquiry is the reduction of makespan. As has been previously analyzed, the complex issue of task scheduling is classified as one of the NP-hard issues. Despite the notable effectiveness of evolutionary algorithms in addressing NP-hard problems, their convergence rate tends to be prolonged due to the exhaustive examination of all probable, plausible solutions. As a result, the prompt achievement of convergence is regarded as a subordinate goal in this research. Given the presence of m VMs and n tasks within the environment, a set of tasks (T) and VMs (V) can be expressed by Eq. (1) and (2).

$$T = \{t_1, t_2, t_3, \dots, t_n\} \quad (1)$$

$$V = \{v_1, v_2, v, \dots, v_m\} \quad (2)$$

The assignment of these tasks to VMs produces a significant number of potential patterns, which can be expressed as nm possible scenarios. Suppose VR represents a collection of VM resources, indicated as $VR = (vr_1, vr_2, vr_3, \dots, vr_k)$. These features include the central processing unit's (CPU) capability to execute Millions of Instructions Per Second (MIPS), the availability of bandwidth, the capacity of Random Access Memory (RAM), and the capacity of storage. The task completion duration depends on the specific allocation of resources to the selected VMs. Cloud service providers utilize specific measurements known as Compute Units (CUs) to measure the capacity of VMs. For example, a solitary Amazon CU possesses processing capabilities that align with a frequency range of up to 1.2 GHz, similar to an Xeon and Opteron processor. The calculation of the duration of task execution and the projected operational expenditure for the workflow depends on CUs. Therefore, the execution time for the i^{th} task is given in the form of Eq. (3).

$$ET_{ij} = \frac{L_i}{CU_j} \quad (3)$$

Where i and j represent indices within integer numbers sets ranging from 1 to n and 1 to m , respectively. Here, CU_j corresponds to the computational unit associated with the j^{th} VM, whereas L_i signifies the execution time of the i^{th} task. The active time of a VM is defined as the interval during which tasks are being processed on the VM. Specifically, the phase of intense utilization throughout the active time of the j^{th} VM is represented by Eq. (4).

$$BT_j = \sum_{i=1}^n ET_{ij} \times c_{ij} \quad (4)$$

c_{ij} is limited to binary values, 0 or 1, x_{ij} represents the relationship between tasks and VMs, where 1 implies that task t_i is allocated to the j^{th} VM. Since VMs operate concurrently, the workflow's overall duration, or makespan, is calculated by the most prolonged time any single VM remains occupied. The makespan can be expressed using Eq. (5).

$$M = \max(BT_j) \quad (5)$$

Evolutionary algorithms aim to find the optimal solution by systematically navigating the problem domain. The time needed for the algorithm to converge depends on the solution space properties and the number of iterations executed. As the solution space expands, the convergence time increases. This relationship between convergence time and the solution space size can be expressed mathematically by Eq. (6).

$$CT \propto (l_x, k) \quad (6)$$

CT refers to the convergence time, k denotes the number of iterations necessary to identify an optimal solution, and l_x represents the length of the optimal solution x . The objective function O , used to determine the solution x , can be formulated based on Eq. (3) and (4) in form of Eq. (7).

$$O(x) = \min(M), \min(CT) \quad (7)$$

IV. PROPOSED METHOD

Harris Hawk cooperative hunting and tracking procedures inspire the HHO algorithm. These birds employ strategic tactics of surprise jumps and seven killings to capture their prey. In

cooperative attacks, some hawks coordinate in pursuit of a rabbit that has exposed itself after revealing its whereabouts for pursuit and quick capture. With hunting, however, there would be successive quick dives next to the prey, based on how it would react and the chance of its fleeing. Harris's hawks have various hunting techniques under their wings, each for different circumstances and different maneuvers of prey to evade them. If the top hawk in a hunting activity fails to track the rabbits, then another member of the team should replace that hawk and foil possible escape. It is here that the rabbit, once the hunt starts, cannot regain its defense mechanism, and the team's combined effort prevents it from escaping. The most experienced hawk makes the final catch of exhausted prey to share among the team members.

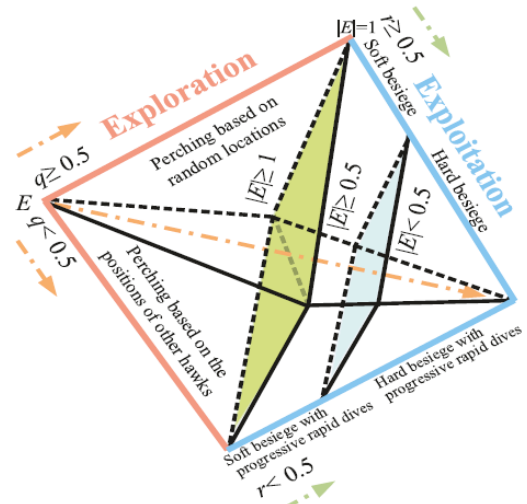


Fig. 3. HHO steps.

Fig. 3 visually represents the different phases of the HHO algorithm and reflects hawk predatory behavior: locating, circling, and ultimately capturing prey. HHO's mathematical formulation is structured accordingly, including the exploration, transition, and exploitation phases. Within this conceptual framework, each Harris's hawk symbolizes a possible solution to a particular problem, while the target prey symbolizes the ideal solution to be identified. Falcons use two exploration strategies to find prey. In the first strategy, hawks choose locations according to other hawks' positions and prey locations. In the second tactic, hawks sit randomly on tall trees. Eq. (8) mathematically models these two exploration methods with equal probability and uses random numbers to simulate their occurrence.

$$x(t+1) = \begin{cases} x_{random}(t) - r_1 |x_{random}(t) - 2r_2 x| & q \geq 0.5 \\ x_{rabbit}(t) - x_{mean}(t) - r_3 (LB + r_4 (UB - LB)) & q < 0.5 \end{cases} \quad (8)$$

Eq. (9) calculates the average position of the hawk population. The algorithm dynamically transitions between exploration and exploitation phases based on a metric termed 'rabbit energy,' defined by Eq. (10). When the rabbit's escaping energy $|E|$ exceeds 1, the hawks engage in a more extensive exploration of the search space; otherwise, the algorithm

transitions to the exploitation phase. Eq. (11) to (14) establish whether the hawks execute a soft siege or a hard siege, depending on the rabbit's energy level and its likelihood of escape. In a soft siege, the hawks simulate the rabbit's successful escape by performing repetitive diving maneuvers. Conversely, a hard siege employs a distinct computational strategy to model the scenario.

$$x_{mean}(t) = \frac{1}{N} \sum_{i=1}^N x_i(t) \quad (9)$$

$$E = 2E_0 \left(1 - \frac{t}{Max_iter}\right) \quad (10)$$

$$x(t+1) = \Delta x(t) - E|J \cdot x_{rabbit}(t) - x(t)| \quad (11)$$

$$\Delta x(t) = x_{rabbit}(t) - x(t) \quad (12)$$

$$J = 2(1 - random) \quad (13)$$

$$x(t+1) = x(t) - E|\Delta x(t)| \quad (14)$$

Eq. (15) to (18) regulate the rapid dives employed during the soft siege, employing Lévy movements to simulate the prey's evasive behavior. Eq. (15) and (16) mathematically model the hawks' actions during the diving phase. Subsequently, Eq. (17) and (18) define the characteristics of the

final rapid dives performed during the soft siege and the associated factors, k and z , utilized throughout the hard siege phase.

$$k = x_{rabbit}(t) - E|J \cdot x_{rabbit}(t) - x(t)| \quad (15)$$

$$z = k + RandomVector.L(dim) \quad (16)$$

$$x(t+1) = \begin{cases} k & \text{if } f(k) < f(x(t)) \\ z & \text{if } f(z) < f(x(t)) \end{cases} \quad (17)$$

$$k = x_{rabbit}(t) - E|J \cdot x_{rabbit}(t) - x_{mean}(t)| \quad (18)$$

The HHO algorithm incorporates four pursuit strategies during the exploitation phase to enhance exploration capabilities. While heightened exploration is beneficial in identifying diverse solution spaces, it can inadvertently precipitate premature convergence and local optima. To counteract this, standard stochastic strategies such as Gaussian random walk, Brownian motion, and Levy flight are often integrated into optimization algorithms. These strategies introduce controlled stochasticity, allowing the algorithm to balance exploitation with exploration. By generating random deviations, these methods keep the algorithm from becoming stuck in suboptimal solutions and boost its overall performance.

Initialize parameters:

- Generate an initial population of hawks with random positions.
- Define maximum iterations (Max_iter), iteration counter ($t = 1$), and necessary constants.
- Define the fitness function for the optimization problem.

Evaluate initial fitness:

- Calculate the fitness of each hawk in the initial population.
- Identify the prey position (the current best solution based on fitness).

While $t \leq Max_iter$, perform the following steps:**a. Calculate Rabbit Energy (E):**

Compute $E = 2E_0(1 - t/Max_iter)$, where E_0 is a random value in $[-1,1]$.

b. Determine phase:

- If $|E| > 1$, proceed with the exploration phase.
- Otherwise, proceed with the exploitation phase.

c. Exploration phase:

- Update the hawks' positions using:
 - Strategy 1: Update positions based on other hawks and the prey location.
 - Strategy 2: Allow hawks to perch randomly on tall trees.
- Use probabilistic rules (Equation 8) to determine the update strategy.

d. Exploitation phase:

- If prey escapes (soft siege):
 - Simulate evasive behavior using rapid dives with Lévy flights (Equations 15-18).
- Else (hard siege):
 - Aggressively update positions based on prey location (Equations 11-14).

e. Dynamic random walk (if fitness stagnates):

- Compute the deviation using Equation 19.
- Update the hawks' positions using Equation 20.

f. Update fitness and prey position:

- Recalculate the fitness for the updated hawk positions.
- Update the prey position if a better solution is found.

g. Increment iteration:

- Increase t by 1.

Output the results:

- Return the best position (prey) and its corresponding fitness value.
-

Fig. 4. Pseudocode of the proposed algorithm.

The paper proposes a dynamic random walk strategy to enhance the HHO algorithm. The pseudocode of the proposed algorithm is depicted in Fig. 4. The magnitude of the random walk deviation decreases over time. This ensures a balance between exploration (larger deviations in early iterations) and exploitation (smaller deviations in later iterations). The random walk is activated only when the fitness value of a hawk remains unchanged compared to the previous iteration. This indicates potential stagnation in the search process. The deviation is calculated using a time-dependent formula involving a random number and the present iteration relative to the maximum number of iterations. The proposed random walk strategy can be mathematically expressed using Eq. (19).

$$Deviation = (c \times rand - c/2) \times \cos(\pi/2 \times (t/T)) \quad (19)$$

Where *Deviation* is the value added to the hawk's position, *c* is a constant controlling the maximum deviation, *rand* is a random number between 0 and 1, *t* is the ongoing iteration, and *T* is the total number of iterations. Eq. (20) is used to model the process.

$$x_m(i) = X(i) + \left(c \times rand - \frac{c}{2}\right) \times \cos\left(\frac{\pi}{2} \times \left(\frac{t}{T}\right)^2\right) \times (X(i) - X_{rabbit}) \quad (20)$$

Experimental results indicate that a value of *c* equal to six yielded optimal performance. Applying the random walk strategy produces a novel position, denoted as $X_m(i)$. A subsequent greedy selection process, as formalized in Eq. (21), determines the most suitable position for the ensuing iteration.

$$X(t+1) = \begin{cases} X_m(t+1), & f(X_m(t+1)) < f(X(t+1)) \\ X(t+1), & f(X_m(t+1)) \geq f(X(t+1)) \end{cases} \quad (21)$$

V. RESULTS

The proposed algorithm (EHHO) algorithm was simulated using the CloudSim toolkit, which offers robust support for on-demand resource provisioning and versatile features, including multi-objective optimization, dynamic resource scaling, application modeling, and cloud deployment simulation. Kafka's built-in load-balancing mechanism was employed. To evaluate EHHO's performance, it was compared against ALO, GA, ACO, PSO, MGWO, and EMVO algorithms using metrics such as execution time, cost, memory storage, and makespan. Experimental parameters are detailed in Table III.

The platform selection for evaluating the EHHO algorithm, including Kafka, Spark, Flink, and Storm, was driven by their unique characteristics that align with the requirements of task scheduling in cloud environments. Kafka was chosen for its real-time reporting capabilities, enterprise-level security, and efficient cloud monitoring, making it ideal for scenarios requiring immediate feedback and load balancing. Spark's in-memory computation and scalability enable high-speed processing for large datasets, while Flink's event-driven architecture supports dynamic and continuous task scheduling. Storm, known for its low-latency processing, is particularly suitable for time-critical scheduling tasks. These platforms were selected to demonstrate EHHO's adaptability and performance across workloads, real-time requirements, and

resource management conditions, ensuring comprehensive evaluation in diverse cloud scenarios.

Tables IV and V show the execution time and cost results for different algorithms and platforms. EHHO consistently demonstrated superior performance, achieving the lowest execution time (610 ms) and cost (60) on the Kafka platform. Tables VI and VII summarize memory storage and makespan results, with EHHO again exhibiting optimal performance, recording minimum makespan values and memory consumption across all platforms. To ensure a fair comparison, all algorithms employed a maximum iteration of 100 and a population size of 100. Specific parameter settings for each algorithm are detailed below:

- EHO: alpha = 0.5, beta = 1, upper bound = 0.9, number of clans = 10, set elitism = 2, lower bound = 0.3.
- MGWO and EMVO: number of appliances = 12, coefficient vector = 1, TDR = 1, WEP = 0.2.
- PSO: maximum initial velocity = 15, minimum initial velocity = 5, alpha = 0.8, beta = 0.8.
- ACO: time factor = 2, saving matrix factor = 2, visibility coefficient = 3, pheromone concentration coefficient = 1.
- GA: mutation probability = 0.02, crossover probability = 0.60, number of demes = 6.
- ALO: number of dimensions = 5, lower bound = 0.1, upper bound = 0.8.

Kafka consistently outperforms other platforms regarding cost, execution time, makespan, and memory storage for all algorithms. Its real-time reporting capabilities, enterprise-level security, efficient cloud monitoring, and superior processing speed contributed to these results. Fig. 5 to 8 provide visual representations of the comparative performance of the algorithms as measured by cost, execution time, memory storage, and makespan, respectively. The simulations validate the superiority of the EHHO algorithm in optimizing resource allocation and performance across various metrics and platforms. Its ability to effectively balance workload and resource utilization resulted in significant improvements compared to traditional optimization algorithms.

TABLE III. SIMULATION PARAMETERS

Element	Parameter	Value
Task	Task length	1000
	Task count	1000
VM	Service provider count	5
	VM count	1000
	MIPS	500
	Bandwidth	500
	Processing element count	2
Datacenter	Datacenter count	10
	Host count	2

TABLE IV. SIMULATION RESULTS FOR COST

Platform	EMVO	MGWO	PSO	ACO	GA	ALO	EHHO
Kafka	151	150	123	178	174	155	60
Spark	175	165	139	189	178	170	79
Flink	186	174	145	202	184	192	85
Storm	191	187	153	190	191	204	101

TABLE V. SIMULATION RESULTS FOR EXECUTION TIME

Platform	EMVO	MGWO	PSO	ACO	GA	ALO	EHHO
Kafka	835	893	792	785	911	774	610
Spark	897	946	862	888	1080	803	649
Flink	906	956	874	901	1123	875	716
Storm	964	979	889	909	1201	895	727

TABLE VI. SIMULATION RESULTS FOR MEMORY USAGE

Platform	EMVO	MGWO	PSO	ACO	GA	ALO	EHHO
Kafka	502	421	530	443	398	382	305
Spark	531	488	631	555	479	457	317
Flink	690	548	659	630	525	536	332
Storm	696	571	722	730	840	514	336

TABLE VII. SIMULATION RESULTS FOR MAKESPAN

Platform	EMVO	MGWO	PSO	ACO	GA	ALO	EHHO
Kafka	109	120	140	124	210	240	52
Spark	121	146	164	142	231	243	55
Flink	140	156	175	185	275	275	82
Storm	143	187	184	191	281	286	94

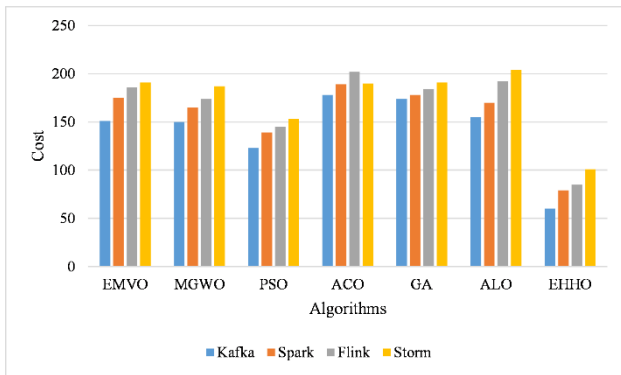


Fig. 5. Cost comparison.

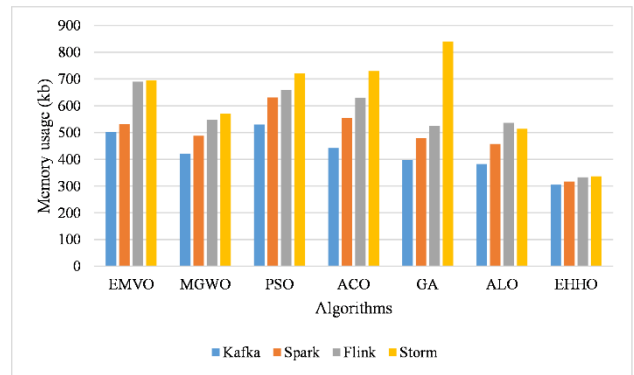


Fig. 7. Memory usage comparison.

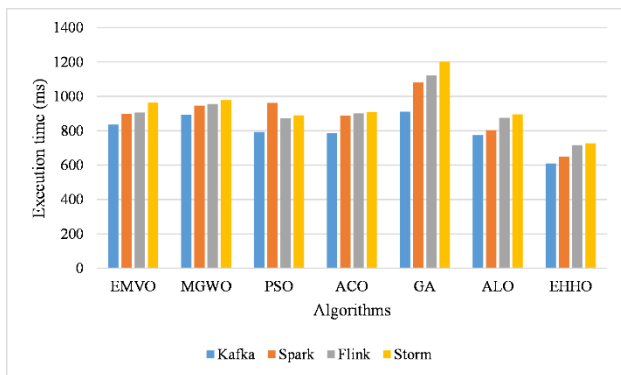


Fig. 6. Execution time comparison.

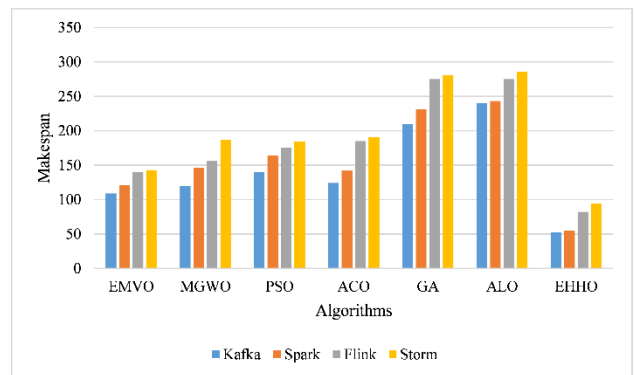


Fig. 8. Makespan comparison.

VI. DISCUSSION

The EHHO algorithm demonstrates significant advancements in task scheduling within cloud computing environments. Integrating a dynamic random walk strategy has notably improved the algorithm's exploration and exploitation power, leading to superior performance in various metrics compared to other optimization algorithms. The experimental findings reveal that EHHO consistently achieves lower execution times, costs, memory usage, and makespan across multiple platforms, including Kafka, Spark, Flink, and Storm. These findings underscore the robustness and efficiency of EHHO in optimizing resource allocation and handling complex scheduling problems in cloud computing.

A key factor contributing to EHHO's success is its ability to avoid premature convergence, a common issue in traditional meta-heuristic algorithms. By incorporating stochastic strategies such as Gaussian random walk, Brownian motion, and Levy flight, EHHO maintains equilibrium between global exploration and local exploitation. This balance ensures that the algorithm can explore diverse solution spaces without falling into a local optimum, thereby enhancing solution quality. The dynamic adjustment of the random walk deviation over time further refines this balance, enabling EHHO to effectively adapt to different stages of the optimization process.

Moreover, the simulation results highlight the exceptional functionality of the Kafka platform concerning makespan, execution time, cost, and memory usage. Kafka's real-time reporting capabilities, enterprise-level security, efficient cloud monitoring, and superior processing speed contribute to these outcomes. These characteristics make Kafka a suitable environment for deploying EHHO, allowing it to fully leverage its optimization potential. The comparative analysis with other platforms reinforces the importance of selecting an appropriate infrastructure to maximize the benefits of advanced optimization algorithms like EHHO in cloud computing.

In summary, the EHHO algorithm effectively responds to the complex task scheduling challenges in cloud computing. Its enhanced exploration and exploitation mechanisms, coupled with the optimal performance on platforms like Kafka, position EHHO as a leading approach for efficient resource management. Researchers could explore ways to improve the EHHO algorithm, such as integrating additional stochastic strategies or refining the random walk parameters, to achieve even greater performance improvements. Additionally, investigating the algorithm's scalability and applicability to other optimization problems could expand its utility in broader contexts.

The EHHO algorithm can seamlessly integrate with popular cloud services such as AWS, Azure, and Google Cloud to optimize task scheduling and resource management. By leveraging these platforms' capabilities, EHHO can enhance the efficiency of IaaS by dynamically allocating VMs and managing compute resources. In PaaS, EHHO can streamline application deployments by optimizing workload distribution across scalable infrastructure. For SaaS, the algorithm ensures reduced latency and cost-effective resource utilization, improving overall service delivery. The ability of EHHO to

adapt to real-time cloud environments and balance workloads makes it a crucial component for maximizing the performance and scalability of cloud-based services, further solidifying its relevance in modern cloud computing ecosystems.

Despite its promising performance in task scheduling, the EHHO algorithm has certain constraints. Its reliance on predefined parameters, such as random walk deviation and iteration limits, may limit adaptability across varying real-time scenarios and dynamic workloads. Additionally, while EHHO demonstrates superior results on metrics like makespan, cost, and memory usage, its scalability to handle significantly larger task datasets or highly heterogeneous environments remains untested. The simulations, primarily conducted using the Kafka platform, suggest a dependency on specific infrastructure capabilities such as real-time reporting and efficient monitoring, raising concerns about performance consistency on less advanced platforms. Furthermore, while the dynamic random walk strategy improves exploration and exploitation, fine-tuning these adjustments for broader applications remains challenging. Addressing these constraints, particularly scalability and infrastructure independence, will be critical for maximizing EHHO's potential in diverse cloud environments.

VII. CONCLUSION

Effective task scheduling is paramount to the optimal performance of cloud computing systems. Unlike traditional computing environments, cloud-based task scheduling necessitates considering diverse parameters, including computational costs, processing capabilities, and task duration. In this research, we introduced the EHHO algorithm to tackle the complex challenge of task scheduling in cloud computing environments. Leveraging the CloudSim toolkit for simulations, EHHO demonstrated superior performance over traditional algorithms like PSO, ACO, GA, ALO, MGWO, and EMVO across critical metrics, including cost, execution time, makespan, and memory storage. Integrating a random walk approach significantly improved the algorithm's exploration capabilities, effectively preventing premature convergence to local optima and ensuring more efficient resource allocation. With its robust load balancing, high security, real-time analysis, and scalability, Kafka's platform further highlighted the algorithm's efficiency. Our findings underscore EHHO's potential for optimizing the operational efficiency of cloud computing systems, making it a viable solution for better task scheduling and resource management in diverse and dynamic cloud environments.

Future research on the EHHO could include focusing on its scalability and adaptability given real-time scheduling scenarios in highly dynamic cloud environments. By integrating adaptive random walk strategies, deviation parameters can dynamically change depending on task complexity and resource availability in real-time. The following extension of EHHO for multi-cloud or hybrid cloud infrastructures with cross-platform scheduling and resource allocation would increase its applicability. Testing its performance with more diverse and larger datasets and optimization of computational efficiency for real-world runtime applications may position EHHO as a more robust and versatile solution to complex challenges in cloud computing.

REFERENCES

- [1] B. Pourghebleh and N. J. Navimipour, "Data aggregation mechanisms in the Internet of things: A systematic review of the literature and recommendations for future research," *Journal of Network and Computer Applications*, vol. 97, pp. 23-34, 2017, doi: <https://doi.org/10.1016/j.jnca.2017.08.006>.
- [2] M. Elrifae, T. Zayed, E. Ali, and A. H. Ali, "IoT contributions to the safety of construction sites: a comprehensive review of recent advances, limitations, and suggestions for future directions," *Internet of Things*, p. 101387, 2024.
- [3] V. Hayyolalam, B. Pourghebleh, A. A. P. Kazem, and A. Ghaffari, "Exploring the state-of-the-art service composition approaches in cloud manufacturing systems to enhance upcoming techniques," *The International Journal of Advanced Manufacturing Technology*, vol. 105, no. 1-4, pp. 471-498, 2019.
- [4] Q. Li, J. Huang, S. Li, and C. Huang, "A Sustainable Data Encryption Storage and Processing Framework via Edge Computing-Driven IoT," *Engineering Letters*, vol. 32, no. 7, 2024.
- [5] V. Hayyolalam, B. Pourghebleh, M. R. Chehrezad, and A. A. Pourhaji Kazem, "Single - objective service composition methods in cloud manufacturing systems: Recent techniques, classification, and future trends," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 5, p. e6698, 2022.
- [6] V. Hayyolalam, B. Pourghebleh, and A. A. Pourhaji Kazem, "Trust management of services (TMOs): investigating the current mechanisms," *Transactions on Emerging Telecommunications Technologies*, vol. 31, no. 10, p. e4063, 2020.
- [7] A. Mohamed et al., "Software-defined networks for resource allocation in cloud computing: A survey," *Computer Networks*, vol. 195, p. 108151, 2021.
- [8] L. Rosa, L. Foschini, and A. Corradi, "Empowering Cloud Computing With Network Acceleration: A Survey," *IEEE Communications Surveys & Tutorials*, 2024.
- [9] C. Wang and D. Wang, "Managing the integration of teaching resources for college physical education using intelligent edge-cloud computing," *Journal of Cloud Computing*, vol. 12, no. 1, p. 82, 2023.
- [10] R. Zolfaghari, A. Sahafi, A. M. Rahmani, and R. Rezaei, "Application of virtual machine consolidation in cloud computing systems," *Sustainable Computing: Informatics and Systems*, vol. 30, p. 100524, 2021.
- [11] T. Sun, C. Ma, Z. Li, and K. Yang, "Cloud Computing-based Parallel Deep Reinforcement Learning Energy Management Strategy for Connected PHEVs," *Engineering Letters*, vol. 32, no. 6, 2024.
- [12] M. D. Tezerjani, M. Khoshnazar, M. Tangestanizadeh, and Q. Yang, "A Survey on Reinforcement Learning Applications in SLAM," *arXiv preprint arXiv:2408.14518*, 2024, doi: <https://doi.org/10.48550/arXiv.2408.14518>.
- [13] B. Pourghebleh, A. A. Anvigh, A. R. Ramtin, and B. Mohammadi, "The importance of nature-inspired meta-heuristic algorithms for solving virtual machine consolidation problem in cloud environments," *Cluster Computing*, pp. 1-24, 2021.
- [14] S. E. Shukri, R. Al-Sayyed, A. Hudaib, and S. Mirjalili, "Enhanced multi-verse optimizer for task scheduling in cloud computing environments," *Expert Systems with Applications*, vol. 168, p. 114230, 2021.
- [15] G. Natesan and A. Chokkalingam, "Task scheduling in heterogeneous cloud environment using mean grey wolf optimization algorithm," *ICT Express*, vol. 5, no. 2, pp. 110-114, 2019.
- [16] J. P. B. Mapetu, Z. Chen, and L. Kong, "Low-time complexity and low-cost binary particle swarm optimization algorithm for task scheduling and load balancing in cloud computing," *Applied Intelligence*, vol. 49, pp. 3308-3330, 2019.
- [17] H. Liu, "Research on cloud computing adaptive task scheduling based on ant colony algorithm," *Optik*, vol. 258, p. 168677, 2022.
- [18] Z. Zhou, F. Li, H. Zhu, H. Xie, J. H. Abawajy, and M. U. Chowdhury, "An improved genetic algorithm using greedy strategy toward task scheduling optimization in cloud environments," *Neural Computing and Applications*, vol. 32, pp. 1531-1541, 2020.
- [19] L. Abualigah and A. Diabat, "A novel hybrid antlion optimization algorithm for multi-objective task scheduling problems in cloud computing environments," *Cluster Computing*, vol. 24, no. 1, pp. 205-223, 2021.
- [20] S. K. Panda, S. S. Nanda, and S. K. Bhoi, "A pair-based task scheduling algorithm for cloud computing environment," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 1, pp. 1434-1445, 2022.
- [21] P. Tamilarasu and G. Singaravel, "Quality of service aware improved coati optimization algorithm for efficient task scheduling in cloud computing environment," *Journal of Engineering Research*, 2023.
- [22] L. Abualigah et al., "Improved Jaya Synergistic Swarm Optimization Algorithm to Optimize Task Scheduling Problems in Cloud Computing," *Sustainable Computing: Informatics and Systems*, p. 101012, 2024.
- [23] I. Behera and S. Sobhanayak, "HTSA: A novel hybrid task scheduling algorithm for heterogeneous cloud computing environment," *Simulation Modelling Practice and Theory*, vol. 137, p. 103014, 2024.
- [24] M. Khademi Dehnavi, A. Broumandnia, M. Hosseini Shirvani, and I. Ahanian, "A hybrid genetic-based task scheduling algorithm for cost-efficient workflow execution in heterogeneous cloud computing environment," *Cluster Computing*, pp. 1-26, 2024.
- [25] R. Gong, D. Li, L. Hong, and N. Xie, "Task scheduling in cloud computing environment based on enhanced marine predator algorithm," *Cluster Computing*, vol. 27, no. 1, pp. 1109-1123, 2024.
- [26] P. Pabitha, K. Nivitha, C. Gunavathi, and B. Panjavarnam, "A chameleon and remora search optimization algorithm for handling task scheduling uncertainty problem in cloud computing," *Sustainable Computing: Informatics and Systems*, vol. 41, p. 100944, 2024.
- [27] V. Parthasaradi, A. Karunamurthy, C. Hussaian Basha, and S. Senthilkumar, "Efficient Task Scheduling in Cloud Computing: A Multiobjective Strategy Using Horse Herd-Squirrel Search Algorithm," *International Transactions on Electrical Energy Systems*, vol. 2024, no. 1, p. 1444493, 2024.

PCE-BP: Polynomial Chaos Expansion-Based Bagging Prediction Model for the Data Modeling of Combine Harvesters

Liangyi Zhong¹, Mengnan Deng², Maolin Shi^{3*}, Ting Lou⁴, Shaoyang Zhu⁵, Jingwen Zhan⁶, Zishang Li⁷, Yi Ding⁸
School of Agricultural Engineering, Jiangsu University, Zhenjiang, China, 212013^{1, 2, 3, 4, 5, 6, 7, 8}
School of Mechanical Engineering, Jiangsu University, Zhenjiang, China, 212013³

Abstract—With the rapid developments of measurement and monitoring techniques, massive amounts of in-situ data have been recorded and collected from the measurement system of combine harvesters in their working process and/or field experiments. However, the relationship between the operation parameters and the performance index such as clearing loss usually changes greatly in different sample subspaces, which makes it difficult for conventional prediction models to model the in-situ data, since most of them assume that the relationship is the same or similar throughout the whole sample space. Therefore, a polynomial chaos expansion-based bagging prediction model (PCE-BP) is proposed in this article. A polynomial chaos expansion-based decision tree is constructed to divide the sample space such that the relationship between the operation parameters and the performance index in the same part is more similar than the others, and bagging is used to ensemble the polynomial chaos expansion-based decision trees to reduce the perturbation and provide robust predictions. The experiments on the mathematical functions show that the proposed prediction model outperforms polynomial chaos expansion, polynomial chaos expansion-based decision tree, and the conventional bagging prediction model. The proposed prediction model is validated through two monitoring datasets from a combine harvester. The experimental results show that the PCE-BP model provides better cleaning loss and impurity rate prediction results than the other prediction models in most experiments, showing the advantages of sample space partitioning and bagging in the data modeling of combine harvesters.

Keywords—Combine harvester; data modeling; polynomial chaos expansion; decision tree; bagging

I. INTRODUCTION

A combined harvester is a critical type of agricultural machinery that has been widely used to harvest grain crops. In the working units of a combine harvester, the grain crops are divided into grains and other materials by cutting, feeding, threshing, and cleaning. Since the interactions between the working units (header, conveying trough, cleaning fan, and sieve) and crops are very complex, it is difficult to construct an accurate theoretical or simulation models to accurately describe the working process of combine harvesters [1], [2], [3], [4]. In recent decades, numerical simulation methods such as computational fluid dynamics have been used to predict and analyze the working process of combine harvesters, which provides useful advice and references for the design, analysis, and optimization of working units [5], [6]. However, the computational cost of numerical simulations is too high to be

accepted. For instance, the computational fluid dynamics simulation of a cleaning fan takes approximately four hours. If 100 simulations are conducted to obtain the mapping function between the design variables and the flow rate, it would take out 400 hours, around 17 days. In the past few years, the intelligent techniques of combine harvester since more operation parameters can be monitored and measured in the operation process and/or field experiments of combine harvesters. The interaction mechanisms and information among the working units and those between the crops and working units are involved in the measured data. In addition, the computational cost of the data-driven model is usually much lower than that of the corresponding numerical simulation. Therefore, the data-driven design, analysis, optimization, and control of combine harvesters are being the topics of interest in recent years [7], [8], [9], [10].

In the data mining tasks of combine harvesters, the first and most crucial step is constructing a prediction model for the response of interest, such as cleaning loss and grain impurity. Compared with hyperparameter prediction models (such as artificial neural networks and support vector regression), polynomial regression-based prediction models offer the advantages of lower computational cost and higher interpretability, which have been widely used in the data modeling of combine harvesters. For example, Zareei and Abdollahpour [7] applied polynomial regression to identify the primary factors influencing header loss and determined the optimal factor combination through experimental design. Mirzazadeh et al. [11] constructed a semi-threshed cluster prediction model using polynomial regression and then optimized the feeding rate, fan speed, and sieve open rate to reduce the impurity rate. However, the interplay between operational factors and the associated performance indicators often changes greatly in different sample subspaces, as the interactions between the working units and crops are very complex, as discussed above. Most conventional polynomial regression-based prediction models assume that the regression relationship is the same or similar throughout the whole sample space, so it is challenging to assess the complex relationship between the operation parameters and the performance index of combine harvesters. On the other hand, the conventional polynomial regression method lacks nonlinear learning ability [12]. To this end, we proposed a prediction method in this work, aiming to solve the first problem by sample space partition based on the interplay between operational factors and the associated

*Corresponding Author.

performance indicators and to solve the second problem by introducing a Gaussian stochastic process to polynomial regression (polynomial chaos expansion, PCE).

Many sample space partition methods have been proposed in the area of machine learning, such as decision trees [13], the k -means algorithm [14], and the fuzzy c -means algorithm [15]. Since the k -means algorithm and fuzzy c -means algorithm cannot provide the partition rules directly, the sample space partition strategy proposed here is developed under the framework of decision tree. In a decision tree, the sample space is recursively partitioned so that the samples in the same subspace at each leaf node have similar/identical classification labels (classification decision tree) or responses (regression decision tree). Decision trees for classification purposes play a prominent role in various applications, including fault detection and identifying patterns. Chaitanya and Yadav [16] proposed a fault identification and location approach for multi-terminal lines based on a decision tree. The proposed method has been applied and validated based on series-compensated transmission lines and double-ended transmission lines. Liu et al. [17] designed a void detection method to assist in the health monitoring of sandwich-structured immersed tube tunnels, where the void classifier was constructed using a decision tree based on the characteristics of impact elastic waves. Muralidharan and Sugumarán [18] used continuous wavelet transform to represent the vibration signals of monoblock centrifugal pumps and applied a decision tree to predict different types of faults. In a regression decision tree, the sample space is recursively partitioned so that the continuous responses in the same subspace are similar. The average response of the samples at the same node is considered as the predicted value for new points. Liang et al. [19] used a regression decision tree to predict the uniaxial compressive strength based on the material parameters and indicated that the regression decision tree outperformed multiple regression in most experimental cases. Waruru et al. [20] analyzed the near infrared diffuse reflectance spectroscopy data of air-dried soil and then used a regression decision tree to estimate the soil aggregation level based on the spectral data. Nieto et al. [21] collected the filter pressure drop data of a micro irrigation system and constructed a pressure drop prediction model using a regression decision tree. In addition, the importance of the input variables is ranked based on the nodes and splitting values of the regression decision tree. In conventional decision trees, the choice of the splitting input variable and the determination of the partitioning threshold for each node hinge on either the classification labels or the mean response exhibited by the samples. Put simply, within each leaf node's subspace, samples share identical classification labels or comparable responses, yet there's no consistent correlation between input variables and the output. The prediction performance of the decision tree frequently undergoes significant variations due to the perturbation in the splitting feature optimization process as well. To solve the first problem, a new decision tree based on polynomial chaos expansion is proposed here, in which the sample space at each node is divided according to the regression relationship of samples. Then, bootstrap aggregation [22], [23], [24], also called bagging, is used to improve the robustness of the prediction results to solve the second problem.

Here's how the remainder of this document is structured. In Section II, the related works of sample space partitioning and polynomial chaos expansion are reviewed, and the motivation and framework of the proposed method are discussed as well. The details of the proposed polynomial chaos expansion-based bagging prediction model are presented in Section III. In Sections IV, several mathematical functions and two in-situ datasets of a combine harvester are used to validate the proposed prediction model. Section VI summarizes the conclusions and viewpoints.

II. POLYNOMIAL CHAOS EXPANSION-BASED BAGGING PREDICTION MODEL (PCE-BP)

A. Polynomial Chaos Expansion

In this study, Polynomial Chaos Expansion (PCE) is employed to assess the correlation between operational parameters and performance indices. In a PCE model, the response of interest y is estimated as follows [25].

$$y = \sum_{\alpha \in N^n} \beta_{\alpha} \Psi_{\alpha}(\mathbf{x}) \quad (1)$$

where \mathbf{x} is the vector of input variables, $\alpha = (\alpha_1, \dots, \alpha_n)$ is an n -dimensional index, β_{α} is the coefficients, and Ψ_{α} is the tensor product of normalized univariate orthogonal polynomials as follows.

$$\Psi_{\alpha}(x) = \prod_{i=1}^n \Psi_{\alpha_i}^i(x_i) \quad (2)$$

Usually, only the p -degree is considered in Eq. (1) to reduce the computation cost, and the response of interest in Eq. (1) is revised as follows.

$$y \cong \sum_{\alpha \in A^{p,n}} \beta_{\alpha} \Psi_{\alpha}(x) \quad (3)$$

$$A^{p,n} = \{\alpha \in N^n : \alpha = \sum_{i=1}^n \alpha_i \leq p\}$$

Eq. (1) can be rewritten as

$$y = \Psi \beta \quad (4)$$

where, y is the vector composed of the responses for the n samples, Ψ is the matrix of Hermite normalized univariate orthogonal polynomials, and β is the vector of the polynomial chaos coefficients. Upon examining the aforementioned equation, it becomes apparent that acquiring knowledge of the coefficients β allows for the derivation of the PCE model. Notably, when the quantity of samples is at least as many as the model's degree, the coefficients β can be estimated utilizing the least squares approach, as outlined below.

$$\beta = (\Psi^T \Psi)^{-1} \Psi^T y \quad (5)$$

B. Proposed Prediction Model

As discussed in Introduction, we proposed a new prediction model based on polynomial chaos expansion, named the polynomial chaos expansion-based bagging prediction model (PCE-BP), in which the sample space is partitioned to enhance prediction precision, while bagging techniques are employed to bolster the stability of the prediction outcomes. In the proposed prediction model, m PCE-based decision trees are generated. The main difference between the proposed PCE-based decision tree and the conventional decision tree is that the node is split according to on the regression relationship, but not the

classification labels or the mean response. At all nodes, the samples (D_{train}) are categorized into the left subset $D_{train-L}$ and the right subset $D_{train-R}$. Based on D_{train} , $D_{train-L}$, and $D_{train-R}$, three PCE models are constructed, named PCE_t , PCE_L , and PCE_R . The training error before and after partition is used to calculate the splitting criterion S .

$$S = R_{after}^2 - R_{before}^2 + \theta$$

$$R_{before}^2 = 1 - \left(\frac{\sum_{i=1}^{n_t} (y_{i,t} - \bar{y}_{i,t})^2}{\sum_{i=1}^{n_t} (y_{i,t} - \bar{y})^2} \right) \quad (6)$$

$$R_{before}^2 = 1 - \left(\frac{\sum_{i=1}^{n_L} (y_{i,L} - \bar{y}_{i,L})^2 + \sum_{i=1}^{n_R} (y_{i,R} - \bar{y}_{i,R})^2}{\sum_{i=1}^{n_t} (y_{i,t} - \bar{y})^2} \right)$$

where $y_{i,t}$, $y_{i,L}$, and $y_{i,R}$ are the real responses of D_{train} , $D_{train-L}$, and $D_{train-R}$, respectively; \bar{y} is the mean response of D_{train} ; $\bar{y}_{i,L}$, $\bar{y}_{i,R}$, and $\bar{y}_{i,t}$ are the PCE predicted responses; n_t , n_L , and n_R are the sample sizes of D_{train} , $D_{train-L}$, and $D_{train-R}$, respectively; and θ is the adjustment coefficient. When $S > 0$, the current node is a leaf node. The construction process of the polynomial chaos expansion-based decision tree is summarized in Fig. 1. Utilizing the classification rules derived from the PCE-based decision tree, samples for prediction undergo classification until they arrive at leaf nodes, at which point the PCE models positioned as those leaf nodes provide the predictive responses.

A popular heuristic algorithm, the gray wolf optimizer [26], is modified to optimize the splitting feature at each node. In the modified optimization algorithm, the split input variable is first

transformed into a latent variable. Given d input variables, the latent variable represents each alternative splitting input variable through the following Table I.

After that, the new quantitative variable (q) and the division point (p_d) are combined into a vector $z = [q, p_d]$ and optimized as follows. In the optimization process, the solution with the best splitting criterion S is set as the alpha (z_α), the beta (z_β) and the delta wolf (z_δ) are worse than z_α , and the other wolves are omega wolves (z_ω). During each iteration, the solution will undergo an update as detailed below.

$$z(t + 1) = z(t) - a \cdot d \quad (7)$$

where $z(t + 1)$ is the updated solution, $z(t)$ is the current solution, a is a coefficient vector, and d is the motion of the wolf relative to the prey (z_{prey}), which is defined as follows:

$$d = |c \cdot z_{prey}(t) - z(t)| \quad (8)$$

where:

$$a = 2a \cdot r_1 - \tau \quad (9)$$

$$c = 2 \cdot r_2 \quad (10)$$

$$\tau = 2 - t \left(\frac{2}{T} \right) \quad (11)$$

TABLE I. THE INPUT VARIABLES AND LATENT VARIABLES

Input variable	1	2	...	$d - 1$	d
Quantitative variable	$[0, \frac{1}{d}]$	$[\frac{1}{d}, \frac{2}{d}]$...	$[\frac{d-2}{d}, \frac{d-1}{d}]$	$[\frac{d-1}{d}, 1]$

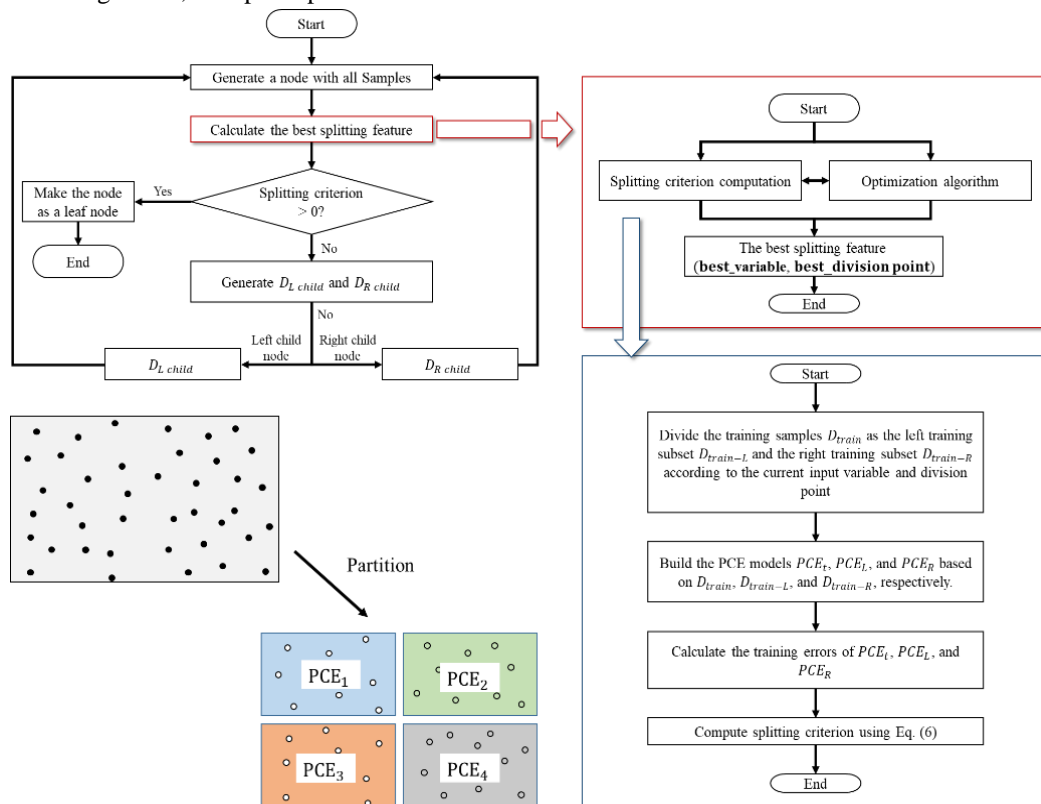


Fig. 1. Polynomial chaos expansion-based decision tree.

Where, T is the maximum number of iterations, and t is the current iteration. The positions of the other wolves (ω) are adjusted based on the three best solutions (z_α , z_β , and z_δ) as follows:

$$\begin{cases} d_\alpha = |c_1 \cdot z_\alpha(t) - z| \\ d_\beta = |c_2 \cdot z_\beta(t) - z| \\ d_\delta = |c_3 \cdot z_\delta(t) - z| \end{cases} \quad (12)$$

$$\begin{cases} z_1 = z_\alpha - a_1 \cdot (d_\alpha) \\ z_2 = z_\beta - a_2 \cdot (d_\beta) \\ z_3 = z_\delta - a_3 \cdot (d_\delta) \end{cases} \quad (13)$$

$$z(t + 1) = \frac{z_1(t+1) + z_2(t+1) + z_3(t+1)}{3} \quad (14)$$

The above process of Eq. (7)-Eq. (14) repeats until the termination criterion is fulfilled, and the z_α of the last iteration is considered the best splitting feature.

From the above equations, it can be found that the random generation of initial solutions have effect on the construction process of the PCE-based decision tree. In other words, the generated decision tree might be slightly different even if the settings are same. Thus, Bagging is introduced to solve this problem, in which where m PCE-based decision trees are generated simultaneously and the final prediction result is estimated by the generated PCE-based decision trees.

$$\widehat{y}^* = \frac{\sum_{i=1}^m \widehat{y}_i^*}{m} \quad (15)$$

Where \widehat{y}^* represents the final estimated response, and \widehat{y}_i^* refers to the response produced by the i -th decision tree based on PCE.

III. EXPERIMENTS ON MATHEMATICAL FUNCTIONS

The effects of the parameter settings including the number of trees (m) and the adjustment coefficient (θ) on the proposed PCE-BP model are studied through two mathematical functions. The proposed model (PCE-BP) is compared with PCE, PCE-based decision tree (PCET), and random forest. The effectiveness of the aforementioned methods is assessed using R -square (R^2), calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \widehat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (16)$$

where y_i is the i -th real response, \widehat{y}_i is the corresponding predicted value, n is the number of testing points, and \bar{y} is the mean of the real responses. The closer R^2 is to 1, the better the performance of the prediction model.

A. Experiments on a Single Two-Dimensional Function

A two-dimensional function is utilized to validate the proposed PCE-BP model, maintaining consistency across the entire space. The function is defined as follows:

$$y = x_1^2 - 5\cos(2\pi x_2) \quad x \in [-1, 1] \quad (17)$$

The impact of the number of trees is examined first. For each number of trees (5, 10, 15, ..., 45, and 50), 20 experiments are conducted, where the parameter θ is set as 0.05. 100 samples are generated using Latin hypercube sampling, and another 2,000

samples are used to validate the prediction models. The obtained mean and variance of R^2 are shown in Fig. 2.

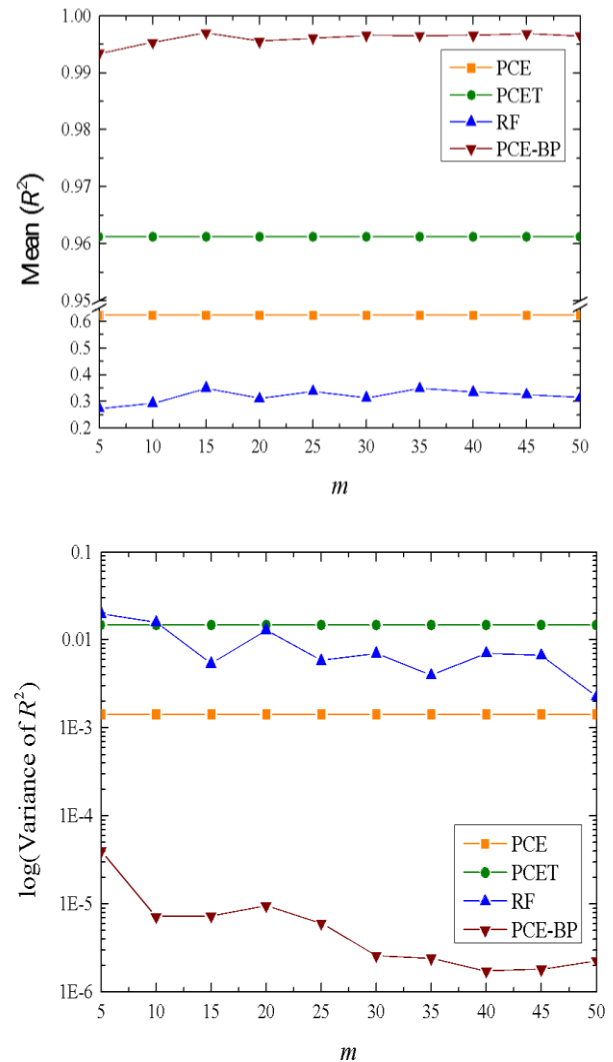


Fig. 2. The prediction results with different m .

From Fig. 2, it is found that the PCE-BP method outperforms the other prediction models in terms of the mean R^2 for all the values of m . The mean R^2 of PCET is higher than that of PCE, which is mainly because the PCET model partitions the sample space into several subspaces so that the regression relationship between the input variables and the response of interest is similar. Compared with the PCET model, the PCE-BP model produces better results with R^2 higher than 0.99. In each PCE-based decision tree, the sample space splitting at each node is influenced by the training samples and the randomly generated initial potential solutions for the splitting feature. The PCE-BP model uses bagging strategy to solve this issue, in which the predicted response is averaged by several PCE-based decision trees. The performance of the PCE-based decision tree varies greatly (the highest variance of R^2 for most experiments is shown in Fig. 2.), so its mean R^2 is smaller than that of the PCE-BP model. The PCE-BP model classifies the samples based on the regression relationship, but the RF model is based on the

mean responses. Thus, the PCE-BP model is also better than RF. The PCE-BP model is much better than the other three models in term of the variance of R^2 as well. The highest value is smaller than 0.0001, showing its robust prediction performance. With the parameter m increasing, the mean R^2 of the PCE-BP model first increases and then changes slightly when m exceeds 10. The variance of R^2 first decreases as parameter m increases and tends to remain stable. A larger m means that more PCE-based decision trees are generated in the PCE-BP model so that the perturbation brought by the splitting feature optimization is eliminated. As a result, the prediction accuracy is increased, as shown in Fig. 3. From the results and analysis above, it can be determined that the adverse effect of the splitting feature optimization is effectively reduced when m exceeds 25. In other words, the proposed PCE-BP model provides competitive prediction results for the single two-dimensional mathematical function tested here when the number of trees exceeds 25.

From Eq. (6), it can be found that the parameter θ is directly correlated with the splitting criterion, which would have an important effect on the prediction results of the PCE-BP model. The number of trees is set as 30, and the parameter θ is set as 0.04, 0.042, ..., 0.058, and 0.06. The mean and variance of R^2 over 20 experiments for each θ are presented in Fig. 3. The PCE-BP model is still better than that of the PCE, PCET, and RF models, showing the advantages of sample space partitioning and bagging. With the parameter θ increasing from 0.04 to 0.058, the mean R^2 increases and then tends to be stable. When the parameter θ is higher than 0.058, the mean R^2 decreases slightly. From Section III, it is known that the larger θ is, the higher the regression error before and after partitioning at each node. Samples within the same subspace show a more similar regression relationship between the input variables and the response of interest than those in different subspaces. The prediction accuracy of the PCE-BP model increases as parameter θ increases. However, when the parameter θ is too high, it is sample space is hard to be divided by the PCE-BP model, so the performance of the PCE-BP model decreases. The variance of R^2 decreases with increasing θ , indicating that the performance of the PCE-BP model is more robust. As a

conclusion, the PCE-BP model produces competitive prediction performance when the parameter θ is approximately 0.05.

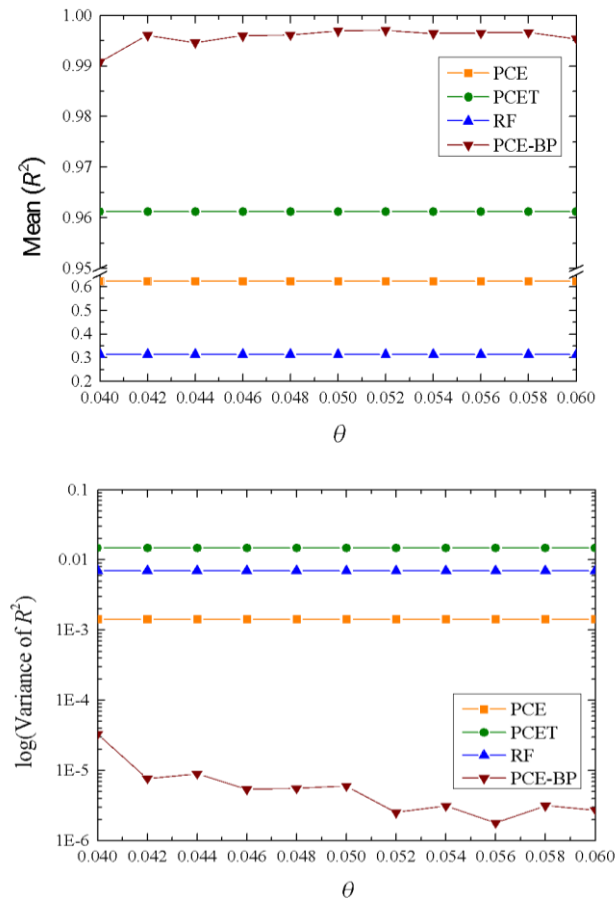


Fig. 3. The prediction results with different θ .

B. Piecewise Four-Dimensional Function

A piecewise four-dimensional function is used to validate the PCE-BP model, in which the mathematical function changes in different subspace, as defined as follows:

$$\begin{cases} y = \sin \sin (2\pi x_1) + x_2^2 + x_3 + x_4, x_1 \in [0, 0.5], x_2, x_3, x_4 \in [0, 1] \\ y = \cos \cos (2\pi x_1) + x_2 + x_3 + x_4, x_1 \in [0.5, 1], x_2, x_3, x_4 \in [0, 1] \end{cases} \quad (18)$$

The parameter θ is 0.05, and m is set as 5, 10, ..., 45, and 50. Two hundred points are generated for the prediction models, then then the model accuracy is validated through another 4,000 samples. The obtained mean and variance of R^2 for 20 experiments are shown in Fig. 4. The mean R^2 of PCET PCE-BP is higher than that of PCE in all experiments. With the help of sample space partition according the relationship between the input variables and the response of interest, the PCE models at the leaf nodes of the PCET and PCE-BP can accurately evaluate the relationship, and the overall prediction accuracy is improved as well. The PCE-BP model outperforms the RF model as well, indicating that the prediction model of each subspace of the PCE-BP model can accurately evaluate the relationship. In addition, the prediction accuracy of the PCET and PCE-BP models both surpass the RF model. The variance of R^2 of the

PCE-BP and RF models is smaller than those of the PCE and PCET models, which is mainly because of the introduction of bagging. With the parameter m increases, the average performance of the PCE-BP model increases and then tends to be stable. When m increases from 35 to 50, the variance of R^2 changes slightly. A larger m means that more PCET trees are generated in the PCE-BP model, so the perturbation brought by the splitting feature optimization can be more effectively eliminated. Thus, as m increases, the prediction accuracy increases, and the performance variance decreases, as shown in Fig. 4. When the parameter m is too high, the perturbation brought by the splitting feature optimization cannot be further reduced, so the mean and variance of the prediction results tend to be stable. Overall, the proposed PCE-BP model outperforms

the other three models and produces competitive prediction performance when the number of trees surpasses 25.

In the following experiments, the parameter m is set as 25; θ is set as 0.04, 0.042, ..., 0.058, and 0.060. The experiments are conducted 20 times for each value of θ , and the mean and variance of R^2 are shown in Fig. 5. From this figure, it is found that the proposed model outperforms the other models. With the parameter θ increasing from 0.040 to 0.046, the mean of R^2 of the proposed model increases, but the variance of R^2 decreases. As θ continuously increases to 0.06, both the mean and variance of R^2 change slightly. From the introduction of the PCE-BP model, it can be found that the initial space at the node is more likely to be partitioned when parameter θ is relatively smaller. In other words, the PCE-based decision tree is deeper, which results in the tendency of decision tree overfitting [13]. Thus, the prediction performance of the proposed model tends to be better when θ is larger. On the other hand, the variation brought by the splitting feature optimization is also reduced since the space at the node is more difficult to partition. Therefore, the variance of R^2 decreases as θ increases. The PCE-BP model provides competitive results for the piecewise four-dimensional mathematical function when θ is approximately 0.05.

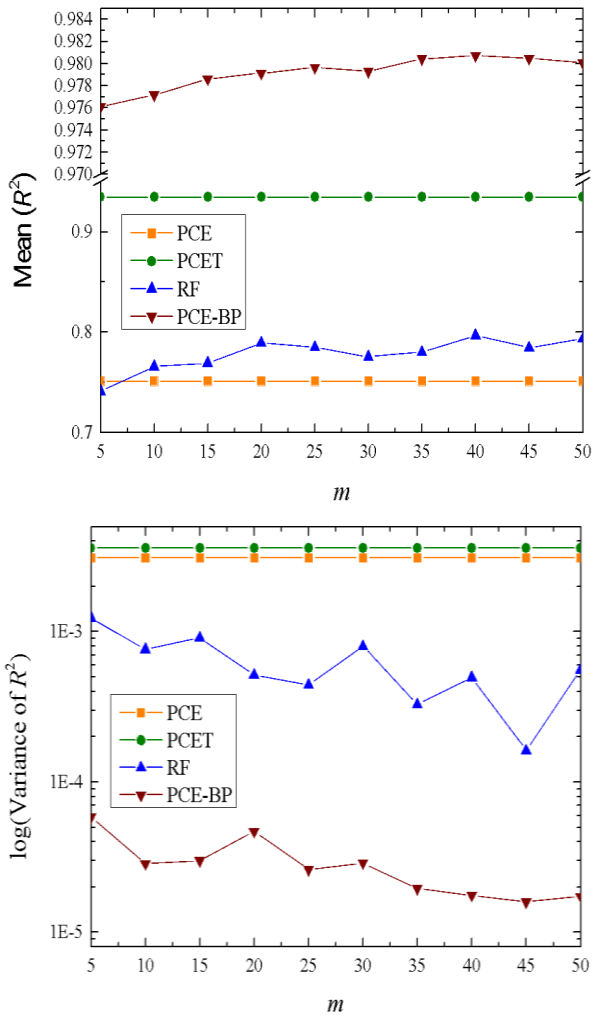


Fig. 4. The prediction results with different m .

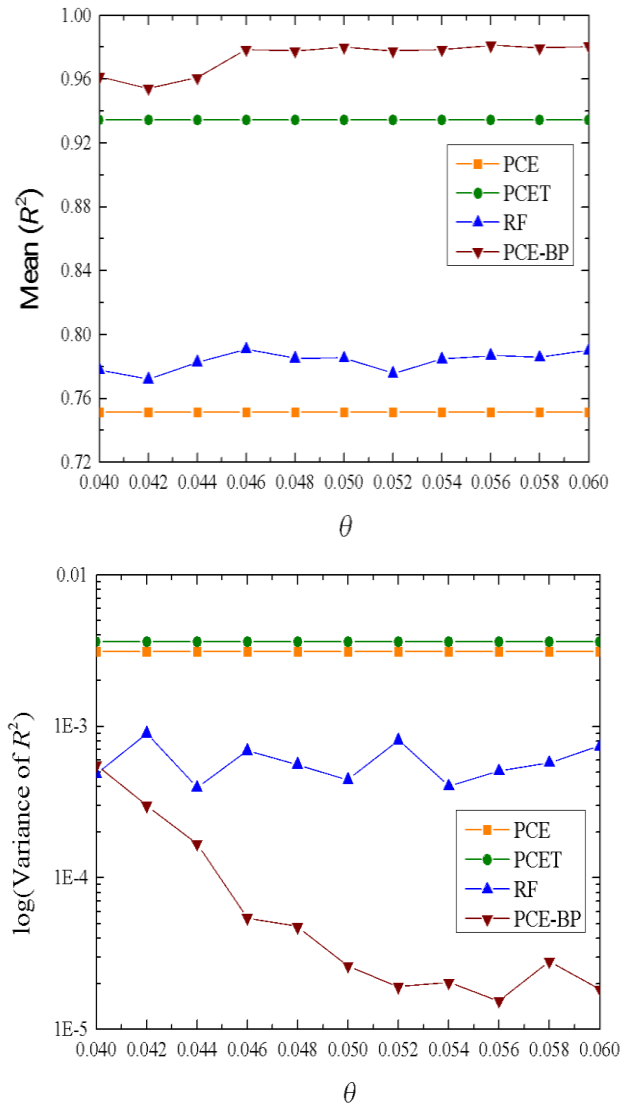


Fig. 5. The prediction results with different θ .

IV. VALIDATION BASED ON THE DATASETS OF A COMBINE HARVESTER

A. Case 1 (Cleaning Loss)

The PCE-BP model is applied to a while-feed combine harvester manufactured in Jiangsu, China (World Ruilong 4LZ-6.0A). The dataset used here comes from the field experiment (Fig. 6), containing 750 samples, including the header height, the open rate of the cleaning fan, the rotation speed of the cleaning fan, the open rate of the sieve, the angle of the guide plate, the rotation speed of the threshing drum, the gap of the threshing drum, and the cleaning loss. In each experiment, ten folds cross-validation experiments are conducted (in each experiment, a segment of the data serves as the basis for verifying the accuracy of the cleaning loss prediction models, whereas the remaining nine segments are dedicated to building the prediction model). The parameters m and θ of the PCE-BP model are set as 30 and 0.05, respectively. Fig. 7 shows the results of ten experiments.



Fig. 6. Field experiment of harvesting rapeseed.

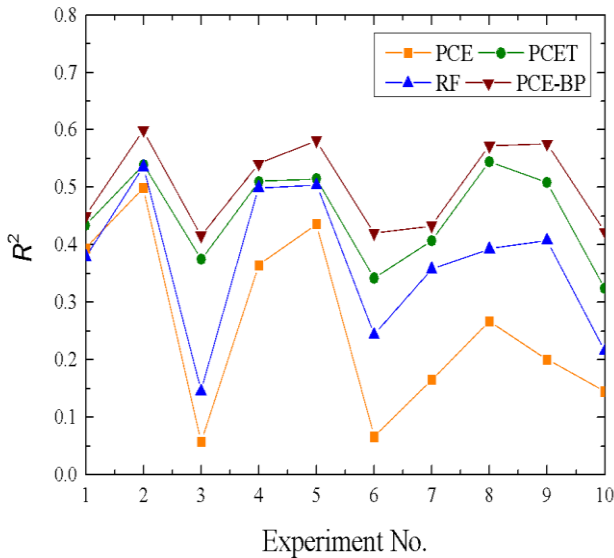


Fig. 7. Experimental results of Case 1.

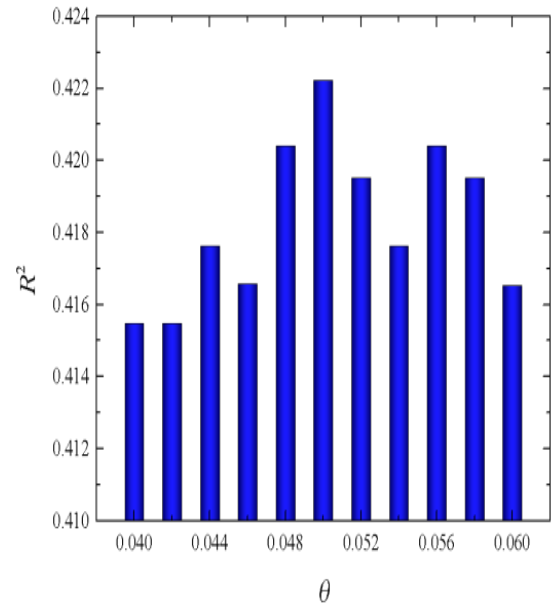
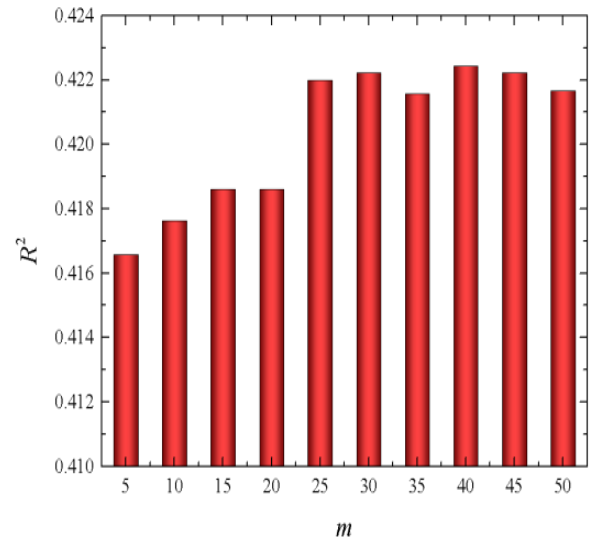


Fig. 8. Effect of parameters m and θ on the PCE-BP model.

It is found that the PCE-BP model is better than the other three models. The R^2 of the PCET model is higher than that of the conventional PCE model, where the PCET's average R^2 is 0.449 and PCE is 0.259. In the PCET model, the sample space is divided into different parts such that the relationship between the operating parameters and the cleaning loss in the same part is more similar than those in the other parts, thus improving the cleaning loss prediction accuracy. Similarly, RF is also better than the PCE model, which can be attributed to the sample space partitioning. With the help of the bagging strategy, the perturbation brought by the feature splitting in PCET is eliminated. Additionally, the RF model divides the sample space according to the mean responses. Thus, the PCE-BP model outperforms the PCET and RF models in all experiments.

The training and testing datasets in the tenth experiment are used to study the number of trees m on the proposed model. The parameter θ is set as 0.05, and the parameter m is set as 5, 10,

..., 45, and 50. Fig. 8 shows the experimental results. It is found that the prediction performance of PCE-BP increases as m increases from 5 to 25. A larger m means that more PCE-based decision trees are generated in the PCE-BP model, which means that the perturbation brought by the splitting feature optimization can be effectively reduced, thus increasing the prediction accuracy. As m continually increases to 50, the R^2 of the PCE-BP model changes slightly, which is mainly because the perturbation cannot be reduced further. The effect of the parameter θ is studied as well, and the results are shown in Fig. 8, where the number of tree is 30. It is observed that R^2 tends to increase and then decrease with increasing parameter θ . From the introduction of the PCE-BP model, it can be found that a deeper PCE-based decision tree would be constructed when the parameter θ is relatively smaller. The generated tree is easy to overfit, so R^2 tends to be lower. When the parameter θ is relatively larger, the space at the node is more difficult to partition, so R^2 is lower. Overall, the PCE-BP model can provide competitive performance for different values of m and θ .

B. Case 2 (Impurity Rate)

Another dataset is used here, which includes 337 samples with recorded operational parameters and impurity rates. Ten folds cross-validation experiments are conducted as well in this subsection. The parameters m and θ are set to 30 and 0.05, respectively. Fig. 9 shows the experiments. It is found that the PCE-BP model outperforms the PCE and RF models, which can be attributed to sample space partitioning based on the relationship between the input variables and the response of interest. The R^2 of the PCE-BP is higher than PCET in most experiments. In experiments 2, 3, and 9, the performance of the PCE-BP model is very close to that of the PCET model. From Fig. 10, it can be found that the mean R^2 of PCE-BP is 0.550, which exceeds those of the other three models, highlighting the benefits of sample space partitioning and bagging.

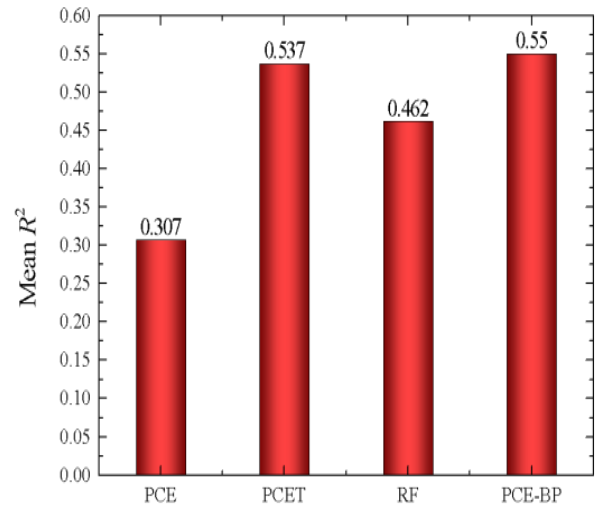


Fig. 9. Experimental results of Case 2.

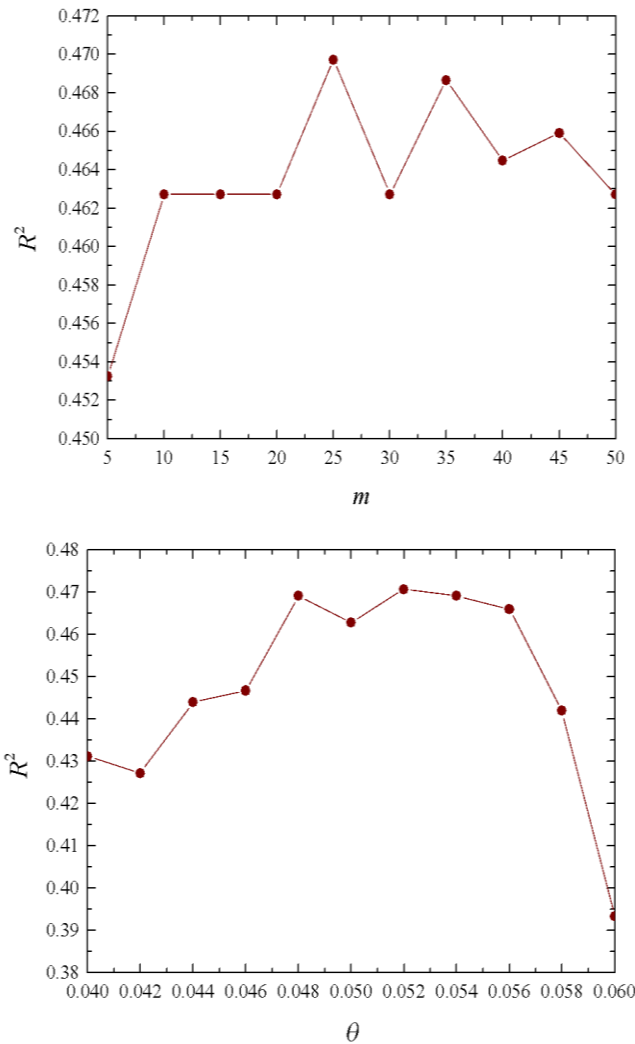
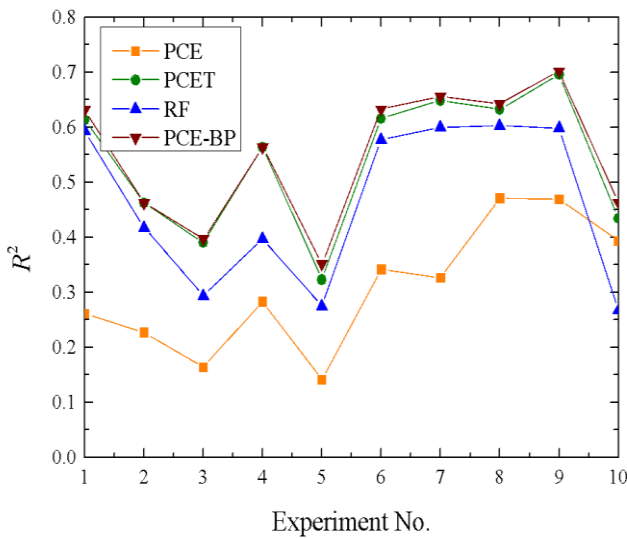


Fig. 10. Effects of parameters m and θ on the PCE-BP model.



The training and testing datasets from the tenth experiment are used to study the effects of the parameters m and θ on the PCE-BP model. The obtained results are shown in Fig. 10 (θ is set as 0.05). It is found that the R^2 with $m = 25\sim 50$ is higher than $m = 5\sim 20$. The parameter m is set as 30, the parameter θ is set as 0.04, 0.042, ..., 0.058, and 0.06, and the results are shown in Fig. 10 as well. R^2 first increases and then decreases as θ increases. When the factor θ is relatively small, a deeper PCE-based decision tree is constructed, which means that the prediction model easily overfits. On the other hand, when the parameter θ is relatively larger, the space at the node is more difficult to partition, so R^2 is lower as well. Overall, the PCE-BP model provides better results than the other three models for most values of parameters m and θ , as shown in Fig. 9 and Fig. 10.

V. CONCLUSION

In this paper, a polynomial chaos expansion-based bagging prediction model (PCE-BP) is proposed for modeling the field data of a combine harvester. In the proposed model, a polynomial chaos expansion-based decision tree is designed to partition the sample space, and bagging is used to ensemble the polynomial chaos expansion-based decision trees. The efficiency of the proposed prediction model is first validated through single and piecewise mathematical functions. The results show that the proposed prediction model outperforms polynomial chaos expansion, polynomial chaos expansion-based decision tree, and the conventional bagging prediction model functions. The proposed model demonstrates excellent prediction performance with 25 trees and the adjustment coefficient of 0.05. The proposed prediction model is further validated through two in-situ datasets of a combine harvester. The PCE-BP model provides more accurate cleaning loss and impurity rate prediction results than the conventional prediction models in most experiments. The experimental results show the advantages of sample space partitioning and bagging in the data modeling of combine harvesters.

From the results of the experiments, we found that the construction time of the proposed model is higher than that based on the conventional polynomial chaos expansion. In future work, the cost reduction of the proposed model will be our research topic. In addition, the other ensemble strategy such as boosting would be introduced into the proposed model as well.

ACKNOWLEDGMENT

This work is supported by the Project funded by China Postdoctoral Science Foundation (Grant No. 2022M711388); the Natural Science Foundation of Jiangsu Province (Grant No. BK20210777); Jiangsu Province and Education Ministry Co-Sponsored Synergistic Innovation Center of Modern Agricultural Equipment (Grant No. XTCX2014); and the Funding of Jiangsu University (Grant No. 20JDG068).

REFERENCES

- [1] I. Badretidinov, S. Mudarisov, R. Lukmanov, V. Permyakov, R. Ibragimov and R. NasYROV. "Mathematical modeling and research of the work of the grain combine harvester cleaning system," *Computers and Electronics in Agriculture*, 2019, vol. 165, p. 104966.
- [2] Z. Liang, Y. Li, J. D. Baerdemaeker, L. Xu and W. Saeys. "Development and testing of a multi-duct cleaning device for tangential-longitudinal flow rice combine harvesters," *Biosystems Engineering*, 2019, vol. 182, p. 95-106.
- [3] J. Pang, Y. Li, J. Ji, & L. Xu. "Vibration excitation identification and control of the cutter of a combine harvester using triaxial accelerometers and partial coherence sorting," *Biosystems Engineering*, 2019, vol. 185, p. 25-34.
- [4] Z. Qiu, G. Shi, B. Zhao, X. Jin, and L. Zhou. "Combine harvester remote monitoring system based on multi-source information fusion," *Computers and Electronics in Agriculture*, 2022, vol. 194, p. 106771.
- [5] C. Fan, T. Cui, D. Zhang and Qu, Z. "Design of Feeding Head Spiral Angle Longitudinal Axis Corn Threshing Separation Device Based on EDEM," 2019 Boston, Massachusetts July 7- July 10, 2019: n. pag.
- [6] Tang, H., Xu, C., Zhao, J., and Wang, J. "Formation and steady state characteristics of flow field effect in the header of a stripping prior to cutting combine harvester with CFD" *Computers and Electronics in Agriculture*, 2023, vol. 211, p. 107959.
- [7] S. Zareei and S. Abdollahpour. "Modeling the optimal factors affecting combine harvester header losses," *Agricultural Engineering International: CIGR Journal*, 2016, vol. 18(2), p. 60-65.
- [8] Z. Guan, Y. Li, S. Mu, M. Zhang, T. Jiang, H. Li, G. Wang and C. Wu. "Tracing algorithm and control strategy for crawler rice combine harvester auxiliary navigation system," *Biosystems Engineering*, 2021, vol. 211, p. 50-62.
- [9] L. Nádaí, F. Imre, S. Ardabili, T. M. Gundoshmian, P. Gergo and A. Mosavi. "Performance analysis of combine harvester using hybrid model of artificial neural networks particle swarm optimization," In 2020 RIVF International Conference on Computing and Communication Technologies (RIVF) (pp. 1-6). IEEE.
- [10] Chen, M., Jin, C., Ni, Y., Yang, T., and Zhang, G "Online field performance evaluation system of a grain combine harvester," *Computers and Electronics in Agriculture*, 2022, vol. 198, p. 107047.
- [11] A. Mirzazadeh, S. Abdollahpour and M. Hakimzadeh. "Optimized Mathematical Model of a Grain Cleaning System Functioning in a Combine Harvester using Response Surface Methodology," *Acta Technologica Agriculturae*, 2022, vol. 25(1), p. 20-26.
- [12] Torre, E., Marelli, S., Embrechts, P., and Sudret, B. "Data-driven polynomial chaos expansion for machine learning regression," *Journal of Computational Physics*, 2019, vol. 388, p. 601-623.
- [13] Costa, V. G., and Pedreira, C. E. "Recent advances in decision trees: An updated survey," *Artificial Intelligence Review*, 2023, vol. 56(5), 4765-4800.
- [14] Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., and Heming, J. "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Information Sciences*, 2023, vol. 622, p. 178-210.
- [15] M. Shi, Z. Liang, J. Zhang, L. Xu and Song, X. "A robust prediction method based on Kriging method and fuzzy c-means algorithm with application to a combine harvester," *Structural and Multidisciplinary Optimization*, 2022, vol. 65(9), p. 1-18.
- [16] B. K. Chaitanya, & A. Yadav. "Decision tree aided travelling wave based fault section identification and location scheme for multi-terminal transmission lines," *Measurement*, 2019, vol. 135, p.312-322.
- [17] R. Liu, S. Li, G. Zhang and W. Jin. "Depth detection of void defect in sandwich-structured immersed tunnel using elastic wave and decision tree," *Construction and Building Materials*, 2021, vol. 305, p. 124756.
- [18] V. Muralidharan and V. Sugumaran. "Feature extraction using wavelets and classification through decision tree algorithm for fault diagnosis of mono-block centrifugal pump," *Measurement*, 2013, vol.46(1), p. 353-359.
- [19] M. Liang, E. T. Mohamad, R. S. Faradonbeh, D. J. Armaghani and S. Ghoraba. "Rock strength assessment based on regression tree technique," *Engineering with Computers*, 2016, vol. 32(2), p. 343-354.
- [20] B. K. Waruru, K. D. Shepherd, G. M. Ndegwa and A. M. Sila. "Estimation of wet aggregation indices using soil properties and diffuse reflectance near infrared spectroscopy: An application of classification and regression tree analysis," *Biosystems Engineering*, 2016, vol. 152, p. 148-164.
- [21] P. J. G. Nieto, E. García-Gonzalo, G. Arbat, M. Duran-Ros, F. R. Cartagena and J. Puig-Bargues. "Pressure drop modelling in sand filters

- in micro-irrigation using gradient boosted regression trees,” *Biosystems engineering*, 2018, vol. 171, p. 41-51.
- [22] R. E. Banfield, L. O. Hall, K. W. Bowyer and W. P. Kegelmeyer. “A comparison of decision tree ensemble creation techniques,” *IEEE transactions on pattern analysis and machine intelligence*, 2006, vol. 29(1), p. 173-180.
- [23] P. Yariyan, S. Janizadeh, T. Van Phong, H. D. Nguyen, R. Costache, H. Van Le and J. P. Tiefenbacher. “Improvement of best first decision trees using bagging and dagging ensembles for flood probability mapping,” *Water Resources Management*, 2020, vol. 34(9), p. 3037-3053.
- [24] S. Moral-García, C. J. Mantas, J. G. Castellano, M. D. Benítez and J. Abellan. “Bagging of credal decision trees for imprecise classification,” *Expert Systems with Applications*, 2020, vol. 141, p. 112944.
- [25] J. Zhang, X. Yue, J. Qiu, L. Zhuo and J. Zhu. “Sparse polynomial chaos expansion based on Bregman-iterative greedy coordinate descent for global sensitivity analysis,” *Mechanical Systems and Signal Processing*, 2021, vol. 157, p. 107727.
- [26] S. Mirjalili, S.M. Mirjalili and A. Lewis. “Grey wolf optimizer,” *Advances in engineering software*, 2014, vol. 69, p. 46-61.

Detecting Emotions with Deep Learning Models: Strategies to Optimize the Work Environment and Organizational Productivity

Cantuarias Valdivia Luis Alberto de Jesús, Gómez Human Javier Junior, Sierra-Liñan Fernando
Facultad de Ingeniería, Universidad Privada del Norte, Lima, Perú

Abstract—This study proposes the implementation of a facial emotion recognition system based on Convolutional Neural Networks to detect emotions in real time, aiming to optimize the workplace environment and enhance organizational productivity. Six deep learning models were evaluated: Standard CNN, AlexNet, VGG16, InceptionV3, ResNet152 and DenseNet201, with DenseNet201 achieving the best performance, delivering an accuracy of 87.7% and recall of 96.3%. The system demonstrated significant improvements in key performance indicators (KPIs), including a 72.59% reduction in data collection time, a 63.4% reduction in diagnosis time, and a 66.59% increase in job satisfaction. These findings highlight the potential of Deep Learning technologies for workplace emotional management, enabling timely interventions and fostering a healthier, more efficient organizational environment.

Keywords—Facial recognition; real-time emotions; convolutional neural networks; work environment; artificial intelligence in human resources

I. INTRODUCTION

Many companies today face the challenge of effectively managing their employees' emotions to improve the work environment. As noted in study [1], mental health is a state of well-being that allows you to manage stress and work effectively. In addition, a high percentage of workers suffer from work-related stress, which affects their mental health, performance and interpersonal relationships, generating significant economic and social costs [2]. This impact is linked to the “emotional sphere”, a model in which emotions influence group processes and outcomes [3].

In this way, it is [4] highlighted that positive socio-emotional interactions foster collaboration and teamwork. For its part, [5] it underlines that emotional management is a crucial skill for work success, since it allows regulating one's own and others' emotions. In Peru, job dissatisfaction is a growing problem, with 78% of Peruvians reporting burnout in 2023 [6]. This problem affects both the quality of life of employees and talent retention, since workers with high levels of stress are 4.5 times more likely to quit [7].

Traditional emotional management methodologies are limited, technologies such as facial recognition integrated with AI emerge as effective solutions to analyze emotions in real time and optimize organizational decisions [8]. Facial recognition has been integrated into multiple industries for its ability to authenticate identities and analyze emotions [9]. However,

identifying similar expressions such as fear, and surprise remains a challenge [10]. Since 55% of communicative information comes from non-verbal elements [11], these tools are essential to address emotional problems in organizations.

In this context, there is a need to implement an intelligent system for the detection of emotions in real time in controlled work environments. This approach seeks to optimize the identification of emotional states with high precision, contributing to the improvement of the work environment and the productivity of organizations.

The paper is organized as follows. Section II includes the literature review, where previous studies and key concepts are presented. Section III describes the methodology used. Section IV presents the results obtained from the experiment performed. Section V addresses the discussion of the findings. Finally, Section VI presents the conclusions and possible future work.

II. LITERATURE REVIEW

Facial recognition, based on computer vision and deep learning techniques, allows for the identification and analysis of emotional expressions with high precision. Artificial Intelligence (AI) algorithms are trained with large data sets to learn distinctive patterns, applicable in areas such as security, education, medicine and marketing [12], [13], [14]. In addition, its non-intrusive capacity and operational autonomy make it versatile technology for machine learning tasks [13].

Although effective, emotion recognition faces significant challenges, cultural differences and expression ambiguity are obstacles highlighting the need for high-quality data [15]. In [16], their DenseNet201 model showed the highest accuracy, with 86.85%, in detecting fake faces, outperforming other convolutional neural network architectures by using advanced transfer learning techniques and medicine benefits from AI to understand emotions [17], although they require advanced methods to address the complexity of facial expressions. Additionally, a BLTSM-based model based on attention mechanisms was shown to be effective in describing emotional attitudes and recognizing emotions [18].

Recent advances have greatly improved accuracy, for example, the “IPSOBSA-QCNN” was developed, a quantum neural network that achieves 98% success in emotion classification [19]. On the other hand, local binary patterns and extreme learning were integrated to maximize effectiveness on data sets such as CK+ and JFFE [20]. In addition, histograms of

oriented gradients (HOG) and fast networks were used [21], achieving 95.04% accuracy.

Specialized applications reinforce the potential of facial recognition, where the “EigenFaces” algorithm was used with libraries such as OpenCV and SKlearn to analyze emotions [22], emotional changes were detected from low-resolution images [23]. In addition, the effectiveness of convolutional neural networks to identify emotions in real time with an accuracy greater than 80% was highlighted [24], [25]; on the other hand, it was shown how scalable models allow the classification of seven emotions in video game, security and education applications [26].

Automatic recognition of facial emotions has seen significant advances thanks to convolutional neural networks (CNNs), which allow for highly accurate classifications. However, the effectiveness of these models largely depends on the quality and diversity of the data sets used for their training, highlighting the need for careful and representative data collection [27]. Recent advances in deep neural networks have

achieved accuracy rates above 90% in emotion classification, while future research aims to develop more robust models that adapt to diverse contexts and environmental conditions [28]. On the other hand, emotion recognition from visual data faces significant challenges due to the subjective nature of human emotions and the complexity of visual information, which has led to the use of convolutional neural networks to improve sentiment classification accuracy [29]; In addition, they have benefited from genetic algorithms (GA) [30] to optimize the hyperparameters of CNN models, achieving an accuracy of 76.11% in the third generation, consolidating the potential of the CNN-GA approach in facial emotion recognition [30].

In summary, these studies demonstrate that AI-based emotion detection has great potential across multiple sectors, although challenges remain related to cultural interpretation, data quality, and ethical applications. Table I compares recent studies on emotion detection using neural networks, highlighting their objectives, methods, datasets, and key results, along with the approach proposed in this work.

TABLE I. WORK RELATED TO EMOTION DETECTION USING DEEP LEARNING

Study	Goals	Method	Evidence	Results
[12]	Emotion detection in low-resolution images using residual networks	Residual network with voting	RAF-DB dataset (low resolution images)	Accuracy: 85.69%
[38]	Feature Extraction in Detection with EfficientNetB0	EfficientNetB0	FER2013 dataset	Accuracy: 95.82%
[31]	Real-time emotion detection for human-robot collaboration in smart factories	DeepFace	Industrial contexts (real-time testing)	High precision in real-time collaboration scenarios
[26]	Emotion recognition in low-resolution images	CNN	FER2013 dataset (low resolution images)	Accuracy: 66.85%
[19]	Quantum CNN for emotion detection in mental health contexts	Quantum CNN (QCNN)	Mental health context	Improved efficiency and reduced training times
Proposed	Real-time emotion detection in controlled work environments	Standard CNN, AlexNet, VGG16, InceptionV3, ResNet152, DenseNet201	Controlled work environment (real-time testing)	High precision in real-time emotion detection

III. PROPOSED METHODOLOGY

In this research, the CRISP-DM model was taken as inspiration as a basis to address the different stages of the project, from understanding to deployment, widely used as a

standard in data mining [32]. This framework, widely accepted in data analysis [33], stands out for its adaptability, allowing its application in areas such as medicine [34], signal processing [35], and the manufacturing industry [36], [37]. Fig. 1 shows the methodological graph that will be followed in this research.

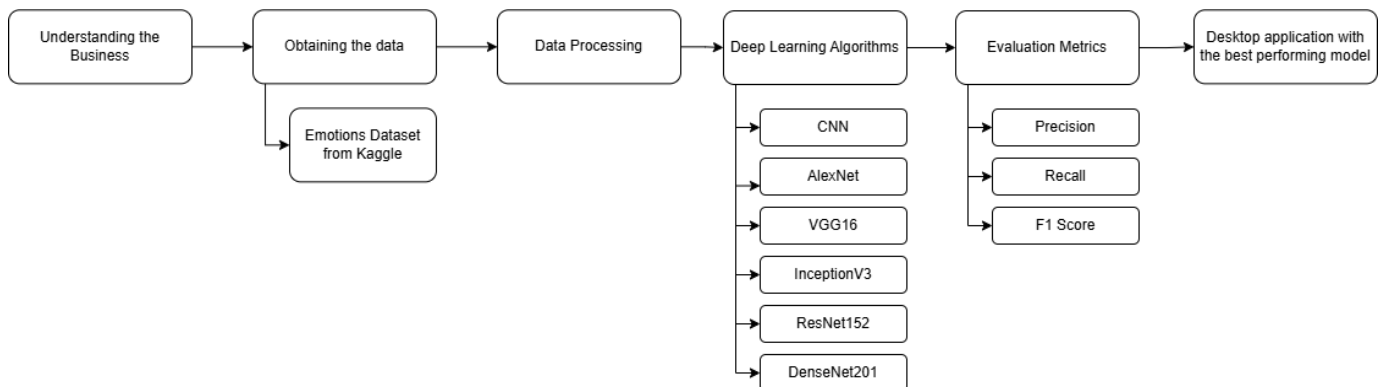


Fig. 1. Diagram of the methodology.

A. Understanding the Business

The present study aims to optimize the work environment and productivity of the organization by implementing a facial

recognition system based on deep learning models. To achieve this, the following key business objectives were identified:

- Improve the emotional well-being of employees in real-time.
- Detect and address negative emotions that may impact team productivity.
- Provide technological tools that support data-driven decision-making.
- Reduce work stress and foster a healthier organizational environment.
- Increase job satisfaction through timely interventions.

Evaluation of the Current Situation: Employee dissatisfaction and stress impact performance. Traditional methods are subjective, requiring an automated solution.

Research Population: A sample of 17 workers was selected for relevant data collection, ensuring that despite constraints.

Expected Impact: Real-time emotion detection enables timely interventions, improving the work environment and decision-making.

B. Obtaining the Data

The development of the facial recognition system for emotion detection began with the collection of high-quality data, using the public Kaggle database where a dataset of 35,960 images of faces with the emotions required for training was found. In addition, special attention was paid to the diversity of the images, ensuring the representation of different demographic groups and emotional expressions (happiness, anger, sadness, surprise, fear, disgust and neutrality). Fig. 2 shows the dataset used in this study.



Fig. 2. Images of dataset.

C. Data Processing

In this phase of development, a series of essential steps were taken to ensure that the data collected from the dataset was suitable for training Deep Learning models, maximizing their effectiveness and precision. First, images were filtered, eliminating those that were of low quality or had characteristics that could hinder learning, such as insufficient resolutions or irrelevant elements in the background. This process made it possible to work with high-quality data.

Subsequently, image resizing was implemented, adjusting all images to a standard size defined for the models, which ensured consistency and homogeneity in the system input. Data normalization was then conducted, adjusting pixel values to a

uniform range, which facilitated processing and improved the model's ability to identify significant patterns during training.

Additionally, data augmentation techniques such as rotation, shifting and horizontal flipping were applied to increase the diversity of the dataset. These modifications helped simulate various capture conditions, increasing the robustness and generalization capacity of the model.

Finally, the processed images were transformed into tensors, ensuring that the emotional labels remained correctly aligned with each image. This set of steps established a solid foundation for efficient training of the models, minimizing errors and improving their predictive ability. Furthermore, the emotions were divided into folders as shown in Fig. 3.

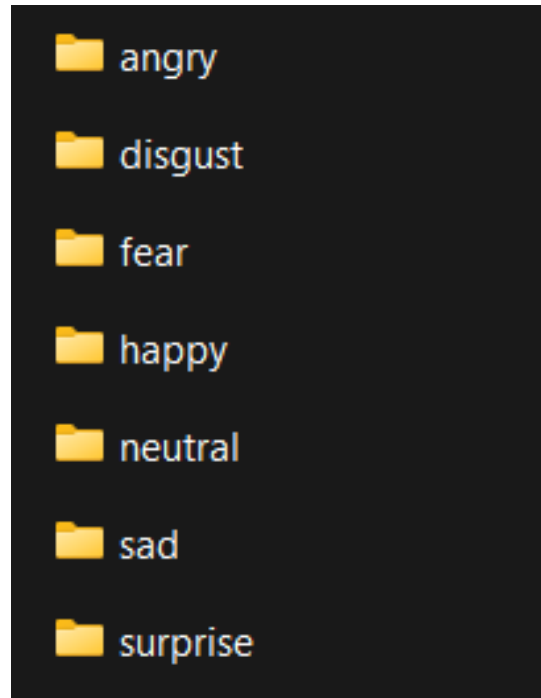


Fig. 3. Organization of folders by emotions.

D. Deep Learning Algorithms

In the modeling stage, six deep learning models were developed and trained for emotion detection from facial images: standard CNN, AlexNet, VGG16, DenseNet201, ResNet152, and InceptionV3. Each model was selected for its characteristics and proven performance in classification and computer vision tasks, allowing the problem to be approached from different architectural perspectives.

The CNN model was implemented in this study to see its level of prediction. On the other hand, AlexNet, known for being a pioneer in the use of deep networks, was included for its ability to extract relevant features through efficient convolutional layers. Likewise, VGG16, with its architecture based on small convolutional layers, allowed capturing fine details of the images, benefiting from greater depth.

As for the advanced models, DenseNet201 stood out for its dense structure, which connects each layer to all the following ones, maximizing the reuse of residual features and redundancy

in learning. Meanwhile, ResNet152 used residual connections to mitigate the gradient vanishing problem, effectively training a deep network. Finally, InceptionV3 incorporated convolutions of multiple sizes in a single layer, capturing information at different spatial scales, making it especially robust against the complexity of emotional expressions. Fig. 4 illustrates the trained models divided by folders.

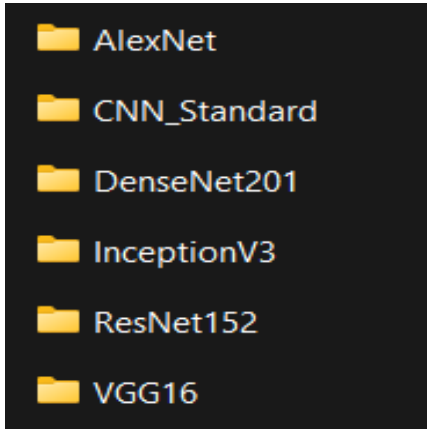


Fig. 4. Folder organization by trained model.

E. Evaluation Metrics

In this phase, the performance of the trained deep learning models for facial emotion detection was evaluated using the following quantitative metrics.

- Precision: Represents the percentage of correct predictions made by the model compared to the total number of predictions.

$$Precision = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

- Recall: Evaluates the model's ability to correctly detect positive instances, minimizing false negatives.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

- F1 Score: Combines precision and sensitivity in a harmonic average.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (3)$$

Where:

- TP: True Positive
- TN: True Negative
- FP: False Positives
- FN; False negatives (False Negative)

F. Performance Evaluation Instrument

To evaluate the effectiveness of the models, an experimental instrument was designed in which the 17 individuals (sample) participated. Each participant performed specific tests in which they were asked to express emotions such as happiness, sadness, anger, surprise, fear, disgust and neutrality towards the system, while the system tried to recognize them. The predictions made by each model were compared with real emotions, generating a confusion matrix for each architecture. This approach allowed recording detailed data on the performance of the models in a controlled environment.

Precision, recall, and F1-score metrics were manually calculated from the results obtained using standard formulas. This procedure ensures that the calculations are accurate and reflect the real performance of the models under experimental conditions.

G. Desktop Application with the Best Performing Model

This section aims to present and analyze the results obtained after the implementation of the face recognition system based on convolutional neural networks. It was first based on a comparison of six main architectures: standard CNN, AlexNet, VGG16, DenseNet201, ResNet152, and InceptionV3; comparing their performance in terms of precision, recall, and F1 score. This comparison allows us to identify the most suitable model for the final implementation, based on its ability to classify emotions accurately and efficiently.

In addition to the comparison of models, the effects of the implementation of the selected system in the work environment were analyzed, which included metrics related to productivity, organizational climate, and diagnosis and data collection times.

Table II shows the results of applying the metrics to each model, with DenseNet201 being the best performing model.

TABLE II. MODEL COMPARISON

Model	Precision	Recall	F1 Score
CNN	78.59%	86.60%	82.45%
AlexNet	77.87%	88.64%	82.90%
VGG16	80.59%	90%	84.98%
DenseNet201	87.70%	96.30%	91.80%
ResNet152	82.79%	93.48%	87.81%
InceptionV3	81.15%	88.17%	84.51%

H. Confusion Matrices

Below are the confusion matrices for each model used. These matrices complement the global metrics presented in the comparison table, providing a more granular view of the classification. The confusion matrices reflect the individual performance of each model in classifying emotions. In each matrix, the balance between correct predictions (main diagonal) and errors can be observed, highlighting the relative precision of the most robust models. The matrices can be seen in Fig. 5.

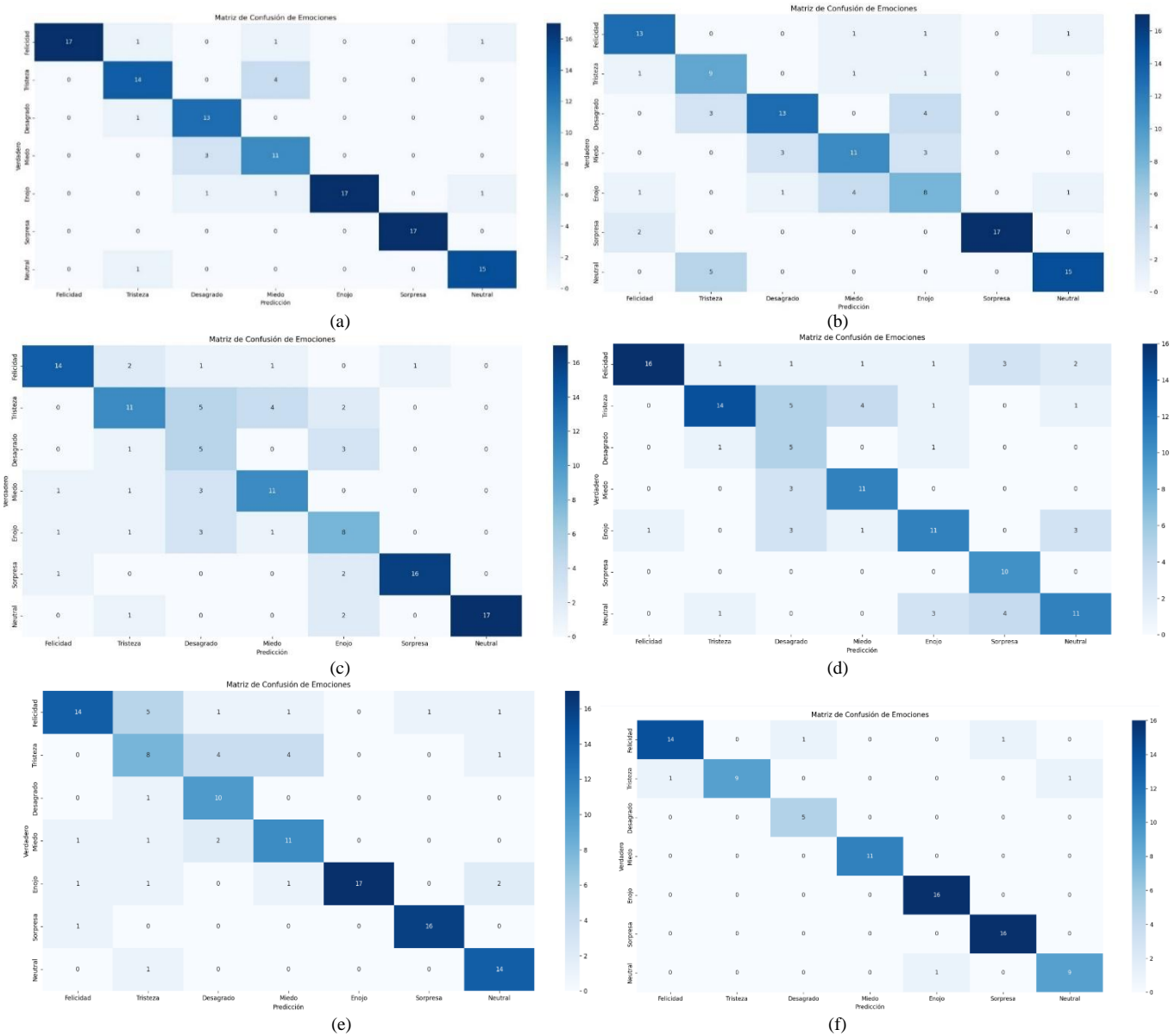


Fig. 5. (a) Confusion Matrix of DenseNet201. (b) Confusion Matrix of ResNet152. (c) Confusion Matrix of InceptionV3. (d) Confusion Matrix of AlexNet. (e) Confusion Matrix of VGG16. (f) Confusion Matrix of CNN.

IV. RESULTS

A. About the Prototype

The emotion detection application was developed using Python and the Tkinter library to create an intuitive and functional graphical user interface (GUI). Tkinter allowed the design of a visual environment where users could interact with the system in real time. The best performing model, DenseNet201, was implemented, pre-trained, and integrated into the application using libraries such as TensorFlow and OpenCV. This combination allowed the predictions to be processed quickly and accurately, ensuring that the user experience was efficient and aligned with the system's objectives (see Fig. 6).



(a)

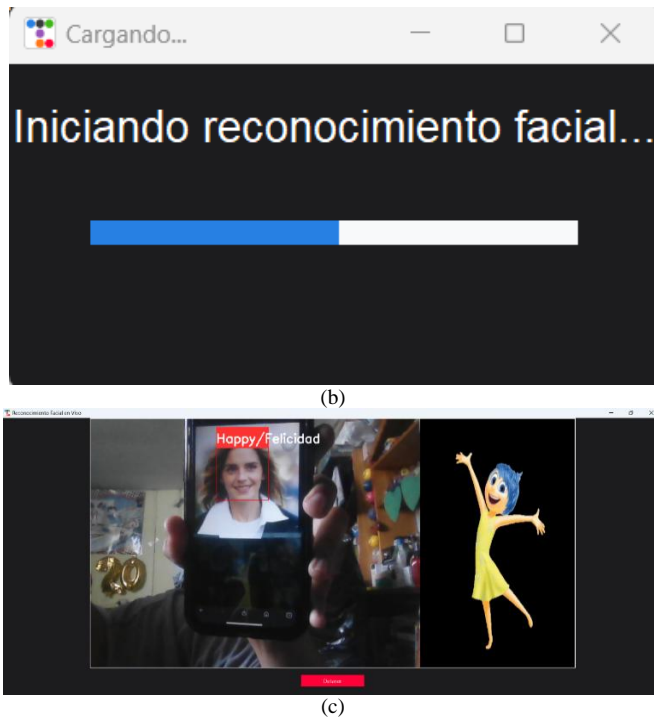


Fig. 6. (a) Startup interface (b) Progress bar (c) Running system.

B. About the Population

The research work approach is quantitative, so there was a population of 17 people, as specified in Table III.

TABLE III. WORKER'S POPULATION

Population	Number
Workers	17

In the present study, the samples require delimiting the population according to the available resources and the time allocated to conduct the research. For this reason, it was decided to use a non-probabilistic convenience sampling approach, which allows selecting participants intentionally, considering criteria such as accessibility, availability and ease of contact with them. This method is especially suitable in contexts where the total population is small. In this case, the same 17 people who make up the target population were selected, who meet the requirements to participate in the study, guaranteeing the obtaining of relevant data within the existing limitations, without compromising the validity of the analysis proposed in this document.

C. About the Indicators

This section presents the results obtained after the implementation of the facial recognition system based on deep learning, with the aim of optimizing the work environment and productivity in the organization. To evaluate the impact of the system, four key performance indicators (KPIs) were defined: i) Data collection time, ii) diagnosis time and iii) job satisfaction. These indicators were measured at two times: before and after the implementation of the system, using a pre-experimental design with a population of 17 workers. The results obtained reflect the effects of the system on operational effectiveness and

organizational well-being, providing empirical evidence on the contribution of emotional recognition technology in the workplace. Next, the results of each KPI are analyzed in detail, highlighting the most significant changes and their implication in organizational management.

1) *KPI - Data collection time:* The implementation of the automated facial recognition system has proven to be an effective solution to address the inefficiencies associated with manual emotional data collection. Before the intervention, the average time required for this process was 98.5 minutes, which implied not only a significant investment of time, but also a high dependence on human factors that could introduce biases and errors. With the integration of the proposed system, this time was drastically reduced to 27 minutes, which constitutes an improvement of 72.59%. This result demonstrates the transformative impact of AI-based technologies on organizational processes.

The optimization achieved is not limited only to time reduction; it also represents an advance in resource allocation, allowing staff to focus on strategic activities with greater added value. In addition, the automation of the process guarantees greater consistency and precision in data collection, eliminating variations derived from subjectivity or human limitations. This change not only improves operational efficiency but also strengthens the organization's ability to make informed decisions based on reliable data.

In terms of organizational impact, this reduction in data collection time has significant implications for overall productivity and responsiveness to emerging emotional issues. As illustrated in Fig. 7, the bar chart clearly compares the average times before and after implementation, providing a visual representation of the positive change achieved. This finding highlights how the adoption of advanced technologies can not only solve technical challenges but also contribute to the well-being of workers by freeing up time and resources to more effectively address their emotional needs in real time.

In summary, these results underline the strategic value of incorporating automated tools into organizational management. The ability to significantly reduce time, together with the improvement in the quality and consistency of the data collected, positions the facial recognition system as a key innovation to optimize both operational processes and the work environment in organizational environments.

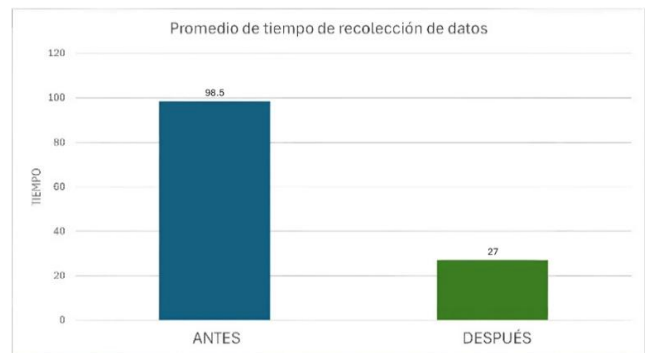


Fig. 7. Before and after bar graph.

2) *KPI - Diagnostic time*: Following the implementation of the facial recognition system, a substantial improvement in the efficiency of emotional diagnosis was seen, with a reduction in the average time from 93 minutes to 34 minutes, representing a 63.4% increase in the speed of the process. This result not only optimized the workflow but also increased the precision in emotion detection by reducing the reliance on traditional methods, such as manual interviews, which are often subject to human bias and variability in results. Automation allowed for greater uniformity in diagnoses, which is key to addressing emotional problems more quickly and effectively.

Furthermore, this advancement directly contributed to operational efficiency by freeing up resources that can now be allocated to higher-value strategic organizational activities. Fig. 8 graphically shows this significant reduction using a bar chart, highlighting the positive impact of the system not only in terms of time, but also on the quality of diagnosis and the organization's responsiveness to complex emotional challenges. This finding underlines the transformative role of artificial intelligence in improving critical organizational processes.

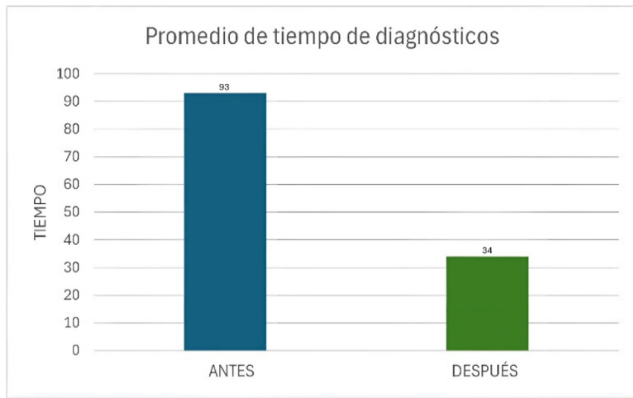


Fig. 8. Before and after bar graph.

3) *KPI - Job satisfaction*: One of the key indicators evaluated was the impact of the facial recognition system on employee job satisfaction, which is considered crucial for emotional well-being and organizational climate. To measure it, questionnaires with a 5-point Likert scale were applied to a sample of 17 employees, before and after implementing the system. The 15 questions in the questionnaire addressed aspects such as communication, leadership and emotional support.

The analysis consisted of comparing the averages of the pre- and post-implementation responses, allowing the identification of quantitative changes in the workers' perception of their work environment.

The graph in Fig. 9 illustrates the general trends, where the orange line (after) reflects a high and stable perception compared to the blue line (before), which shows lower and less consistent scores. This graph highlights the positive impact of the system on the organizational climate and employee well-being. Furthermore, the figures show that, before implementation, employee responses were more diverse. After the changes, there was evidence of higher job satisfaction and uniformity in the aspects evaluated. The decrease in negative

perceptions suggests an improvement in emotional stability and trust in the work environment, with a 66.59% increase in satisfaction.

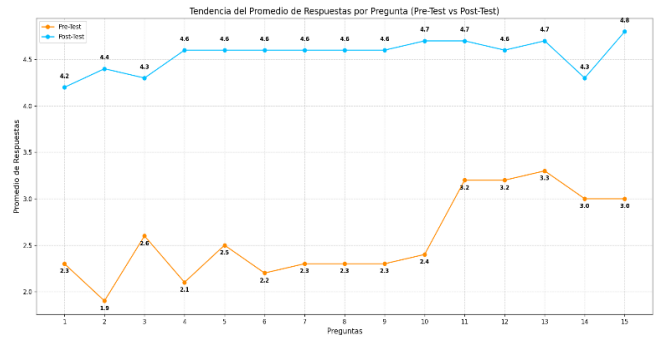


Fig. 9. Before and after response trends.

D. About the Survey and Expert Evaluation

To ensure the validity of the instrument designed to measure job satisfaction, the questionnaire was reviewed by three experts with experience in the development and evaluation of applied technological systems. The experts analyzed the structure, clarity and relevance of the questions, providing feedback to ensure that the instrument met the objectives of the study. This process allowed key adjustments to be made that strengthened the validity of the questionnaire, ensuring that the data collected accurately reflected the perception of the participants. Table IV presents the indicators that were used on the validity test approved by the experts. Where the following footage was graded:

- Poor (0 - 20%)
- Average (21 - 40%)
- Good (41 - 60%)
- Very Good (61 - 80%)
- Excellent (81 - 100%)

E. About the Methodology

The research was guided by the CRISP-DM model, recognized for its flexibility and focus on data mining and machine learning projects. This model divides the process into well-defined stages, allowing systematic progress from problem identification to the deployment of a functional solution.

The choice to take inspiration from CRISP-DM was based on its ability to adapt to the specific challenges of the research, such as handling large volumes of data and the need to train highly accurate deep learning models. Throughout this process, each stage was adapted to address key aspects of the project, such as the selection and preparation of the dataset, the training of advanced neural network models, and the implementation of a facial recognition system that meets the objectives of optimizing the work environment and productivity.

Table V presents the stages of the CRISP-DM-inspired methodology along with the activities carried out in each of them. This allows a clear visualization of how this approach was applied to ensure the effectiveness and reproducibility of the project.

TABLE IV. INDICATORS OF VALIDITY

Indicator	Criterion	Score		
		Expert 1	Expert 2	Expert 3
Clarity	It is formulated with appropriate language	80%	85%	90%
Objectivity	It is expressed in a coherent and logical manner	85%	85%	85%
Currentness	It is appropriate for advances in technology.	90%	85%	90%
Organization	There is a logical organization of variables and indicators.	80%	80%	80%
Sufficiency	It is coherent between indicators and dimensions	95%	90%	80%
Intentionality	It is appropriate to values and aspects related to the topic.	100%	82%	85%
Consistency	It is considered that the items used in this instrument are all and each one is specific to the field being investigated.	100%	90%	80%
Coherence	It is considered that the structure of this instrument is appropriate to the type of user to whom the instrument is directed.	85%	80%	90%
Methodology	The strategy responds to the purpose of the research.	85%	80%	85%
Relevance	It is appropriate to deal with the research topic.	80%	81%	85%

TABLE V. METHODOLOGY USED

Stage	Activity Completed
Understanding the Business	Problem identification: Job dissatisfaction and productivity; literature review to define objectives and approaches.
Obtaining the Data	Selection of the emotional images dataset (Kaggle), analysis of its structure and quality, initial data cleaning.
Data Processing	Normalization of pixel values, image resizing, data augmentation (rotation, flip, shift) and organization into folders according to emotions.
Deep Learning Algorithms	Training models such as standard CNN, DenseNet201, VGG16, ResNet152, AlexNet and InceptionV3.
Evaluation Metrics	Comparison of key metrics: accuracy, sensitivity and F1 score; analysis of confusion matrices to validate the effectiveness of each model.
Desktop App with the best performing model	Implementation of the DenseNet201 model in a functional prototype with a graphical interface, using Python and Tkinter.

This approach allowed theory to be integrated with practice in a structured manner, facilitating the achievement of the

research objectives and the validation of the results obtained. The table summarizes the essential steps that led to the success of the project, highlighting the rigor at each stage of the proposed methodology.

V. DISCUSSION

A. About KPI's

The results of this study show significant improvements in the operational efficiency of this company in the wholesale sector after the implementation of the facial recognition system based on convolutional neural networks. Significant reductions were recorded in data collection times with an improvement of 72.59%, emotional diagnosis with an improvement of 63.4% and job satisfaction with 66.59%, optimizing the company's internal processes. In addition, the analysis showed an increase in productivity, reflecting a positive impact on employee performance.

B. About the Models

This chapter analyzes the results obtained with the system implemented using DenseNet201, highlighting both productivity and work environment. The findings are compared with previous research, such as those that achieved 85.69% accuracy in RAF-DB using residual networks applied to low-resolution environments [12], and studies that achieved 85.82% accuracy with EfficientNetB0 on the FER2013 dataset [38]. Likewise, the integration of DeepFace in smart factories was explored, evidencing its ability to adapt to real-time scenarios [31]. In contrast, other research reported 66.85% accuracy with CNN in FER2013, highlighting the challenges associated with generalizing less-represented emotions [26].

On the other hand, the Bayesian CNN-LSTM model demonstrated outstanding performance on metrics such as accuracy, sensitivity, and F1-score, underlining its effectiveness in correlating emotions expressed in forums with the dropout rate in MOOCs [39]. Similarly, the model proposed in this study employs sequential and residual identity blocks, allowing for high-accuracy facial feature extraction, outperforming other state-of-the-art methods, especially in distance education contexts [40]. Furthermore, CNN-based architecture designed for gender and emotion classification have achieved accuracy levels above 98%, successfully addressing challenges such as emotional changes reflected in vocal features [41].

Finally, quantum convolutional networks (QCNNs) have shown significant improvements in the efficiency of emotional detection systems in mental health contexts [19]. These advances highlight the importance of factors such as controlled environments and robust computational resources to maximize the performance of these systems. Furthermore, recurrent neural networks (RNNs) have recorded 95% accuracy in classifying emotions in videos, evidencing their ability to capture complex temporal patterns in dynamic data [42]. Together, these studies illustrate how advanced technologies can transform emotional detection, successfully addressing the limitations and challenges present in this field.

C. About Limitations

First, the research was conducted in a controlled setting with a small sample size of 17 participants, limiting the

generalizability of the findings to larger populations. While the data set ensured multicultural representation, the system was evaluated within a single cultural context.

VI. CONCLUSIONS AND FUTURE WORK

In conclusion, the implementation of a facial recognition system based on convolutional neural networks (DenseNet201) in the context of a wholesale company has proven to be an effective tool to comprehensively address the challenges related to emotional management in work environments. This system allowed real-time monitoring of employees' emotions, which facilitated the early identification of negative emotions and enabled the implementation of timely interventions aimed at optimizing the organizational climate and promoting a healthier and more productive work environment.

A significant decrease was also observed in the time associated with data collection and emotional diagnosis, which translated into a substantial improvement in the efficiency of processes related to the management of workplace well-being. These findings highlight the transformative potential of AI-based technologies to promote more resilient, efficient and human-centered work environments.

Future studies could focus on evaluating the adaptability and applicability of this system in different sectors and work contexts, with the aim of validating its effectiveness and exploring new areas for improvement. Additionally, it is recommended that they can further enhance the system's capacity to address the dynamic challenges of emotional management in the organizational setting.

ACKNOWLEDGMENT

To our parents for giving us the opportunity to study and fulfill our dreams of becoming honest professionals with values; in addition to always being our guide in this difficult stage and always giving us the hope that the world can change if one is willing to give it there; to them, thank you very much.

To the Universidad Privada del Norte (UPN) for providing us with quality education during these five years of the computer systems engineering degree.

REFERENCES

- [1] World Health Organization, "Mental health: strengthening our response," World Health Organization. Accessed: Dec. 09, 2024. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>
- [2] R. Y. Curiel Gómez, J. Chiquillo Rodelo, and D. Muñoz Rojas, "Management of public policies in mental health in the Colombian work context," *Venezuelan Journal of Management*, vol. 29, no. 106, pp. 847–864, Mar. 2024, doi: 10.52080/RVGLUZ.29.106.25.
- [3] JW Bonny, "Self-report and facial expression indicators of team cohesion development," *Behav Res Methods*, vol. 55, no. 1, pp. 1–15, Jan. 2023, doi: 10.3758/S13428-022-01799-3/TABLES/3.
- [4] S. Zhang, J. Chen, Y. Wen, H. Chen, Q. Gao, and Q. Wang, "Capturing regulatory patterns in online collaborative learning: A network analytic approach," *Int J Comput Support Collab Learn*, vol. 16, no. 1, pp. 37–66, Mar. 2021, doi: 10.1007/S11412-021-09339-5/TABLES/15.
- [5] M. Bascuñana, "Managing emotions in the workplace," *Affor Health*. Accessed: Sep. 27, 2024. [Online]. Available: <https://afforhealth.com/managing-emotions-in-the-workplace-health/>
- [6] L. Chávez Quispe, "Burnout continues to rise in Peru: 78% of workers say they experience it," *Forbes Peru*. Accessed: Sep. 01, 2024. [Online]. Available: <https://forbes.pe/capital-humano/2023-11-10/el-burnout-sigue-en-ascenso-en-peru-el-78-de-trabajadores-afirma-experimentarlo>
- [7] M. Ríos, "Workers with job stress: How high is their intention to quit? | Burnout syndrome | People management," *Gestión*, NEWS MANAGEMENT, May 28, 2024. Accessed: Sep. 01, 2024. [Online]. Available: <https://gestion.pe/economia/management-empleo/trabajadores-con-estres-laboral-que-tan-alta-es-su-intencion-de-renunciar-sindrome-de-burnout-gestion-de-personas-recursos-humanos-noticia/>
- [8] L. Montag, R. Mcleod, L. De Mets, M. Gauld, F. Rodger, and M. Pelka, "The rise and rise of biometric mass surveillance in the EU," 2021.
- [9] Alicio, "What is facial recognition and what is it for?," *Alice Biometrics*. Accessed: Jun. 10, 2024. [Online]. Available: <https://alicebiometrics.com/para-que-sirve-el-reconocimiento-facial/>
- [10] X. Wu et al., "Emotion Recognition Using Convolutional Neural Network (CNN)," *J Phys Conf Ser*, vol. 1962, no. 1, p. 012040, Jul. 2021, doi: 10.1088/1742-6596/1962/1/012040.
- [11] M. Wang, P. Tan, X. Zhang, Y. Kang, C. Jin, and J. Cao, "Facial expression recognition based on CNN," *J Phys Conf Ser*, vol. 1601, no. 5, Aug. 2020, doi: 10.1088/1742-6596/1601/5/052027.
- [12] JL Gómez-Sirvent, F. López de la Rosa, MT López, and A. Fernández-Caballero, "Facial Expression Recognition in the Wild for Low-Resolution Images Using Voting Residual Network," *Electronics* 2023, Vol. 12, Page 3837, vol. 12, no. 18, p. 3837, Sep. 2023, doi: 10.3390/ELECTRONICS12183837.
- [13] S. Teja Chavali, C. Tej Kandavalli, T. M. Sugash, and R. Subramani, "Smart Facial Emotion Recognition With Gender and Age Factor Estimation," *Procedia Comput Sci*, vol. 218, pp. 113–123, Jan. 2023, doi: 10.1016/J.PROCS.2022.12.407.
- [14] UR Villanueva, JCG Delión, and FP Larroca, "Recognition of facial expressions and personal characteristics as a tool to identify people in a public transport system," *Industrial Engineering*, pp. 261–277, Apr. 2022, doi: 10.26439/ING.IND2022.N.5811.
- [15] WB Putra and F. Arifin, "Real-Time Emotion Recognition System to Monitor Student's Mood in a Classroom," *J Phys Conf Ser*, vol. 1413, no. 1, p. 012021, Nov. 2019, doi: 10.1088/1742-6596/1413/1/012021.
- [16] J. Atwan, M. Wedyan, D. Albashish, E. Aljaafrah, R. Alturki, and B. Alshawi, "Using Deep Learning to Recognize Fake Faces," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 1, pp. 1144–1155, 2024, doi: 10.14569/IJACSA.2024.01501113.
- [17] Á. González-Almansa Laredo, "Speaker-independent emotion identification system based on convolutional neural networks," *Polytechnic University of Madrid*, May 2023.
- [18] C. Wang, "Emotion Recognition of College Students' Online Learning Engagement Based on Deep Learning," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 17, no. 06, pp. 110–122, Mar. 2022, doi: 10.3991/IJET.V17I06.30019.
- [19] S. Hossain, S. Umer, RK Rout, and H. Al Marzuqi, "A Deep Quantum Convolutional Neural Network Based Facial Expression Recognition For Mental Health Analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 1556–1565, 2024, doi: 10.1109/TNSRE.2024.3385336.
- [20] M. Wafi, FA Bachtiar, and F. Utaminigrum, "Feature extraction comparison for facial expression recognition using adaptive extreme learning machine," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 1, pp. 1113–1122, Feb. 2023, doi: 10.11591/IJECE.V13I1.PP1113-1122.
- [21] MRM Asemawi, MH Mutar, EH Ahmed, HO Hanoosh, and AH Abbas, "Emotions recognition from human facial images based on fast learning network," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 3, pp. 1478–1487, Jun. 2023, doi: 10.11591/IJECS.V30.I3.PP1478-1487.
- [22] AP Canazas, JJR Blaz, PDT Martínez, and XJ Mamani, "Emotion identification system through facial recognition using artificial intelligence," *Innovation and Software*, vol. 3, no. 2, pp. 140–150, Sep. 2022, doi: 10.48168/innosoft.s9.a74.
- [23] O. Shaughnessy et al., "Facial Emotion Recognition for Photo and Video Surveillance Based on Machine Learning and Visual Analytics," *Applied*

- Sciences 2023, Vol. 13, Page 9890 , vol. 13, no. 17, p. 9890, Aug. 2023, doi: 10.3390/APPI13179890.
- [24] AT Lopes, E. de Aguiar, AF De Souza, and T. Oliveira-Santos, "Facial expression recognition with Convolutional Neural Networks: Coping with few data and the training sample order," *Pattern Recognit* , vol. 61, pp. 610–628, Jan. 2017, doi: 10.1016/J.PATCOG.2016.07.026.
- [25] X. Shaoyuan, Y. Cheng, Q. Lin, and J. Allebach, "Emotion recognition using convolutional neural networks," *IS and T International Symposium on Electronic Imaging Science and Technology* , vol. 2019, no. 8, Jan. 2019, doi: 10.2352/ISSN.2470-1173.2019.8.IMAWM-402.
- [26] X. Wu et al. , "The application of deep learning in computer vision: Facial emotion recognition based on convolutional neural network," *J Phys Conf Ser* , vol. 2646, no. 1, p. 012020, Dec. 2023, doi: 10.1088/1742-6596/2646/1/012020.
- [27] AAH Qutub and Y. Atay, "Deep learning approaches for emotion recognition classification based on facial expressions," *Nexo Scientific Journal* , vol. 36, no. 05, pp. 1–18, Nov. 2023, doi: 10.5377/NEXO.V36I05.17181.
- [28] M. Miranda-Leon and RA Toala-Dueñas, "Emotion detection in speeches using machine learning," *593 Digital Publisher CEIT* , vol. 9, no. 4, pp. 72–101, Jul. 2024, doi: 10.33386/593dp.2024.4.2367.
- [29] S.G. Tejashwini and D. Aradhana, "Multimodal Deep Learning Approach for Real-Time Sentiment Analysis in Video Streaming," *International Journal of Advanced Computer Science and Applications* , vol. 14, no. 8, pp. 730–736, 2023, doi: 10.14569/IJACSA.2023.0140881.
- [30] M. Sam'an, Safuan, and M. Munsarif, "Convolutional neural network hyperparameters for face emotion recognition using genetic algorithm," *Indonesian Journal of Electrical Engineering and Computer Science* , vol. 33, no. 1, pp. 442–449, Jan. 2024, doi: 10.11591/ijeecs.v33.i1.pp442-449.
- [31] A. Chiurco et al. , "Real-time Detection of Worker's Emotions for Advanced Human-Robot Interaction during Collaborative Tasks in Smart Factories," *Procedia Comput Sci* , vol. 200, pp. 1875–1884, Jan. 2022, doi: 10.1016/J.PROCS.2022.01.388.
- [32] C. Schröer, F. Kruse, and JM Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput Sci* , vol. 181, pp. 526–534, Jan. 2021, doi: 10.1016/J.PROCS.2021.01.199.
- [33] C. Schröer, F. Kruse, and JM Gómez, "A Systematic Literature Review on Applying CRISP-DM Process Model," *Procedia Comput Sci* , vol. 181, pp. 526–534, Jan. 2021, doi: 10.1016/J.PROCS.2021.01.199.
- [34] N. Azadeh-Fard, FM Megahed, and F. Pakdil, "Variations of length of stay: A case study using control charts in the CRISP-DM framework," *International Journal of Six Sigma and Competitive Advantage* , vol. 11, no. 2–3, pp. 204–225, 2019, doi: 10.1504/IJSSCA.2019.101418.
- [35] A. Dåderman and S. Rosander, "Evaluating Frameworks for Implementing Machine Learning in Signal Processing A Comparative Study of CRISP-DM, SEMMA and KDD," *Vetenskap Och Konst* , 2018.
- [36] E. Gholamzadeh Nabati and KD Thoben, "On applicability of big data analytics in the closed-loop product lifecycle: Integration of CRISP-DM standard," *IFIP Adv Inf Commun Technol* , vol. 492, pp. 457–467, Mar. 2017, doi: 10.1007/978-3-319-54660-5_41/TABLES/3.
- [37] F. Schafer, C. Zeiselmaier, J. Becker, and H. Otten, "Synthesizing CRISP-DM and Quality Management: A Data Mining Approach for Production Processes," *IEEE International Conference on Technology Management, Operations and Decisions, ICTMOD 2019* , pp. 190–195, Apr. 2019, doi: 10.1109/ITMC.2018.8691266.
- [38] M. Anand and S. Babu, "Multi-class Facial Emotion Expression Identification Using DL-Based Feature Extraction with Classification Models," *International Journal of Computational Intelligence Systems* , vol. 17, no. 1, pp. 1–17, Dec. 2024, doi: 10.1007/S44196-024-00406-X/TABLES/5.
- [39] K. Mrhar, L. Benhiba, S. Bourekache, and M. Abik, "A Bayesian CNN-LSTM Model for Sentiment Analysis in Massive Open Online Courses MOOCs," *International Journal of Emerging Technologies in Learning (iJET)* , vol . 16, no. 23, pp. 216–232, Dec. 2021, doi: 10.3991/IJET.V16I23.24457.
- [40] ABS Salamh and HI Akyüz, "A New Deep Learning Model for Face Recognition and Registration in Distance Learning," *International Journal of Emerging Technologies in Learning (iJET)* , vol. 17, no. 12, pp. 29–41, Jun. 2022, doi: 10.3991/IJET.V17I12.30377.
- [41] TM Taha, Z. Ben Messaoud, and M. Frikha, "Convolutional Neural Network Architectures for Gender, Emotional Detection from Speech and Speaker Diarization," *International Journal of Interactive Mobile Technologies (ijim)* , vol. 18, no. 03, pp. 88–103, Feb. 2024, doi: 10.3991/IJIM.V18I03.43013.
- [42] Prathwini and Prathyakshini, "DeepEmoVision: Unveiling Emotion Dynamics in Video Through Deep Learning Algorithms," *International Journal of Advanced Computer Science and Applications* , vol. 15, no. 3, pp. 885–892, 2024, doi: 10.14569/IJACSA.2024.0150388.

Sentiment and Emotion Analysis with Large Language Models for Political Security Prediction Framework

Liyana Safra Zaabar, Adriana Arul Yacob, Mohd Rizal Mohd Isa,
Muslihah Wook, Nor Asiakin Abdullah, Suzaimah Ramli, Noor Afiza Mat Razali*
Faculty of Defence Science & Technology, National Defence University of Malaysia, Kuala Lumpur

Abstract—The increasing spread of textual content on social media, driven by the rise of Large Language Models (LLMs), has highlighted the importance of sentiment analysis in detecting threats, racial abuse, violence, and implied warnings. The subtlety and ambiguity of language present challenges in developing effective frameworks for threat detection, particularly within the political security domain. While significant research has explored hate speech and offensive content, few studies focus on detecting threats using sentiment analysis in this context. Leveraging advancements in Natural Language Processing (NLP), this study employs the NRC Emotion Lexicon to label emotions in a political-domain social media dataset. *TextBlob* is used to extract sentiment polarity, identifying potential threats where anger and fear intensities exceed a threshold alongside negative sentiment. The Bidirectional Encoder Representations from Transformers (BERT) was applied to enhance threat detection accuracy. The proposed framework achieved an Area Under the ROC Curve (AUC) of 87%, with the BERT model achieving 91% accuracy, 90.5% precision, 81.3% recall and F1-score of 91%, outperforming baseline models. These findings demonstrate the effectiveness of sentiment and emotion-based features in improving threat detection accuracy, providing a robust framework for political security applications.

Keywords—Political security; large language models; sentiment analysis; emotion analysis; BERT; threat prediction

I. INTRODUCTION

Today, cyberspace has proven to be a very powerful tool and can impact national security. As new risks emerge, there is interest in developing more advanced defence strategies[1]. The current response to this problem is too limited, as it cannot capture the scale of information sharing that big data analyses. In this cyber arena, various platforms become addresses for various types of transactions including emotional transactions among the public [1]. These feelings can trigger great security concerns, such as the real-world example at the beginning of the Arab Spring, where the spread of false information online exacerbated negative attitudes and led to social instability that endangered national security[1], [2]. The events of the Arab Spring are one round of examples showing how emotions affect social and political stability. People are emotional creatures, and the emotion of anger or fear can mobilize people towards collective action, even disruptive to society. For example, anger has been demonstrated to lead more often to approach behaviour (increasing social movement participation), while fear leads most frequently to avoidance

[2]. Indeed, these emotional responses to political will unravel even relatively stable societies as seen in the Arab Spring protests. The collapse of stability in the Middle East and North Africa region caught almost all Western policy makers off guarded, but perhaps a paralytic counselling should be devoted to emotional matters within political and social regulation [3].

Many platforms accommodate a wide variety of different types of data exchange in cyberspace, including a wide range of public emotional expressions. These emotions can pose security risks, as evidenced by events like the Arab Spring, where negative sentiments were fuelled by misinformation online, which eventually led to societal unrest that threatened national security. As such, promptly detecting disruptive sentiments like these is essential for authorities to effectively manage crises. However, existing methodologies for emotional evaluation regarding national security are inadequate [9]. While most researchers explore various techniques for classifying human emotions, there is insufficient attention given to connecting these emotions to security threats and developing appropriate measurement mechanisms. Despite the capability of sentiment analysis methods to ascertain word polarity, their application in predicting threats remains largely unexplored, particularly in the realm of political security[5], [9].

Existing research on sentiment analysis largely focuses on hate speech and offensive content, with limited attention to detecting political security threats. While sentiment analysis effectively monitors public sentiment, it often fails to connect emotions like anger and fear to security threats within dynamic environments such as social media. Furthermore, traditional models struggle to capture the contextual and sequential dependencies necessary for identifying evolving threats. This study addresses these gaps by leveraging a BERT-based framework enhanced with the NRC Emotion Lexicon to improve the accuracy of political threat detection. By integrating advanced sentiment and emotion analysis, this research provides a robust model for enhancing political security prediction and contributes to bridging critical gaps in existing literature.

This paper is organized as follows: Section II reviews related work, establishing the relevance of the study and highlighting the contributions of the proposed approach. Section III describes the methods and materials, including details on the dataset, data preprocessing, word embedding

*Corresponding Author.

techniques, and the proposed BERT-based model for detecting political security threats. Section IV presents the results, offering a comparative analysis of the model's performance relative to existing approaches. Section V discusses the comparative findings and explores infrastructure requirements and future directions for deploying large language models in secure and efficient environments. Finally, Section VI concludes with the key findings and their implications for advancing political security threat prediction frameworks.

II. RELATED WORK

Authors in study [13] developed a global cyber-threat intelligence system using Conventional Neural Network and employed sentiment analysis techniques to detect global threats. In this study, the Bidirectional Encoder Representations from Transformers (BERT) model is used to address limitations inherent in traditional Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks when capturing long-term dependencies in sequential data.

Traditional RNNs suffer from a vanishing gradient problem, in which gradients will diminish exponentially over time, making it difficult for the network to learn long-range dependencies [14]. While LSTMs mitigate this issue by introducing a memory cell with a sophisticated structure [12], BERT outperforms both RNNs and LSTMs by leveraging a transformer-based architecture that effectively models contextual relationships across the entire sequence, eliminating the constraints of sequential processing [15].

Emotions are known to have an important impact on human cognitive processes and decision-making [4]. Intelligence has been influenced by emotion, as they contribute for several underlying cognitive skills such as salience detection, decision making and adaptation in a critical way [4]. An interesting discovery is that non-compliant behaviour against security procedures can be influenced by emotions, with four main emotional traits in the field of information security domain including rage, trust fear and stress. Fear has been recognised as a foundational principle for keeping people using security measures. The emotion-based security threat detection is not meant to completely replace the traditional measures but can act as a complement strategy by providing an additional layer of intelligence and adaptability [5].

Recent studies highlight those emotions detected through sentiment analysis, specifically from social media, play an important role in identifying potential security threats. For example, the massive use of sentiment analysis on user-generated content such as tweets or Facebook posts has proven useful in monitoring public sentiment that may signal social instability, the spread of disinformation, or other factors that could threaten national security [6]. Emotional sentiment can be an early indicator of chaos, allowing authorities to intervene early before tensions escalate, as is the case in events such as the Arab Spring [7], [8]. Advances in sentiment analysis have leveraged machine learning and artificial intelligence (AI) technologies to not only detect emotional changes in social media content but also predict potential security concerns. By applying natural language processing (NLP) techniques, the

system can detect changes in emotional tone or intensity among a broad population, identifying issues such as anger, fear, or frustration that may lead to greater social threats [8].

Recent research has pointed out that emotions detected with sentiment analysis, especially those emanating from social media, become highly important in the process of identifying potential security threats. The high-volume applications of sentiment analysis on user-generated content such as tweets or Facebook posts have proven useful in the monitoring of public sentiments indicative of social instability, disinformation dissemination, and other factors that can threaten national security [6]. Therefore, it may be said that emotional sentiment can act as an early warning signal for chaos, and the authorities can intervene at an early stage before tensions escalate, as in events like the Arab Spring [7], [8]. Advanced sentiment analysis harnessed the power of machine learning and AI technologies to identify not only emotional changes in content on social media but also predict possible security concerns. Using the methods of NLP, it can observe changes in emotional tone or intensity of the greater population and pinpoint problems such as anger, fear, or frustration that might be a larger threat to society.

Negative emotions such as anger, fear, disgust and anxiety have been identified as potential indicators of security threats, especially when these emotions are widely expressed in public or digital spaces [9]. For example, anger is often associated with social discontent and can trigger collective actions such as protests or riots, especially in politically unstable contexts. According to studies, anger can be a trigger for aggressive behaviour when an individual or group feels marginalized or oppressed. This shift from emotion to action has been observed in many cases of social instability, where negative sentiment on social media is closely linked to violence or instability in the real world [8], [17].

Fear and related emotions such as extreme fear and anxiety also play an important role in predicting threats. While fear is not an immediate threat, it can cause destabilizing reactions, such as panic buying, mass evacuation, or rioting, especially when influenced by the spread of false information or rumours [18]. Studies in behavioural psychology and security frameworks emphasize that fear increases the perception of vulnerability, making it a useful tool for predicting crises and emergency response strategies [19]. Disgust, often fuelled by moral or ethical outrage, can also increase social division and instability, further contributing to security risks. Sentiment analysis tools, when used to monitor these emotions, have become increasingly accepted for predicting and mitigating threats before they escalate [5]. Author in [9] proposed that emotion is a key variable in determining an opinion or sentiment. Emotions such as anger, fear, disgust, fear, and anxiety can serve as emotional indicators in determining the existence of political security threats in text data. These emotions, as studies have determined, are closely related to the political security domain and have the potential to trigger political events such as riots, coups, terrorism, international wars, civil wars, and political elections, which can lead to negative sentiments or opinions. These opinions or sentiments can be analysed to predict threats in the political security domain [9], [20].

The selection of the proposed BERT-based framework is driven by its ability to overcome the key constraints of existing models, such as traditional RNNs and LSTMs, which face difficulties in addressing long-term dependencies and understanding of context. Unlike this approach, BERT uses a transformer architecture to effectively model data sequentially and capture more subtle relationships in text content. Additionally, while previous models primarily focused on general sentiment analysis or the detection of specific offensive content, they lacked the ability to link emotional intensity such as fear and anger to political security threats. By integrating Lexicon's NRC Emotion and sentiment polarity analysis, this proposed framework bridges the gap, providing better accuracy and adaptability for identifying threats. To prove the effectiveness of this framework, comprehensive verification measures and comparative analysis with existing methods were conducted, emphasizing its superiority in addressing the complexity of predicting political security threats.

III. METHODS AND MATERIALS

In our research, we developed the BERT LLM Political Security Model to predict various threats from online platforms. Our model comprises several key stages, including data pre-processing and cleansing, word embedding and threat classification and detection. In this stage, BERT is used for threat detection by adopting sentiment analysis method. Fig. 1 illustrates the Workflow of BERT LLM Political Security Model.

A. Dataset

In this experimental design, we used the labelled dataset provided by [9]. This dataset was originally created by manually gathering various Malaysian online news sources, such as The Star, New Straits Times (NST), and Free Malaysia Today (FMT), and more. The dataset comprises 250 texts from

online news sources. Out of these texts, 163 of them are categorized as positive, while the remaining 87 are categorized as negative. These positive and negative classifications serve as markers to ascertain the presence of threats within the sample texts.

B. Pre-processing

In Natural Language Processing (NLP), text pre-processing is a process that will enhance classifier performance and reduce feature complexity [10]. In this process, unnecessary elements such as punctuation, HTML codes, and symbols are removed, and the gathered text data is then transformed to lowercase and normalized. The normalization process consists of two main steps. First, the unstructured text dataset is converted into a structured word vector, and then, the feature vector's dimensionality is reduced by eliminating unwanted words and stemming them to their original forms. Stemming refers to reducing words to their roots, while lemmatization is the act of utilizing a lexical knowledge base to convert words to their base forms by rooting verbs. At the end of the process, words will be encoded into numerical formats [11].

C. Word-Embedding

Word embedding is a technique used in NLP and deep learning to represent words as dense vectors of real numbers [12]. It is a way to map words to vectors in a continuous vector space, where similar words are represented by similar vectors. This experimental layer carries max_words as an input dimension, with 50 as the output dimension (embedding size), and max_len as the input length. This layer creates a low dimensional vector that deals with each word in the input sequences and directly replaces them with their dense vector representation. These vectors are then multiplied from the embedding layer container and are then sent to the BERT layer for further processing.

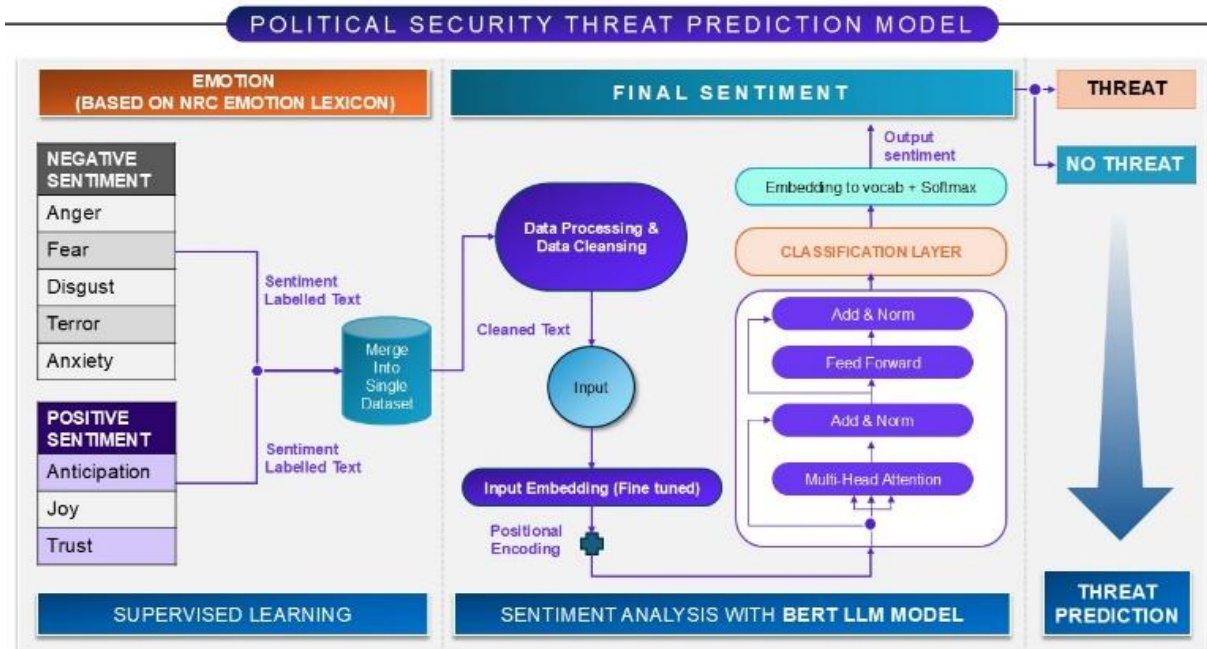


Fig. 1. Workflow of BERT LLM political security model.

D. Threat Prediction Using BERT

To validate our proposed model, an experimental analysis was conducted. The dataset for this research is constructed by collecting text data from various online news platforms, and the proposed model seeks to open new research avenues at the intersection of sentiment analysis and national security. The model will achieve this to enhance emotion measurement and threat prediction in cyberspace.

Fig. 2 illustrates the political security prediction model leveraging NRC Emotion Lexicon [15] and BERT [16] model for threat classification and prediction.

An experiment was conducted on a PC with an Intel® Core™ i7-8650U CPU @ 1.90GHz, 2.11 GHz, 8GB of RAM, a 64-bit OS, and an x64-based processor to test the developed model in Python 3.11.1 environment. 250 sentences from Malaysian online news sources were used as the main labelled dataset after the cleansing and data preparation process. To gauge the efficacy of our models, we compared them to the model developed by the researchers cited in reference [9]. We selected their model for comparison because it addresses the same task as our study, which is identifying threats in the political domain, and because it also utilizes the same dataset for evaluation, which is data that is derived from Malaysian online news.

The final phase of the research design is to validate the analyzed data. This study demonstrated a comparative performance evaluation to validate the proposed theoretical framework in this phase. The results of the evaluation test compare precision, recall, accuracy and F-measure [21]. The performance measure involves calculation of the accuracy, precision and recall value of the test dataset. The evaluation process commenced after the labelling of data into either

positive or negative classes, or the imbalance between the class proportions was addressed. A random subset of sentences was selected, to train and test it with the BERT transformer-based model. The employment of confusion matrix and the computation of accuracy, precision, and recall are the strong indicators to evaluate the performance of the chosen model based on the training data. The formula for accuracy, as shown in Eq. (1), is the proportion of correctly predicted opinions out of all input opinions to the classifier. This proportion is determined by true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision is shown in Eq. (2) and is the percentage of true cases of an opinion (of an instance) among all the classified cases of the opinions (of all instances). To determine the accuracy, true positive rate (TP) was used, as shown in the formula below.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Recall is defined as the proportion of properly categorized occurrences of a polarity over the total number of correct instances of the polarity. The formula to calculate the recall values using TP and FN is shown in Eq. (3):

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

The F-score is calculated by dividing the number of true positives by the sum of true positives and false positives, as shown in Eq. (4).

$$F - score = \frac{2TP}{2TP+FP+FN} \quad (4)$$

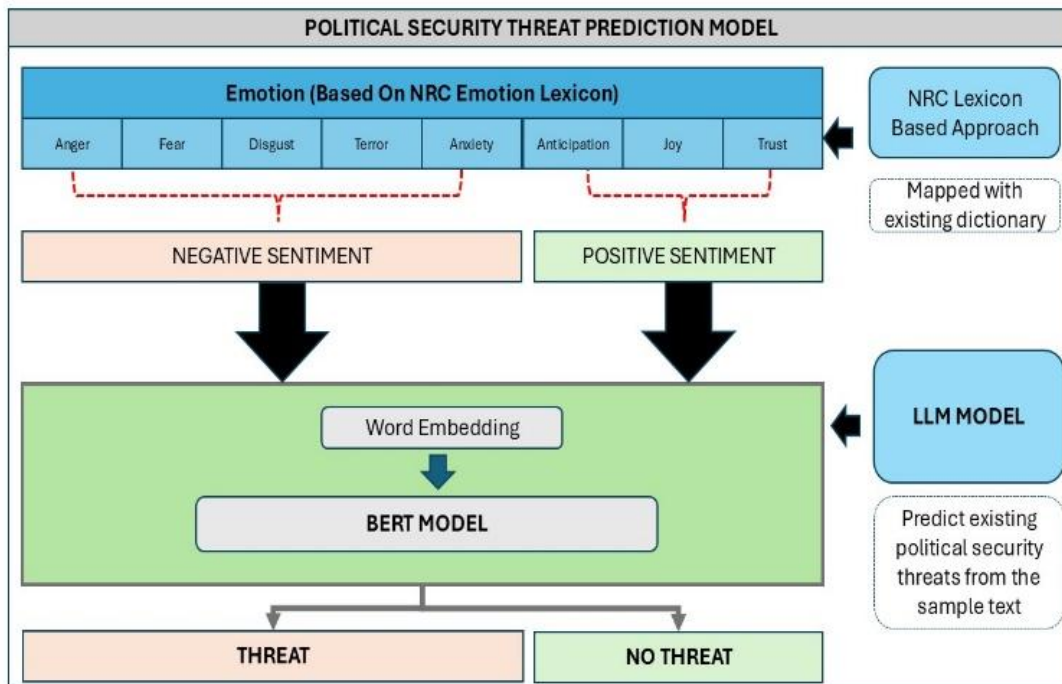


Fig. 2. Political security threat prediction model.

IV. RESULTS

A. Comparative Output and Benchmarking

The performance of the current model, BERT LLM, proposed model in political security domain, was benchmarked against the baseline hybrid model (Lexicon + Decision tree) by [9] as well as LSTM model. In Table I and Fig. 3, the BERT LLM model is compared to other currently existing approaches across key metrics: accuracy, precision, recall and F-score [22]. The findings indicate that the BERT LLM model surpasses the hybrid methods in performance, and that the LSTM model yields the least favorable results.

TABLE I. COMPARATIVE OUTPUT OF THE BERT LLM MODEL WITH OTHER METHODS

Methods	Accuracy	Precision	Recall	F-Score
Baseline Model (Lexicon + Decision Tree)	66.14%	98.86%	40.65%	57.62%
LSTM	81.30%	94.00%	71.80%	81.30%
BERT	91.00%	90.50%	81.30%	91.00%

Table I compares the performance of the BERT model with the Baseline Model (Lexicon + Decision Tree) and LSTM in terms of accuracy, precision, recall, and F-score. The BERT model outperforms both alternatives, achieving the highest accuracy (91.00%) and F-score (91.00%), demonstrating its superior ability to understand and classify threat effectively. While the Baseline Model shows high precision (98.86%), its low recall (40.65%) leads to a significantly lower F-score (57.62%), indicating limited generalizability. The LSTM model balances performance better, achieving 81.30% accuracy and F-score with substantial improvements in recall (71.80%) over the Baseline. However, BERT surpasses LSTM in all metrics, particularly in accuracy and F-score, highlighting its robustness in capturing contextual dependencies and delivering precise and balanced predictions.

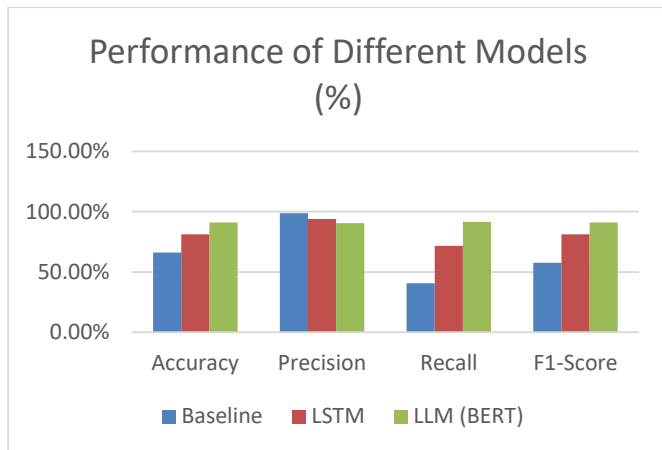


Fig. 3. Comparative performance of BERT LLM model and other methods.

V. DISCUSSION

A. Units Area Under Precision Recall (AUC-PR)

In Fig. 4, an AUC-PR of 0.87 indicates that there is high precision being recalled at different thresholds, suggesting that

the model performs well in separating positive classes from negative classes. The curve demonstrates a high area under the curve (AUC-PR = 0.87), which reflects BERT's ability to maintain a strong balance between precision (90.50%) and recall (81.30%). This high value indicates that BERT effectively minimizes false positives while accurately capturing true positives, even in scenarios with imbalanced datasets. The gradual decline in precision with increasing recall emphasizes the model's robustness in handling trade-offs between these metrics. The shaded region under the curve quantifies the AUC-PR, underscoring the superior contextual understanding and classification efficiency of the BERT model compared to the Baseline and LSTM methods. This result further highlights BERT's utility in tasks requiring precise and consistent predictions.

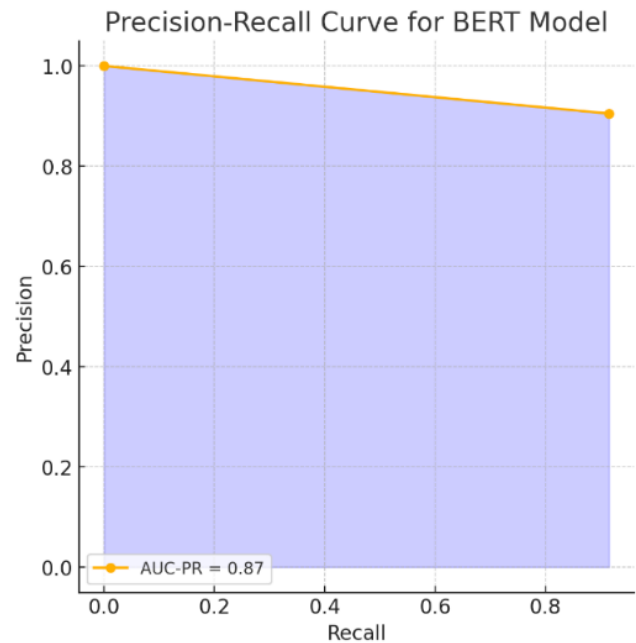


Fig. 4. Precision-recall curve of BERT-LLM model.

B. Training and Validation Loss Curve

The curves in Fig. 5 show the training and validation loss of the model. Validation Loss and Accuracy are calculated on a separate validation dataset and serve as indicators of how well the model generalizes unseen data [15]. The graph shown shows the train loss (red line) and validation loss (green line) over 100 epochs, which illustrates the model learning process.

The red line, which represents losses during training, decreases consistently, signifying that the model is getting better at understanding the patterns and characteristics present in the training data. This continued decline also indicates that the model is successfully reducing prediction errors on the training data, which is a sign that the model is learning well and becoming more efficient at performing predictions.

In addition, validation loss (green line) also shows a steady decline throughout the exercise. This decrease which is almost parallel to the train loss shows that the model not only learns well on the training data but also has good generalization capabilities on the validation data, which has never been seen

during training. The fact that these two losses are almost parallel indicates that the model does not suffer from overfitting, where it manages to avoid overlearning on training data alone, instead works well on the new data being tested.

Both lines show a good downward trend, and there is no significant difference between training loss and confirmation loss. This suggests that the model learns well without relying too much on training data alone. These results show that the model can be effectively used to make accurate predictions on new data. The model shows good generalizations, where it not only gives good results on the training data, but also on the validation data, which is important to ensure that the model does not fit too well with the training data alone.

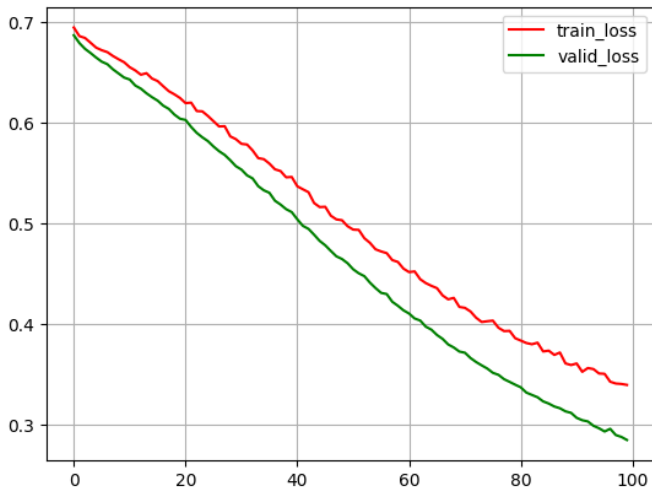


Fig. 5. BERT LLM model loss curves.

C. Future Environment and Infrastructure to Support Large Language Models

This research was conducted in a low-performance environment, utilizing a labeled dataset containing only 250 texts. Given the high computational demands of large language models, it is recommended to leverage high-performance cloud infrastructure in the future.

The success of future Large Language Models approaches relies on a robust cloud infrastructure and stringent security measures to safeguard sensitive data. Consequently, Transformer-based model training, equipped with advanced optimizers, must be performed within high-performance cloud infrastructure. Ensuring the security of the cloud infrastructure, including the underlying bare-metal, firmware, and software technologies, is critical to maintain the overall integrity and resilience of the training environment [23]. A thorough understanding of cloud computing security implications is essential to safeguard data and systems and to ensure the accuracy and reliability of the training data [24], [25].

Cloud security challenges, such as access control issues and the integration of blockchain technologies, including decentralized access control frameworks, must be addressed [26]. The recent adoption of advanced technologies like blockchain for cloud access control through smart contracts also necessitates careful consideration [27]. Additionally,

human factors, such as the acceptance of cloud computing, should be considered to mitigate security risks and enhance the delivery of training and predictive analyses [28], [29]. Additionally, consideration should be given to the human perspective, including the acceptance and trust in cloud computing, to mitigate security risk factors that could affect the delivery of training and predictive analysis of threat [30].

VI. CONCLUSION

This research study shows that combining the transformer-based BERT model significantly enhances deep learning models' abilities to predict political threats. This model is not only capable of reshaping the political security threat prediction landscape but can also support researchers' future studies within the national security field. The synergy between the sentiment analysis, word embedding and BERT networks offers enhanced accuracy and robustness in national security scenarios by adaptively adjusting learning rates, while also capturing long dependencies that are essential in detecting evolving threats. To summarize, the transformer-based BERT model introduces new and effective ways to enhance the accuracy, efficiency, and versatility of political threat prediction, making it a significant advancement in the domain of national security.

ACKNOWLEDGMENT

The authors fully acknowledge Ministry of Higher Education Malaysia (MOHE) and National Defence University of Malaysia (NDUM) which makes this important research viable and effective.

REFERENCES

- [1] P. Datta, N. Lodinger, A. S. Namin, and K. S. Jones, "Predicting Consequences of Cyber-Attacks," in Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 2073–2078. doi: 10.1109/BigData50022.2020.9377825.
- [2] S. Hatab, "Threat perception and democratic support in Post-arab spring Egypt," *Comp Polit*, vol. 53, no. 1, pp. 69–91, 2020, doi: 10.5129/001041520X15822914282706.
- [3] G. Wolfsfeld, E. Segev, and T. Sheaffer, "Social Media and the Arab Spring: Politics Comes First," *International Journal of Press/Politics*, vol. 18, no. 2, pp. 115–137, Apr. 2013, doi: 10.1177/1940161212471716.
- [4] H. Strömfelt, Y. Zhang, and B. W. Schuller, "Emotion-Augmented Machine Learning: Overview of an Emerging Domain," 2017.
- [5] N. A. M. Razali et al., "Opinion mining for national security: techniques, domain applications, challenges and research opportunities," *J Big Data*, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00536-5.
- [6] M. Suhairi et al., "Social Media Sentiment Analysis and Opinion Mining in Public Security: Taxonomy, Trend Analysis, Issues and Future Directions."
- [7] G. Wolfsfeld, E. Segev, and T. Sheaffer, "Social Media and the Arab Spring: Politics Comes First," *International Journal of Press/Politics*, vol. 18, no. 2, pp. 115–137, Apr. 2013, doi: 10.1177/1940161212471716.
- [8] A. Arora, A. Arora, and J. McIntyre, "Developing Chatbots for Cyber Security: Assessing Threats through Sentiment Analysis on Social Media," *Sustainability (Switzerland)*, vol. 15, no. 17, Sep. 2023, doi: 10.3390/su151713178.
- [9] N. A. M. Razali et al., "Political Security Threat Prediction Framework Using Hybrid Lexicon-Based Approach and Machine Learning Technique," *IEEE Access*, vol. 11, pp. 17151–17164, 2023, doi: 10.1109/ACCESS.2023.3246162.

- [10] L. Safra Zaabar, M. R. Yaakub, M. Iqbal, and A. Latiffi, "Combination of Lexicon Based and Machine Learning Techniques in the Development of Political Tweet Sentiment Analysis Model."
- [11] M. Ridzwan Yaakub, M. Iqbal Abu Latiffi, and L. Safra Zaabar, "A Review on Sentiment Analysis Techniques and Applications," in IOP Conference Series: Materials Science and Engineering, Institute of Physics Publishing, Aug. 2019. doi: 10.1088/1757-899X/551/1/012070.
- [12] K. Ohtomo, R. Harakawa, M. Iisaka, and M. Iwahashi, "AM-Bi-LSTM: Adaptive multi-modal Bi-LSTM for sequential recommendation," IEEE Access, 2024, doi: 10.1109/ACCESS.2024.3355548.
- [13] F. Sufi, "A global cyber-threat intelligence system with artificial intelligence and convolutional neural network," Decision Analytics Journal, vol. 9, Dec. 2023, doi: 10.1016/j.dajour.2023.100364.
- [14] Y. Touzani and K. Douzi, "An LSTM and GRU based trading strategy adapted to the Moroccan market," J Big Data, vol. 8, no. 1, Dec. 2021, doi: 10.1186/s40537-021-00512-z.
- [15] A. Saeed and E. Al Solami, "Fake News Detection Using Machine Learning and Deep Learning Methods," Computers, Materials and Continua, vol. 77, pp. 2079–2096, 2023, doi: 10.32604/cmc.2023.030551.
- [16] N. J. Prottasha et al., "Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning," Sensors, vol. 22, no. 11, Jun. 2022, doi: 10.3390/s22114157.
- [17] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," Nov. 01, 2021, Springer. doi: 10.1007/s42979-021-00815-1.
- [18] D. Wollebæk, R. Karlsen, K. Steen-Johnsen, and B. Enjolras, "Anger, Fear, and Echo Chambers: The Emotional Basis for Online Behavior," Social Media and Society, vol. 5, no. 2, 2019, doi: 10.1177/2056305119829859.
- [19] J. Kruger, B. Chen, S. Heitfeld, L. Witbart, C. Bruce, and D. L. Pitts, "Attitudes, Motivators, and Barriers to Emergency Preparedness Using the 2016 Styles Survey," Health Promot Pract, vol. 21, no. 3, pp. 448–456, May 2020, doi: 10.1177/1524839918794940.
- [20] T. G. Coan, J. L. Merolla, E. J. Zechmeister, and D. Zizumbo-Colunga, "Emotional Responses Shape the Substance of Information Seeking under Conditions of Threat," Polit Res Q, vol. 74, no. 4, pp. 941–954, Dec. 2021, doi: 10.1177/1065912920949320.
- [21] Md. N. Y. Ali, Md. G. Sarowar, Md. L. Rahman, J. Chaki, N. Dey, and J. M. R. S. Tavares, "Adam Deep Learning With SOM for Human Sentiment Classification," International Journal of Ambient Computing and Intelligence, vol. 10, no. 3, pp. 92–116, Jul. 2019, doi: 10.4018/IJACI.2019070106.
- [22] S. Nawaz, "A Comparative Study of Machine Learning Algorithms for Sentiment Analysis," 1999. [Online]. Available: www.ijrpr.com
- [23] Proceedings, 2020 16th IEEE International Colloquium on Signal Processing & its Application (CSPA 2020) : 28th-29th February 2020 : conference venue, Hotel Langkawi, Lot 1852 Jalan Penarak, Kuah 07000 Langkawi, Kedah, Malaysia. IEEE, 2020.
- [24] M. Noorafiza, H. Maeda, R. Uda, T. Kinoshita, and M. Shiratori, "Vulnerability analysis using network timestamps in full virtualization virtual machine," in ICISSP 2015 - 1st International Conference on Information Systems Security and Privacy, Proceedings, SciTePress, 2015, pp. 83–89. doi: 10.5220/0005242000830089.
- [25] M. Noorafiza, H. Maeda, T. Kinoshita, and R. Uda, "Virtual machine remote detection method using network timestamp in cloud computing," in 2013 8th International Conference for Internet Technology and Secured Transactions, ICITST 2013, IEEE Computer Society, 2013, pp. 375–380. doi: 10.1109/ICITST.2013.6750225.
- [26] W. N. Wan Muhamad et al., "Enhance multi-factor authentication model for intelligence community access to critical surveillance data," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer, 2019, pp. 560–569. doi: 10.1007/978-3-030-34032-2_49.
- [27] N. M. Noor, N. A. M. Razali, S. N. S. A. Sham, K. K. Ishak, M. Wook, and N. A. Hasbullah, "Decentralised Access Control Framework using Blockchain: Smart Farming Case," International Journal of Advanced Computer Science and Applications, vol. 14, no. 5, pp. 566–579, 2023, doi: 10.14569/IJACSA.2023.0140560.
- [28] M. R. A. Bakar, N. A. M. Razali, M. Wook, M. N. Ismail, and T. M. T. Sembok, "Exploring and Developing an Industrial Automation Acceptance Model in the Manufacturing Sector Towards Adoption of Industry4.0," Manufacturing Technology, vol. 21, no. 4, pp. 434–446, 2021, doi: 10.21062/mft.2021.055.
- [29] M. R. Abu Bakar, N. A. Mat Razali, M. Wook, M. N. Ismail, and T. M. Tengku Sembok, "The Mediating Role of Cloud Computing and Moderating Influence of Digital Organizational Culture Towards Enhancing SMEs Performance," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer Science and Business Media Deutschland GmbH, 2021, pp. 447–458. doi: 10.1007/978-3-030-90235-3_39.
- [30] K. Khalil Ishak et al., "Smart Cities' Cybersecurity and IoT: Challenges and Future Research Directions."

Text-to-Image Generation Method Based on Object Enhancement and Attention Maps

Yongsen Huang, Xiaodong Cai, Yuefan An

School of Information and Communication, Guilin University of Electronic Technology, Guilin, China

Abstract—In the task of text-to-image generation, common issues such as missing objects in the generated images often arise due to the model's insufficient learning of multi-object category information and the lack of consistency between the text prompts and the generated image contents. To address these challenges, this paper proposes a novel text-to-image generation approach based on object enhancement and attention maps. First, a new object enhancement strategy is introduced to improve the model's capacity to capture object-level features. The core idea is to generate difficult samples by processing the object mask maps of tokens, followed by dynamic weighting of the attention map using latent image embeddings. Second, to enhance the consistency between the text prompts and the generated image contents, we enforce similarity constraints between the cross-attention maps and the attention-weighted mask feature maps, penalizing inconsistencies through a loss function. Experimental results demonstrate that the Stable Diffusion v1.4 model, optimized using the proposed method, achieves significant improvements on the COCO instance dataset and the ADE20K instance dataset. Specifically, the MG metrics are improved by an average of 12.36% and 6.55%, respectively, compared to state-of-the-art models. Furthermore, the FID metrics show a 0.84% improvement over the state-of-the-art model on the COCO instance validation set.

Keywords—Multi-object category; text-to-image generation; object enhancement; attention maps

I. INTRODUCTION

The rapid advancement of artificial intelligence has propelled text-to-image generation into the forefront of research at the intersection of computer vision and natural language processing. This emerging field aims to automatically generate visual content that aligns with natural language descriptions. Beyond its theoretical significance, this technology holds vast potential for application in diverse areas, including virtual reality, game design, artistic creation, and human-computer interaction. However, a major challenge persists: efficiently translating textual semantics into high-quality images while ensuring a high degree of consistency between the generated images and their corresponding textual descriptions.

Recent advances in deep learning have opened new avenues for text-to-image generation tasks. Generative Adversarial Networks (GANs) [1], as an early foundational technology, enabled the initial mapping from text to image via adversarial training between generators and discriminators. However, the quality and semantic alignment of the generated images still require significant improvement. The subsequent development of autoregressive and diffusion models has invigorated the field. Autoregressive models generate high-quality, semantically

consistent images through pixel-by-pixel synthesis but suffer from slow training and inference speeds. In contrast, diffusion models generate images through a process of iterative denoising, yielding not only high-quality images but also enhanced diversity, positioning them as the prevailing approach in contemporary research.

The objective of diffusion models can be summarized as reversing the gradual degradation process of data, which consists of a forward process that follows a Markov chain and a reverse diffusion process. In the forward process, noise is gradually added to the original data, causing it to degrade into nearly isotropic Gaussian noise, thereby corrupting the original data. In contrast, the reverse diffusion process utilizes a neural network to learn how to recover the original data from Gaussian noise. It is important to note that the input and output dimensions of the reverse diffusion process must remain consistent.

Although recent text-to-image diffusion models [2][3][4][5][6][7][8][9] have achieved notable progress in generating images with increasing levels of quality, resolution, realism, and diversity, a significant challenge remains in maintaining consistency between text prompts and the content of the generated images.

Several studies have addressed the challenge of generating images containing multiple objects. In 2022, Robin Rombach et al. [10] introduced Stable Diffusion, a high-resolution image synthesis method based on Latent Diffusion Models (LDMs), designed to overcome the computational inefficiencies of traditional diffusion models in high-resolution image generation. The model achieves diverse high-resolution image generation by integrating components such as variational autoencoders (VAE), conditional text encoders, and U-Net. In the same year, Liu et al. [11] proposed Composable Diffusion, a method leveraging multiple diffusion models to generate complex scenes by separately generating different objects, each handled by a specialized model. That same year, Liew et al. [12] introduced MagicMix, a technique that uses pre-trained, text-conditioned diffusion models to blend two distinct semantic concepts into a single image. The process begins by generating a rough semantic layout, followed by content matching the text description, and concludes by merging the semantic information of the two objects. In 2023, Chefer et al. [13] developed Attend-and-Excite, a method aimed at enhancing the semantic fidelity of text-to-image diffusion models. By optimizing the cross-attention mechanism, this approach ensures that the generated images better reflect the input text prompts. Directed Diffusion [14], also in 2023, introduced a novel approach to controlling the positioning of multiple elements in the image by manipulating attention maps at the word and word-position

levels, improving the model's focus on the relevant areas of the image. In the same year, Zirui Wang et al. [15] presented TokenCompose, a method that incorporates token-level supervision to improve performance in multi-object composition tasks and enhances the photorealism of generated images. More recently, in 2024, Tobias Lingenberg et al. [16] proposed DIAGen, an image enhancement method for few-shot learning scenarios. By combining generative models with text prompts, the method effectively increases the diversity of generated images.

However, most prior research has been limited to simple augmentation techniques, such as flipping, rotation, or basic enhancement operations on image data. These methods often fail to adequately capture the object features, leading to issues such as missing objects in the generated images.

Moreover, while much of the previous research has focused on the spatial layout of the cross-attention maps between text prompts and generated image contents during image generation, it has often overlooked the importance of enhancing the understanding of the spatial layout of the object cross-attention maps during the model's training phase.

To address the aforementioned issues, inspired by the literature [10] [15], this paper proposes a novel text-to-image generation method based on object enhancement and attention maps (TI-OEAM). By constructing difficult samples, incorporating a dynamic residual gating mechanism, and applying an attention maps guidance approach, this paper optimizes the model and significantly enhances the consistency between the text prompts and the generated images, leading to a substantial improvement in image quality.

The subsequent sections of this paper will provide a detailed exploration of this research. The TI-OEAM model section will describe the proposed method, focusing on the design and implementation of the object enhancement strategy, as well as how attention map optimization is employed to improve the consistency between text prompts and the generated image content. The experimental section will outline a series of experiments conducted to evaluate the effectiveness and performance of the proposed approach, including comparative studies with existing methods and ablation experiments. Finally, the conclusion section will summarize the key findings and contributions of this work, discuss its limitations, and propose directions for future research.

II. THE PROPOSED FRAMEWORK

A. An Overview of the Proposed Framework

As shown in Fig. 1, in the object enhancement module, the similarity between the object mask map and all other object mask maps in the image is first evaluated through a similarity calculation. Object mask maps with similarity values exceeding a predefined threshold are then filtered out. Gaussian noise is then applied to these selected mask maps to introduce random perturbations, generating difficult samples. Subsequently, latent noisy image embeddings with learnable parameters are utilized as dynamic residual weighting terms in the denoising U-Net, which are integrated into the model's training process. Building on this, attention map optimization is performed by imposing similarity constraints between the cross-attention map and the

attention-weighted masked feature map. Finally, the model is optimized using a loss function.

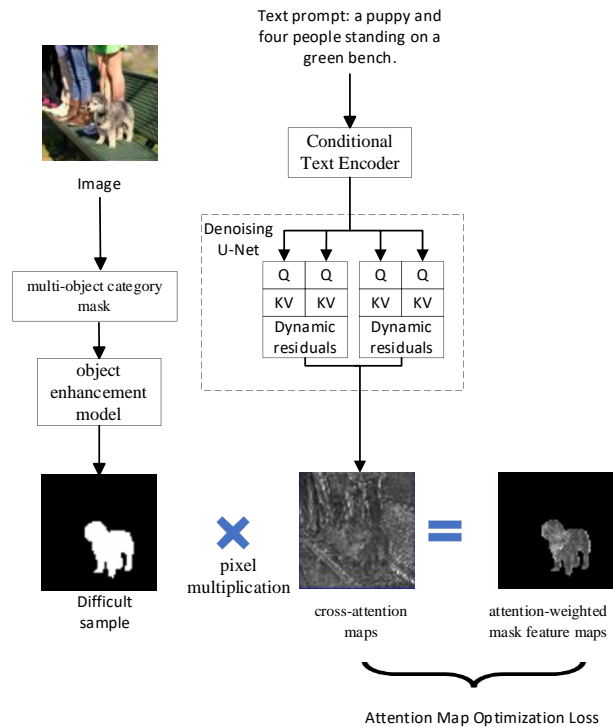


Fig. 1. TI-OEAM overall framework.

B. Object Enhancement Strategies using Masking and Residuals

To fully capture the diverse object feature information in images, a novel object enhancement strategy is proposed, consisting of two key steps. The first step involves constructing difficult samples, while the second step introduces a dynamic residual gating mechanism.

1) *Difficult sample generation*: In an image, multiple objects often exhibit similar visual characteristics, particularly when they belong to the same category or share similar appearances. Such similarities may hinder the model's ability to accurately distinguish between objects, potentially resulting in missed or incorrect generation of specific objects during image generation. To address this issue, we propose calculating the similarity between object masks and introducing noisy interference to highly similar masks. This approach forces the model to focus more on the subtle differences between similar objects. The interference thus encourages the model to learn how to differentiate and generate diverse features of similar objects, ultimately enhancing its capacity to generate objects across multiple categories more effectively.

Specifically, the first step involves designing an object enhancement module that calculates the pixel-level feature similarity between the object mask map in the image and all other object mask maps. Object mask maps with similarity values exceeding a predefined threshold are then filtered out. Subsequently, Gaussian random interference is applied to these mask maps to generate difficult samples.

The specific calculation process is as follows:

$$f = \frac{1}{G-1} \sum_k^{G-1} \cos(\mathbf{M}, \mathbf{M}_k) \quad (1)$$

Where \mathbf{M} denotes the object mask map in the image, G denotes the number of object mask maps the image contains, and f indicates the average pixel-level feature similarity between the object mask map and all other object mask maps in the image.

Then Gaussian random interference is applied to the object mask maps above the set threshold in the following process:

$$\begin{cases} \mathbf{M}' = \mathbf{M} + \Delta', f > \varphi \\ \mathbf{M}, f \leq \varphi \end{cases} \quad (2)$$

Where the threshold φ is set to 0.8, the added noise vector Δ' obeys $\|\Delta\|_2 = \delta$, and δ is a small constant. \mathbf{M}' is called the difficult sample, where Δ process is as follows:

$$\Delta = \omega, \omega \in \mathbb{R}^d \square U(0, 1e-3) \quad (3)$$

The incorporation of difficult samples encourages the model to focus on object mask maps that exhibit high similarity and are challenging to distinguish from other objects in the image. This strategy enables the model to more effectively learn the distinctive features of each object.

Dynamic Residual Gating: In this study, the definition proposed by Zirui Wang et al. [15] is adopted. For a given image $x \in \mathbb{R}^{H \times W \times 3}$ in RGB space, it is first processed by the encoder part of the VAE to obtain its latent image embedding $z_0 = \mathcal{E}(x_0)$. Subsequently, based on this potential image embedding, Gaussian random noise is injected with time t to obtain the potential noisy image embedding z_t containing the noise. Additionally, a conditional text encoder is employed to convert the text prompts y into an embedding $\tau_\theta(y)$ with the neural network parameter θ . Let $\mathbf{K} \in \mathbb{R}^{H \times L_{\tau_\theta(y)} \times d_k}$ represent the embedding of the text prompts corresponding to the token, where $L_{\tau_\theta(y)}$ denotes the length of the text prompts embedding $\tau_\theta(y)$. Let $\mathbf{Q} \in \mathbb{R}^{H \times L_{z_t} \times d_k}$ represent the latent noisy image embedding, with L_{z_t} indicating the length of the latent noisy image embedding. d_k denotes the dimension of \mathbf{K} . The cross-attention map of text prompts and images is computed as follows:

$$\mathbf{Q}^{(h)} = \mathbf{W}_Q^{(h)} \cdot \varphi(z_t) \quad (4)$$

$$\mathbf{K}^{(h)} = \mathbf{W}_K^{(h)} \cdot \tau_\theta(y) \quad (5)$$

$$\mathbf{V}^{(h)} = \mathbf{W}_V^{(h)} \cdot \tau_\theta(y) \quad (6)$$

$$\mathbf{A} = \frac{1}{H} \sum_h^H \text{soft max} \left(\frac{\mathbf{Q}^{(h)} (\mathbf{K}^{(h)})^T}{\sqrt{d_k}} \right) \cdot \mathbf{V} \quad (7)$$

Where $h \in \{1, \dots, H\}$ denotes each head in the multi-head cross-attention. $\mathbf{W}_Q^{(h)}$, $\mathbf{W}_K^{(h)}$ and $\mathbf{W}_V^{(h)}$ are the learnable projection matrices in the cross-attention layer in each head. $\varphi(z_t)$ denotes the function that flattens the 2D latent image embedding to 1D.

This study identifies a limitation in most current approaches, which directly use the output of the pre-trained U-Net model as the input for the VAE-generated images. This practice often results in insufficient learning of multi-category object information. To address this, the second step involves designing a dynamic residual gating mechanism.

In this paper, learnable parameters are employed to adjust the weights of the residual connections, thereby allowing the model to flexibly control the flow of information based on varying contextual conditions. This dynamic adjustment mechanism enhances the model's nonlinear learning capability, enabling more accurate representation of image features. By adopting this approach, the model can optimize the learning process during image generation in accordance with the specific characteristics of the image, thus mitigating the issue of insufficient feature capture that may arise when the weights of residual connections are fixed.

Specifically, the latent image embedding is used to preserve image features, thereby allowing the model to concentrate more effectively on these features. The residual cross-attention computation process is as follows:

$$\mathbf{A}' = \mathbf{A} + \alpha * z_t \quad (8)$$

Where α is the learnable parameter. The dynamic residual gating mechanism enhances and refines the cross-attention maps by dynamically adjusting the weights of the residual connections. This mechanism improves the model's ability to capture features, enabling it to more accurately learn the features of the data.

C. Optimization of Attention Map

Robin Rombach et al. [10] and Zirui Wang et al. [15] learned noisy features by using a denoising function, L_{DM} , which takes into account the lack of direct optimization correlation between tokens and image contents. To compensate for this deficiency, they introduced a token-level attention loss, L_t , which monitors the activation region of cross-attention. In addition, to prevent attention from being overly focused on certain sub-regions of the target region, they also propose a pixel-level attention loss L_p .

Inspired by the work of Robin Rombach, Zirui Wang, and others, we aim to better manage the spatial arrangement and interactions of multiple objects during the image generation process, ensuring that the generated image accurately includes

all specified categories and objects. In this paper, we propose a novel attention optimization method that constrains the similarity between the cross-attention map and the attention-weighted masked feature map. The objective is to improve the model's sensitivity to information within specific attention regions, without compromising its ability to capture global

context. Let i denote the noun token of a text prompt. Let A'_i represent the cross-attention map between the latent noisy image embeddings and the embedding of a token.

First, the attention-weighted mask feature map is computed as follows:

$$A'_{(i,u)} \square M'_{(i,u)} = A'_{M'(i,u)} \quad (9)$$

Where \square denotes the pixel multiplication. u is the spatial location of the cross-attention map. $M'_{(i,u)}$ denotes the object mask map of text token i at spatial position u . Let $A'_{(i,u)}$ denote the cross-attention map formed by the latent noisy image embedding of the i -th token at the spatial location u of $A'_i \in R^{L_{z_i}}$.

Subsequently, the cosine similarity function is calculated in accordance with the following procedure:

$$S(A'_M, A') = \frac{A'_{M'(i,u)} \cdot A'_{(i,u)}}{|A'_{M'(i,u)}| \cdot |A'_{(i,u)}|} = \sum_i^{L_{z_i}} \sum_u^{L_i} \frac{A'_{M'(i,u)} A'_{(i,u)}}{\sqrt{A'_{M'(i,u)}^2} \sqrt{A'_{(i,u)}^2}} \quad (10)$$

Where $S(A'_M, A')$ denotes the similarity between the cross-attention map and the attention-weighted mask feature map.

The two are then brought into closer alignment in the pixel domain using a loss function, which is calculated as follows:

$$L_s = 1 - S(A'_M, A') \quad (11)$$

Cosine similarity function $S(A'_M, A')$ measures how close the cross attention map is to the attention-weighted masked feature map in feature space. The loss function L_s effectively balances the learning of both local and global information, overcoming the limitations of traditional methods that tend to over-focus on the target region or neglect other areas of the image. This balance improves the alignment between text prompts and generated image content, thereby enhancing both image consistency and overall quality.

D. Loss Function

Finally, L_s and L_{DM}, L_t and L_p jointly trained and computed as follows:

$$L = L_{DM} + \sum_d^D (\gamma L_s + \eta L_t + \psi L_p) \quad (12)$$

Where η , γ and ψ are the scaling factors. In this paper, we set γ as 1e-3, η as 1e-3, ψ as 5e-5. The value of D represents the number of training layers.

III. EXPERIMENT

A. Datasets

In order to evaluate the effectiveness of the proposed model in the text-to-image generation task, this paper utilizes the COCO dataset [15] for training experiments. This dataset consists of 4,526 image-caption pairs and their corresponding binary mask maps. The choice of COCO is motivated by the relatively low ambiguity in its visual language and the rich diversity of object categories represented in each image, which effectively supports the task of learning and generating multi-category objects. To further assess the model's performance in multi-category instance combination, we conduct experiments on the COCO instance dataset [17], which includes 80 categories, and the ADE20K instance dataset [18][19], which includes 100 categories. Additionally, to evaluate the distributional differences between the images generated by the model and real images, we randomly sampled 10,000 image-caption pairs from the COCO instance validation set (C) and 1,000 image-caption pairs from the Flickr30K instance validation set (F) [20] for comparative analysis.

B. Evaluation Metrics

In this paper, we use the MULTIGEN [15] metric and the FID [21] metric to assess the ability of the model in combining instances of multiple categories and to analyse the distributional differences between the images generated by the model and the real images.

MULTIGEN is a challenging metric used to evaluate multi-category instance combinations. Specifically, given a set of N distinct instance categories, five categories (e.g., A, B, C, D, and E) are randomly selected and formatted into a sentence (e.g., "A photo of A, B, C, D, and E"). This sentence serves as the conditional input for the text-to-image diffusion model to generate the corresponding image. Subsequently, a robust open-vocabulary detector [15] is employed to assess whether the specified categories are accurately represented in the generated image.

Specifically, for each dataset, 1,000 text prompts were generated by randomly sampling 1,000 instances from each of the 80 COCO categories and 100 ADE20K categories, which were then used as inputs for the multi-category instance combinations. For each text prompt, 10 rounds of image generation were performed, resulting in a total of 10×1000 images for each dataset's category combination. Each generated image was subsequently analyzed using a detector to count the number of category instances present. Based on these detection results, the MG2 to MG5 metrics were computed for each round.

The mean and standard deviation (denoted in parentheses) of the MG2-5 success rates across the 10 rounds are presented in Table I.

TABLE I. COMPARISON OF EXPERIMENTAL RESULT OF VARIOUS MODELS

Method	Multi-category Instance Composition↑								Photorealism↓	
	COCO INSTANCES				ADE20K INSTANCES				FID (C)	FID (F)
	MG2	MG3	MG4	MG5	MG2	MG3	MG4	MG5		
SD	90.72 _{1.33}	50.74 _{0.89}	11.68 _{0.45}	0.88 _{0.21}	89.81 _{0.40}	53.96 _{1.14}	16.52 _{1.13}	1.89 _{0.34}	20.88	71.46
Composable	63.33 _{0.59}	21.87 _{1.01}	3.25 _{0.45}	0.23 _{0.18}	69.61 _{0.99}	29.96 _{0.84}	6.89 _{0.38}	0.73 _{0.22}	-	75.57
Layout	93.22 _{0.69}	60.15 _{1.58}	19.49 _{0.88}	2.27 _{0.44}	96.05 _{0.34}	67.83 _{0.90}	21.93 _{1.34}	2.35 _{0.41}	-	74.00
Structured	90.40 _{1.06}	48.64 _{1.32}	10.71 _{0.92}	0.68 _{0.25}	89.25 _{0.72}	53.05 _{1.20}	15.76 _{0.86}	1.74 _{0.49}	21.13	71.68
Attn-Exct	93.64 _{0.76}	65.10 _{1.24}	28.01 _{0.90}	6.01_{0.61}	91.74 _{0.49}	62.51 _{0.94}	26.12 _{0.78}	5.89 _{0.40}	-	71.68
TokenCompose	98.08 _{0.40}	76.16 _{1.04}	28.81 _{0.95}	3.28 _{0.48}	97.75 _{0.34}	76.93 _{1.09}	33.92 _{1.47}	6.21 _{0.62}	20.19	71.13
TI-OEAM	98.54_{0.19}	79.43_{0.91}	32.55_{0.96}	4.32 _{0.36}	97.91_{0.18}	79.39_{0.85}	37.13_{2.39}	7.04_{0.51}	20.02	71.94

The FID metric is used to assess the distributional disparity between two datasets: 10,000 image-caption pairs from the COCO instance validation set (C) and 1,000 image-caption pairs from the Flickr30K instance validation set (F), in comparison with the model-generated images.

C. Experimental Settings

The experimental setup is as follows: The operating system is Ubuntu 18.04.5 LTS; the hardware configuration includes an NVIDIA 3090 GPU; the deep learning framework used is PyTorch; and the programming language is Python.

The primary experiments in this paper are conducted using the Stable Diffusion v1.4 [22] model and the TokenCompose model. Stable Diffusion is a widely used text-to-image diffusion model for high-quality generation. AdamW is employed as the optimizer [23], with a global learning rate of 5e-6 and a total of 2400 steps. The image resolution is set to 512. Training was performed on a single GPU using a single batch and four gradient accumulation steps across the entire U-Net.

D. Experimental Results and Analysis

To validate the effectiveness of the TI-OEAM model, this study compares its performance with several representative text-to-image generation methods. These include Stable Diffusion (SD) [22], which addresses the high computational cost of traditional diffusion models in high-resolution image generation; Composable Diffusion [11], which generates complex scenes by combining multiple diffusion models; Layout [24], which manipulates the cross-attention layer in diffusion models to achieve precise spatial layout control; Structured [25], which enhances composability and attribute binding in text-to-image tasks by integrating linguistic structures with cross-attention layers; Attend-and-Excite [13], which improves semantic fidelity in text-to-image generation; and TokenCompose [15], which enhances text-to-image models through token-level supervision. The experimental results are presented in Table I, where MG, C, and F represent the MULTIGEN metrics, the COCO instance validation set, and the Flickr30k instance validation set, respectively.

The experimental results show that the baseline method, TokenCompose, achieves an average improvement of 12.8% over the Attend-and-Excite model across all the improved MG metrics. Since the FID metrics are reliable only when comparing 10,000 images, a comparison of the FID metrics on a dataset of

10,000 image-caption pairs sampled from the COCO validation set (C) shows a 3.3% reduction in the metrics of TokenCompose compared to the Attend-and-Excite model. The illustrative analysis results from the TokenCompose model further highlight the 12.8% improvement in MG metrics and the 3.3% reduction in FID metrics, representing substantial advancements in performance.

As shown in Table I, the proposed TI-OEAM model significantly outperforms all baseline methods across most evaluation metrics for these datasets. Compared to the state-of-the-art performance of current mainstream model, TokenCompose, on the COCO instance dataset, TI-OEAM achieves an average improvement of 12.36% on the MG2, MG3, MG4, and MG5 metrics. On the ADE20K instance dataset, TI-OEAM improves by an average of 6.55% on the same metrics. Additionally, TI-OEAM demonstrates a 0.84% reduction in the FID score compared to the 10,000 image-caption pairs in the COCO instance validation set. This performance enhancement is attributed to the model's efficacy, particularly its ability to learn information across multiple object categories through difficult sample construction and dynamic residual gating methods. Furthermore, TI-OEAM incorporates similarity constraints between the cross-attention map and the attention-weighted mask feature map, further improving the consistency between text prompts and the content of the generated images, resulting in significant gains across all performance metrics.

In comparison to the 1,000 image-caption pairs from the Flickr30K instance validation set, the FID metric for the images generated by the TI-OEAM model, as shown in Table I, is 71.94. The discrepancy in these experimental results can be attributed to the model's failure to pass the safety-checker detection during image generation based on the 1,000 captions. This issue led to the generation of a black image, which significantly affected the experimental outcomes.

E. Ablation Study

To validate the impact of the proposed OE and AN components on model performance, this paper conducts MG metrics ablation experiments on the COCO instance dataset and the ADE20K instance dataset, as well as FID metrics ablation experiments on 10,000 image-caption pairs sampled from the COCO instance validation set. Table II presents the experimental results, while Fig. 2 provides effectiveness analysis.

TABLE II. COMPARISON OF ABLATION EXPERIMENTAL RESULTS

Component	COCO INSTANCES			ADE20K INSTANCES			FID (C)
	MG3	MG4	MG5	MG3	MG4	MG5	
TI-OEAM	79.43 _{0,91}	32.55 _{0,96}	4.32 _{0,36}	79.39 _{0,85}	37.13 _{2,39}	7.04 _{0,51}	20.02
OE	77.62 _{1,08}	31.62 _{1,26}	3.93 _{0,46}	77.54 _{0,77}	35.82 _{1,30}	6.44 _{0,63}	19.90
AM	78.48 _{0,96}	31.44 _{1,62}	3.91 _{0,80}	79.18 _{0,79}	37.80 _{0,77}	7.62 _{0,84}	20.25

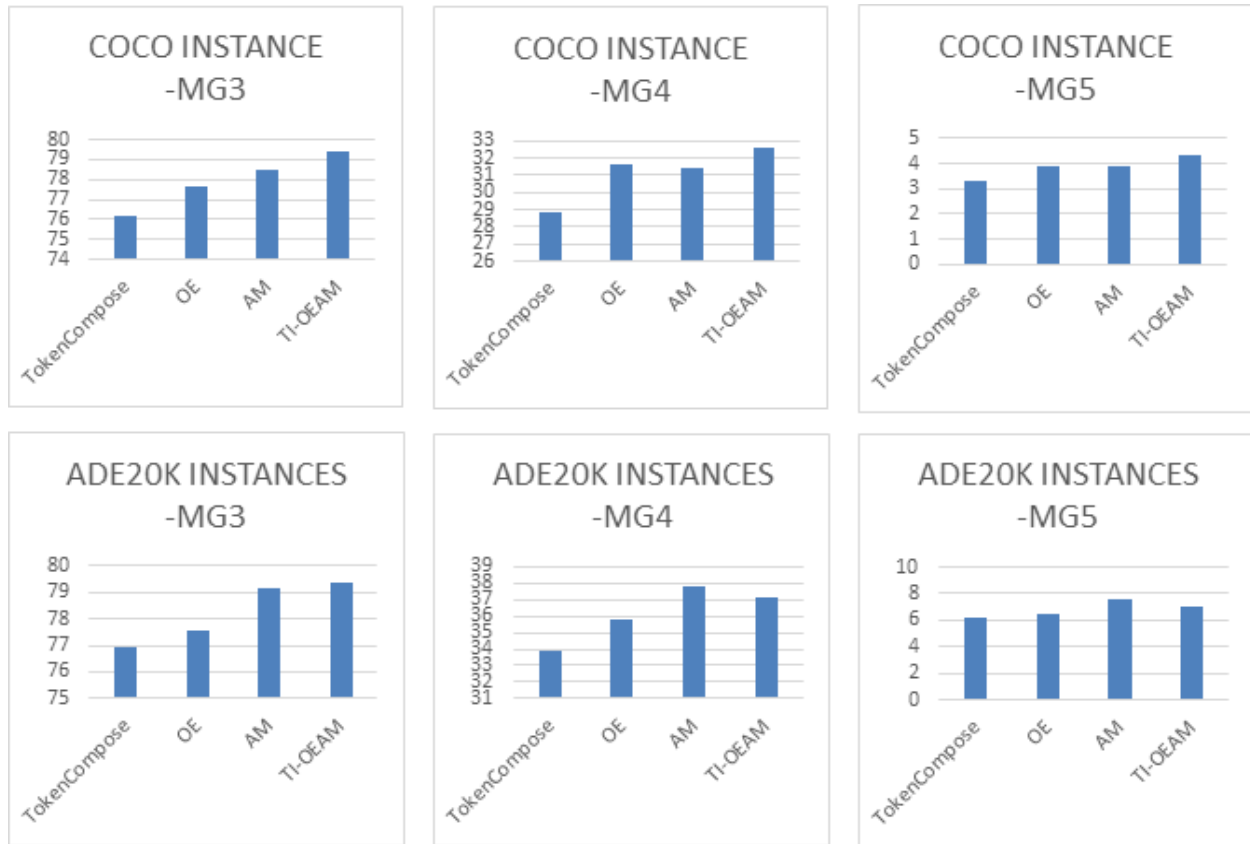


Fig. 2. Effectiveness analysis OE and AM.

Firstly, the OE module was independently validated, and the experimental results are presented in the "OE" section of Table II. Compared to the TokenCompose model, OE achieved improvements of 6.93% in the MG3, MG4, and MG5 metrics, respectively. This enhancement demonstrates that the OE module has made significant progress in extracting object features, thereby further enhancing the model's ability to learn object characteristics within the image.

Secondly, the AM method is validated in isolation, with the results presented in the "AM" section of Table II. Compared to the TokenCompose model, the AM method yields improvements of 11.41% in the MG3, MG4, and MG5 metrics. The AM method demonstrates significant improvements across all metrics, indicating that it enhances the model's understanding of the distribution of objects in images in alignment with text prompts. Furthermore, this method

improves the consistency between text and image content, resulting in images with greater object accuracy.

In conclusion, the OE and AM modules demonstrate significant improvements in the metrics associated with the text-to-image generation task.

F. Visual Presentation Analysis

To facilitate a comprehensive visual comparison between the proposed TI-OEAM model and the benchmark TokenCompose model, a set of six images was generated using identical text prompts by both models. The generated images are presented in Fig. 3 on the following page for direct comparison. In order to ensure the fairness and consistency of the comparison, the initial latent space values, which serve as the starting point for both models, were held constant across all experiments. This approach minimizes potential bias arising from variations in latent representations, thereby allowing for an objective assessment of the models' performance.



Fig. 3. Qualitative comparison between TI-OEAM and baseline.

To rigorously evaluate the quality and alignment of the generated images with the text prompts, this study involved a manual assessment conducted by 11 researchers from diverse fields, including natural language processing, big data, and computer science. The researchers were tasked with evaluating the consistency between the provided text prompts and the corresponding generated images, using a well-defined criterion to ensure objectivity. The results of the evaluation revealed that, out of the total 66 votes cast, 26 votes were in favor of the images generated by the TokenCompose model, while 40 votes were in favor of those generated by the TI-OEAM model. This indicates that the images produced by the TI-OEAM model demonstrated a significantly higher degree of consistency with the text prompts compared to those generated by TokenCompose, highlighting the effectiveness of the proposed model in faithfully translating textual descriptions into visual representations.

IV. CONCLUSION AND FUTURE WORK

This paper proposes a text-to-image generation method based on object enhancement and attention maps. The generation of high-quality images is achieved through the construction of challenging samples, the implementation of a dynamic residual gating mechanism, and the optimization of the model via an attention map guidance approach. These strategies work together to enhance the consistency between text prompts and generated images. Experimental results demonstrate that the

proposed method, which integrates difficult sample design, dynamic residual gating, and attention map optimization, yields more significant improvements than state-of-the-art models in both MG metrics and the consistency of text-image information for text-image generation tasks. Future work will continue to address the consistency issue between text prompts and image contents, with a particular focus on more complex text prompts. The aim is to provide practical solutions for this challenge in the field.

ACKNOWLEDGMENT

This work is partially supported by Guangxi Innovation Driven Development Project (AA20302001).

REFERENCES

- [1] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[J]. Communications of the ACM, 2020, 63(11): 139-144.
- [2] Feng Z, Zhang Z, Yu X, et al. Ernie-vilg 2.0: Improving text-to-image diffusion model with knowledge-enhanced mixture-of-denoising-experts[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 10135-10145.
- [3] Hertz A, Mokady R, Tenenbaum J, et al. Prompt-to-prompt image editing with cross attention control[J]. arXiv preprint arXiv:2208.01626, 2022.
- [4] Ramesh A, Pavlov M, Goh G, et al. Zero-shot text-to-image generation[C]//International conference on machine learning. Pmlr, 2021: 8821-8831.
- [5] Ruiz N, Li Y, Jampani V, et al. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation[C]//Proceedings of the

- IEEE/CVF conference on computer vision and pattern recognition. 2023: 22500-22510.
- [6] Saharia C, Chan W, Saxena S, et al. Photorealistic text-to-image diffusion models with deep language understanding[J]. Advances in neural information processing systems, 2022, 35: 36479-36494.
- [7] Xue Z, Song G, Guo Q, et al. Raphael: Text-to-image generation via large mixture of diffusion paths[J]. Advances in Neural Information Processing Systems, 2024, 36.
- [8] Zhang L, Rao A, Agrawala M. Adding conditional control to text-to-image diffusion models[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 3836-3847.
- [9] Chen J, Yu J, Ge C, et al. Pixart- α : Fast training of diffusion transformer for photorealistic text-to-image synthesis[J]. arxiv preprint arxiv:2310.00426, 2023.
- [10] Rombach R, Blattmann A, Lorenz D, et al. High-resolution image synthesis with latent diffusion models[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 10684-10695.
- [11] Liu N, Li S, Du Y, et al. Compositional visual generation with composable diffusion models[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 423-439.
- [12] Liew J H, Yan H, Zhou D, et al. Magicmix: Semantic mixing with diffusion models[EB/OL]. (2022-10-28)[2024-09-25].<https://arxiv.org/pdf/2210.16056v1.pdf>.
- [13] Chefer H, Alaluf Y, Vinker Y, et al. Attend-and-excite: Attention-based semantic guidance for text-to-image diffusion models[J]. ACM Transactions on Graphics (TOG), 2023, 42(4): 1-10.
- [14] Ma W D K, Lahiri A, Lewis J P, et al. Directed diffusion: Direct control of object placement through attention guidance[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38(5): 4098-4106.
- [15] Wang Z, Sha Z, Ding Z, et al. Tokencompose: Grounding diffusion with token-level supervision[EB/OL]. (2023-12-06)[2024-09-25].<https://arxiv.org/pdf/2312.03626v2.pdf>.
- [16] Lingenberg T, Reuter M, Sudhakaran G, et al. DIAGen: Diverse Image Augmentation with Generative Models for Few-Shot Learning[EB/OL].(2024-08-26)[2024-09-25].<https://arxiv.org/pdf/2408.14584v1.pdf>.
- [17] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740-755.
- [18] Zhou B, Zhao H, Puig X, et al. Scene parsing through ade20k dataset[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 633-641.
- [19] Zhou B, Zhao H, Puig X, et al. Semantic understanding of scenes through the ade20k dataset[J]. International Journal of Computer Vision, 2019, 127: 302-321.
- [20] Plummer B A, Wang L, Cervantes C M, et al. Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models[C]//Proceedings of the IEEE international conference on computer vision. 2015: 2641-2649.
- [21] Heusel M, Ramsauer H, Unterthiner T, et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium[J]. Advances in neural information processing systems, 2017, 30.
- [22] Vaswani A. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017.
- [23] Loshchilov I. Decoupled weight decay regularization[EB/OL].(2019-01-04)[2024-09-25].<https://arxiv.org/pdf/1711.05101v3.pdf>.
- [24] Chen M, Laina I, Vedaldi A. Training-free layout control with cross-attention guidance[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024: 5343-5353.
- [25] Feng W, He X, Fu T J, et al. Training-free structured diffusion guidance for compositional text-to-image synthesis[EB/OL].(2023-02-28)[2024-09-25].<https://arxiv.org/pdf/2212.05032v3.pdf>.

Enhanced Traffic Congestion Prediction Using Attention-Based Multi-Layer GRU Model with Feature Embedding

Sreelekha M¹, Dr. Midhunchakkaravarthy Janarathanan²

Research Scholar, Faculty of Engineering, Lincoln University College, Malaysia¹
Faculty of Computer Science and Multimedia, Lincoln University College, Malaysia²

Abstract—Intelligent Transportation Systems (ITS) are crucial for managing urban mobility and addressing traffic congestion, which poses significant challenges to modern cities. Traffic congestion leads to increased travel times, pollution, and fuel consumption, impacting both the environment and quality of life. Traditional traffic management solutions often fall short in predicting and adapting to dynamic traffic conditions. This study proposes an efficient deep learning (DL) model for predicting traffic congestion, utilizing the strengths of an attention-based multilayer Gated Recurrent Unit (GRU) network. The dataset used for this study includes 48,120 hourly vehicle counts across four junctions and additional weather data. Temporal and lagged features were engineered to capture daily and historical traffic trends and categorical data were considered by employing feature embedding. The attention-based GRU model integrates an attention mechanism to focus on relevant historical data, improving predictive performance by selectively emphasizing crucial time steps. This model architecture, consisting of two hidden layers and attention mechanisms, allows for nuanced traffic predictions by handling temporal dependencies and variations effectively. The performance was evaluated using various error metrics. The results demonstrate the model's ability to predict traffic congestion with MSE of 0.9678, MAE of 0.4322, R² of 0.8686, MAPE of 6% offering valuable insights for traffic management and urban planning.

Keywords—Intelligent transportation system; traffic congestion; urban mobility; deep learning; gated recurrent unit

I. INTRODUCTION

Traffic congestion is a persistent and growing problem in urban areas globally. It leads to significant challenges that affect economic productivity, environmental sustainability, and the quality of life [1]. As cities continue to expand and urbanization accelerates, the demand for road space often surpasses the available infrastructure, resulting in slower traffic speeds, longer travel times, and increased vehicular queuing. Several factors contribute to this issue, including the rapid rise in vehicle ownership, insufficient public transportation systems, and inadequate urban planning. The surge in private vehicles, due to the rising incomes and population growth, places immense pressure on existing road networks, intensifying congestion, especially during peak hours. In many cities, the public transportation infrastructure is either lacking or inefficient, prompting people to rely heavily on private car. Additionally, poor urban planning—such as poorly designed

road networks, insufficient parking, and a lack of infrastructure for pedestrians and cyclists further intensifies traffic blocks [2].

The impacts of traffic congestion are wide-ranging and severe. Economically, it leads to significant costs, including wasted fuel, vehicle maintenance, and lost productivity as people spend more time in traffic. Businesses suffer due to delays in goods transportation and reduced employee efficiency [3]. Environmentally, traffic congestion contributes to higher emissions of greenhouse gases and pollutants, as vehicles emit more while idling in traffic jams than when traveling smoothly. This not only worsens air quality but also accelerates climate change. Socially, congestion reduces the quality of life, as long travels lead to stress, less time for personal activities, and overall frustration. The safety concerns are also notable, with increased stop-and-go traffic raising the likelihood of accidents and making roads more dangerous for both drivers and pedestrians. One of the most critical consequences is its impact on emergency services [4]. Congested roads can significantly delay ambulances, fire trucks, and police vehicles, potentially leading to life-threatening situations due to increased response times.

Addressing traffic congestion requires a multi-faceted approach. Enhancing public transportation is a key strategy, as efficient and reliable public transit systems can reduce the number of private vehicles on the road. Investments in bus transit systems, light rail networks, and better integration of transport modes can make public transport a more attractive option [5]. Additionally, traffic management schemes that use real-time data to enhance traffic flow, such as adaptive traffic signal controls, can alleviate congestion at critical points. Urban planning also plays a crucial role; cities need to adopt designs that encourage high-density, mixed-use developments that reduce the need for long shuttles. Promoting smart mobility solutions, such as ride-sharing, autonomous vehicles, and Mobility-as-a-Service (MaaS), can also help by optimizing road usage. Behavioral changes, supported by public awareness campaigns and incentives for carpooling or telecommuting, are essential in reducing the number of vehicles on the road during peak times [6]. In summary, while traffic congestion is a complex problem with significant impacts, a combination of improved infrastructure, smart technology, and policy measures can help mitigate its effects and create more sustainable and functional urban environments. A DL model for traffic congestion prediction is proposed in this study. The main contributions of the study are given below:

- To develop a DL-based traffic congestion prediction model.
- To compare the effectiveness of the suggested model with existing models.
- To evaluate the efficiency through various error metrics.

The remaining portion of the paper is organized as: Section II provides a comprehensive literature review emphasizing the need for the current research. Section III details the methodology and the deep learning model architecture for effective traffic prediction. Section IV presents the results and discussion, highlighting the potential of the suggested model. Section V concludes the paper by summarizing the key contributions.

II. LITERATURE REVIEW

Li et al. [7] introduced the AST3DRNet model, which incorporated a 3D residual network with a self-attention mechanism. This approach utilized a 3D convolutional module and employs a spatio-temporal attention module to dynamically adjust the impact of these relationships. Experiments conducted using a real-world traffic dataset from Kunming demonstrated that AST3DRNet outperformed existing baseline methods, achieving accuracy improvements of 59.05%, 64.69%, and 48.22% for short-term predictions at 5, 10, and 15 minutes, respectively. Despite its innovations, the model's dependency on convolutional neural networks (CNN) and residual networks was a limitation.

Tsalikidis et al. [8] evaluated various models for multi-step forecasting of traffic flow, particularly in areas with limited historical data. The methodology involved assessing a range of interpretable predictive algorithms, including Ensemble Tree-Based (ETB) regressors like Light Gradient Boosting Machine (LGBM) and comparing them with traditional deep learning methods. Results indicated that ETB models generally outperformed DL approaches, particularly for longer forecasting horizons, achieving high accuracy even at extended prediction steps. The study demonstrated that feature selection and engineering, incorporating temporal and weather data, improved model performance. The study was limited by its reliance on the statistical characteristics of the specific dataset, which could affect the efficiency of the algorithms. High data volume and complexity also posed challenges, impacting model training and performance.

Jiang et al. [9] introduced Congestion Prediction Mixture-of-Experts (CP-MoE) to improve prediction accuracy for dynamic traffic scenarios. The methodology involved developing a Mixture of Adaptive Graph Learners (MAGLs) with a sparsely-gated mechanism and congestion-aware biases, complemented by two specialized experts designed to identify stable trends and periodic patterns. This model was rigorously tested on real-world datasets, demonstrating its superiority over existing spatio-temporal prediction methods. Notably, CP-MoE was successfully integrated into DiDi's system, enhancing travel time estimation reliability. A key limitation identified was the utility of CP-MoE's application to other aspects of ride-hailing services.

Hao et al. [10] presented a fuzzy logic system based on the Greenshields model, designed to predict highway traffic congestion without requiring extensive training data. The methodology involved processing vehicle speed and traffic flow inputs using specified membership functions and applying fuzzy rules guided by Greenshields theory. The approach was validated through a comparative analysis with a polynomial regression model using real-world data from the Sun Yat-Sen Highway in Taiwan, demonstrating consistent prediction results. The fuzzy logic system proved effective in estimating congestion levels and adapting to various road conditions with minimal data preparation. A noted limitation was the potential for reduced precision in highly dynamic traffic scenarios where the fuzzy logic system's fixed rules not capture complex variations as effectively.

Zhang et al. [11] introduced a deep marked graph process (DMGP) model that combined a spatiotemporal convolutional graph network with a traditional point process model to predict congestion indices and occurrence times for large signalized road networks. This hybrid approach utilized the simplicity of the point process model and the advanced capabilities of graph neural networks to model the evolution of traffic congestion. Experiments using real-world traffic data demonstrated that the DMGP model outperformed existing baseline methods, achieving superior prediction accuracy and computational efficiency. While the model showed promise in supporting advanced traffic management and traveler information systems, a significant limitation was its reliance on high-quality citywide traffic data, which had not been available for all road segments.

Jasim et al. [12] analyzed the efficiency of several machine learning (ML) algorithms for congestion detection and prediction within Vehicular Ad hoc Networks (VANETs). The study focused on Support Vector Machines (SVM), Ensemble Learning classifiers, K-Nearest Neighbors (KNN), and Decision Trees (DT). The methodology involved training these algorithms with historical traffic congestion data and applying advanced feature engineering techniques. The study found that SVM, along with KNN and Ensemble Learning classifiers, achieved high classification accuracies. The study was limited by the dependence on precise feature selection and model optimization techniques, which required careful tuning.

Arabiat et al. [13] addressed the challenge of predicting traffic congestion using data mining and ML techniques. The study compared the performance of two open-source software tools, WEKA and Orange, in predicting traffic congestion in Amman, Jordan. Various classifiers, including SVM, KNN, Logistic Regression (LR), and Random Forest (RF), were tested using data from the Greater Amman Municipality for the year 2018. Results revealed that Orange excelled with high prediction accuracy. The study highlighted the superior performance of Orange over WEKA, particularly in handling different classifiers. A notable limitation was the reliance on specific data mining tools, which are not generalize across all types of traffic data or scenarios, potentially affecting the applicability of the findings.

Chahal et al. [14] tackled the challenge of traffic flow prediction using a hybrid model that combines Seasonal Auto-

Regressive Integrated Moving Average (SARIMA) with Bidirectional Long Short-Term Memory (Bi-LSTM) and Back Propagation Neural Network (BPNN). This approach aimed to address both linear and non-linear components of traffic data from the CityPulse EU FP7 project. The hybrid model demonstrated superior performance with the lowest MAE of 0.499 compared to single SARIMA, LSTM, and other models. The study was limited by the feature set considered, as the model did not account for external factors like weather or peak hours, which could affect traffic predictions.

Jin et al. [15] introduced a spatio-temporal graph neural point process (STGNPP) framework specifically designed to predict traffic congestion events that occur sporadically over time. The model incorporated a spatio-temporal graph learning component to efficiently capture long-term dependencies from historical traffic data and road network information. This information is then processed through a continuous GRU to model congestion evolution patterns, with a periodic gated mechanism enhancing the intensity function to account for periodic variations. Extensive experiments conducted on two large-scale real-world datasets demonstrated that STGNPP significantly outperformed existing methods in predicting both the timing and duration of congestion events. However, the model's reliance on historical data may limit its adaptability to sudden or unprecedented traffic disruptions.

Pan et al. [16] presented Ising-Traffic, a dual-model framework that employs the Ising model to address traffic management. Unlike conventional approaches that struggle with balancing algorithmic complexity and computational efficiency, Ising-Traffic combines two distinct Ising models: Predict-Ising and Reconstruct-Ising. Reconstruct-Ising utilized advanced Ising machines to handle traffic uncertainties with reduced latency and lower energy consumption, while Predict-Ising uses conventional processors to project future traffic congestion, requiring only 1.8% of the computational resources compared to existing methods. The proposed framework demonstrated an average speed up of 98 \times and a 5% accuracy improvement over conventional solutions when evaluated on real-world traffic datasets. A notable limitation of this approach was the dependency on specific hardware for Reconstruct-Ising, which affect its scalability and adaptability across different computational platforms.

Zhang et al. [17] explored urban traffic condition prediction and congestion control by integrating improved particle swarm optimization (IPSO) with radial basis function (RBF) networks and a fusion model of LSTM networks and SVM. The

proposed feature fusion model demonstrated superior performance in predicting traffic states, validated by experiments using regional traffic data from Shenyang Station, with the model achieving the lowest RMSE compared to other algorithms. For congestion control, a traffic allocation-based method was developed and tested using VISSIM simulation, showing effective congestion management. The primary limitation of the study was the reliance on simulation models to fully capture the complexities of real-world traffic dynamics and thus limit the applicability of the proposed methods in varied urban settings.

Wang et al. [18] addressed urban traffic congestion by developing a prediction model called Spatio-Temporal Transformer (STTF), which utilizes DL techniques. Traditional models struggled with the growing complexity of urban traffic networks, prompting the introduction of STTF. This model integrated traffic speed data, road network structure, and spatio-temporal correlations to enhance prediction accuracy. The STTF employed an information embedding module to convert both spatial and temporal data into feature vectors, which were then processed through spatial and temporal attention modules. The model was tested on real-world datasets, demonstrating substantial improvements in prediction accuracy compared to existing methods. Despite its advancements, the STTF model's main limitation was its dependence on comprehensive feature engineering and attention mechanisms, which increase computational complexity and impact its efficiency in real-time applications. Table I provides the summary of the existing traffic congestion prediction models.

While existing models, such as traditional spatio-temporal graph-based methods and hybrid approaches, have made significant strides, they often struggle with real-time efficiency and adaptability to rapidly changing traffic conditions. Additionally, existing approaches frequently lack the capability to adaptively weigh the importance of different time steps or traffic features, leading to suboptimal predictions. A critical gap in current traffic congestion prediction methods lies in the need for more advanced deep learning models that can seamlessly integrate and process complex spatio-temporal dependencies and dynamic factors. Deep learning models offer promising avenues for capturing intricate patterns in traffic data and improving prediction accuracy. Addressing these gaps requires the development of deep learning frameworks that can balance high accuracy with operational efficiency, effectively handling large-scale, dynamic data while being robust to varying traffic conditions and external influencing factors.

TABLE I. SUMMARY OF EXISTING TRAFFIC CONGESTION PREDICTION MODELS

Ref. No.	Works Carried Out in the Reference Papers	Advantages	Disadvantages
[7]	AST3DRNet model with a 3D residual network and spatio-temporal attention module for traffic prediction.	Achieved accuracy improvements of 59.05%, 64.69%, and 48.22% for 5, 10, and 15-minute predictions.	Dependency on CNNs and residual networks limits the scalability and adaptability of the model.
[8]	Ensemble Tree-Based (ETB) regressors (e.g., LGBM) compared with traditional DL methods for multi-step forecasting.	ETB models outperformed DL approaches, especially for long-term predictions; feature engineering improved performance.	Results relied on statistical characteristics of the dataset; complexity in high-volume data handling.
[9]	CP-MoE with Mixture of Adaptive Graph Learners and congestion-aware biases for dynamic traffic prediction.	Outperformed baseline methods; integrated into DiDi's system to enhance travel time reliability.	Limited applicability to other ride-hailing service aspects; utility depends on specific use cases.
[10]	Fuzzy logic system using Greenshields model for highway congestion prediction with minimal data preparation.	Consistent results; adaptable to various road conditions with minimal data.	Lower precision in dynamic scenarios with complex variations.
[11]	Deep Marked Graph Process (DMGP) combining spatiotemporal graph network and point process model for congestion prediction.	Superior prediction accuracy and computational efficiency.	Reliance on high-quality, citywide traffic data limits scalability to less monitored areas.
[12]	ML algorithms (SVM, KNN, Ensemble Learning, DT) with advanced feature engineering for VANET congestion detection.	High classification accuracy achieved with SVM, KNN, and Ensemble Learning classifiers.	Dependence on precise feature selection and careful optimization of models.
[13]	Comparison of WEKA and Orange tools for traffic prediction using ML classifiers like SVM, KNN, LR, and RF.	Orange achieved superior prediction accuracy over WEKA.	Limited generalizability across different traffic datasets and tools.
[14]	Hybrid SARIMA-Bi-LSTM-BPNN model for traffic flow prediction, addressing both linear and non-linear components.	Lowest MAE (0.499) compared to single models; superior performance on CityPulse EU FP7 data.	Did not account for external factors like weather or peak hours affecting predictions.
[15]	STGNPP framework with spatio-temporal graph learning and periodic gated mechanism for sporadic traffic event prediction.	Outperformed existing methods in predicting timing and duration of congestion events.	Limited adaptability to sudden or unprecedented disruptions due to reliance on historical data.
[16]	Ising-Traffic framework combining Predict-Ising and Reconstruct-Ising models for traffic management.	Achieved 98× speedup and 5% accuracy improvement; reduced latency and energy consumption.	Dependency on specific hardware for Reconstruct-Ising limits scalability.
[17]	IPSO-RBF network fusion model with LSTM and SVM for traffic prediction and congestion control via traffic allocation.	Superior performance in RMSE; effective congestion management in simulations.	Limited real-world validation; reliance on simulations restricts practical applicability.
[18]	Spatio-Temporal Transformer (STTF) integrating traffic speed, road networks, and spatio-temporal correlations.	Substantial accuracy improvements compared to existing methods.	High computational complexity due to feature engineering and attention mechanisms.

III. MATERIALS AND METHODS

Traffic congestion prediction is crucial for optimizing traffic flow and enhancing commuter experience by enabling more efficient traffic management and route planning. It helps reduce economic losses associated with delays, fuel consumption, and vehicle wear, benefiting both individuals and

businesses. Accurate predictions also contribute to minimize the idle time and slow-moving traffic, thereby supporting environmental sustainability. Additionally, it provides valuable data for urban planners to design better infrastructure and improve overall urban mobility. Thus, this study proposes an efficient DL technique for traffic congestion prediction. Fig. 1 shows the detailed block diagram of the proposed traffic congestion prediction model.

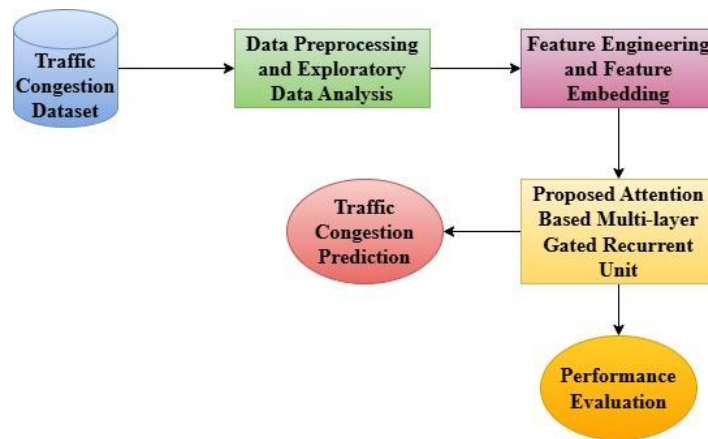


Fig. 1. Block diagram of proposed traffic congestion prediction model.

A. Dataset Description

The study utilized traffic data from Kaggle repository [19]. The dataset comprises 48,120 observations of hourly vehicle counts across four different junctions during the periods of November 1, 2015, to July 1, 2017, with observations distributed across different months and years. The data was collected by sensors placed at each junction, though these

sensors operated at different times, resulting in traffic data from various time periods. The extensive range of hourly traffic counts across multiple junctions provides a valuable resource for modeling and predicting traffic congestion. The key features in the dataset are tabulated in Table II.

TABLE II. KEY FEATURES IN THE DATASET

Features	Description
Date and Time	The specific date and time of the observation.
Junction	The identifier for the junction where the data was collected.
Vehicles	The number of vehicles counted at the junction during the specified hour.
ID	A unique identifier for each observation.

The study incorporated weather data [20] of same period of time for creating more accurate, responsive, and comprehensive traffic management systems. It helps in optimizing traffic flow, improving safety, enhancing emergency response, and supporting long-term infrastructure planning. By understanding how weather influences traffic patterns, cities can better prepare for and mitigate the impacts of both regular and extreme weather conditions on their transportation networks. The dataset provides the temperature alongside two types of radiation measurements: direct horizontal radiation and diffuse horizontal radiation. These metrics are essential for understanding the overall weather conditions, as direct radiation measures the sunlight that reaches the ground without scattering, while diffuse radiation accounts for sunlight scattered by the atmosphere.

specific range, allowing for easy identification of common traffic volume ranges and patterns. The plot also helps in detecting any skewness in the data, understanding the spread of traffic volumes, and spotting potential outliers.

B. Preprocessing and Exploratory Data Analysis

Preprocessing and EDA are crucial in the data analysis pipeline, with preprocessing ensuring the data is suitable for analysis and EDA providing insights and understanding of the data [21]. Fig. 2 provides the statistics of the data statistics.

The histogram in Fig. 3 provides a visual representation of the distribution of traffic volumes in the dataset. It displays how frequently different ranges of vehicle counts occur by dividing the data into 30 bins. Each bar in the histogram represents the frequency of vehicle counts falling within a

	Junction	Vehicles	ID
count	48120.000000	48120.000000	4.812000e+04
mean	2.180549	22.791334	2.016330e+10
std	0.966955	20.750063	5.944854e+06
min	1.000000	1.000000	2.015110e+10
25%	1.000000	9.000000	2.016042e+10
50%	2.000000	15.000000	2.016093e+10
75%	3.000000	29.000000	2.017023e+10
max	4.000000	180.000000	2.017063e+10

Fig. 2. Data statistics.

The line plot shown in Fig. 4 illustrates the average traffic volume over time by resampling the data on a daily basis. It shows how the average number of vehicles varies across different days, providing insights into daily traffic trends and fluctuations. The boxplots in Fig. 5 compare traffic volumes across different junctions, highlighting variations in vehicle counts. Each boxplot visualizes the distribution of traffic volumes for each junction, showing median values, interquartile ranges, and any outliers.

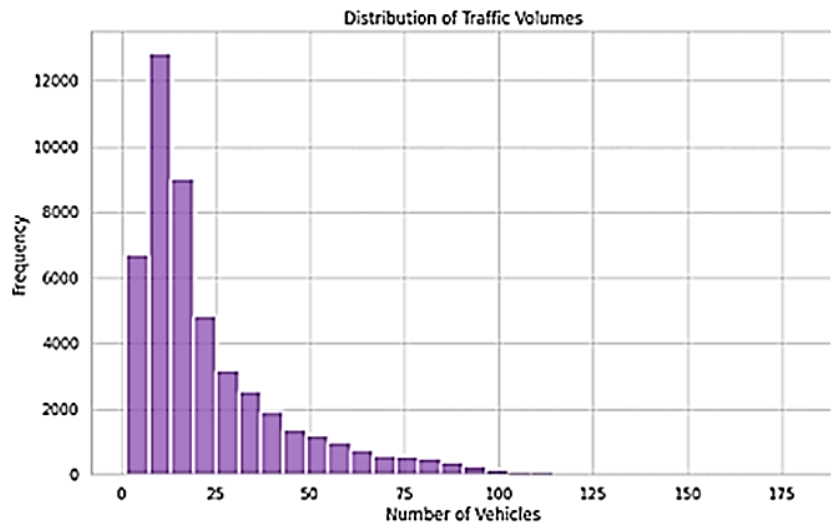


Fig. 3. Distribution of traffic volumes.

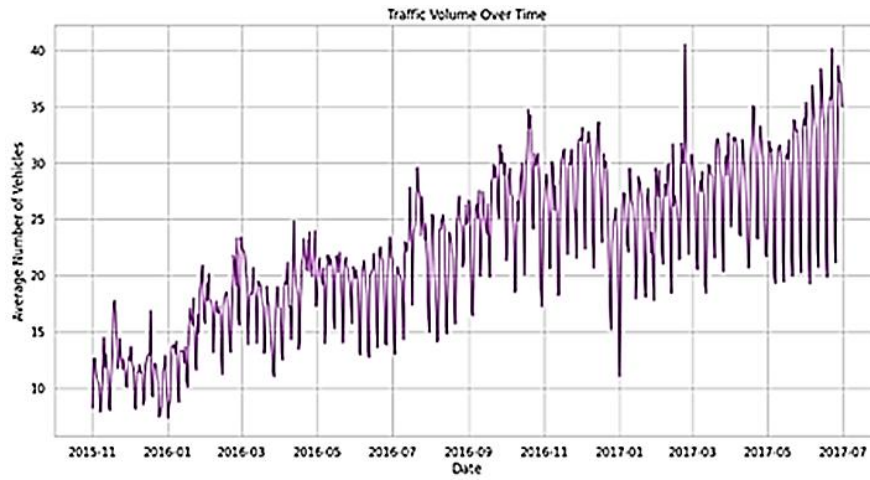


Fig. 4. Traffic volume over time.

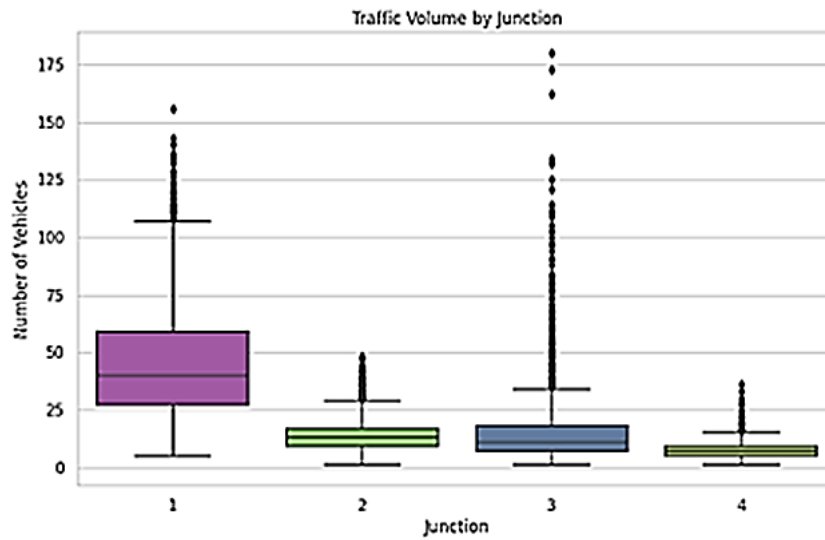


Fig. 5. Traffic volume by junction.

The average traffic volume for each hour and each day of the week is visualized in Fig. 6, revealing variations in traffic patterns across different days. These visualizations help in

identifying peak traffic times and understanding daily and weekly traffic patterns.

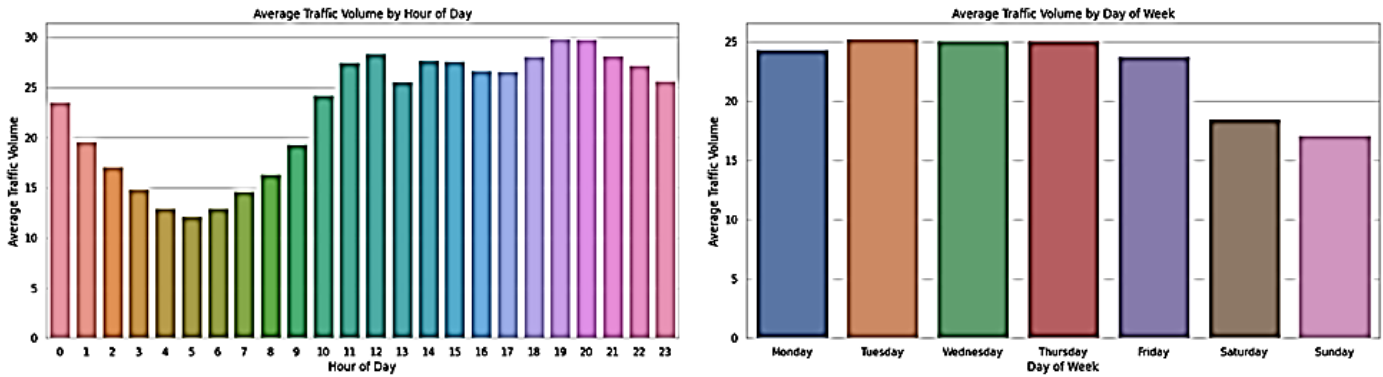


Fig. 6. Average traffic volume for (a) hour of the day (b) day of the week.

The analysis involves identifying peak and off-peak hours for traffic management by calculating the average traffic volume for each hour and each day of the week, highlighting times of high and low traffic as in Fig. 7.

The trend component shows long-term changes in traffic volume, indicating periods of increase or decrease, but without providing conclusive trends due to its variability. The residuals, representing the noise after accounting for trend and seasonality, are scattered around zero with a few outliers. Using a correlation matrix in Fig. 9, the correlation analysis examines at the associations between traffic volume and variables like the day of the week and hour of the day.

Additionally, time series decomposition is performed to understand the underlying patterns in traffic volume data. By breaking down the data into trend, seasonal, and residual components, this approach reveals long-term trends, recurring seasonal effects, and irregular variations, providing a clear scenario of traffic dynamics over time as in Fig. 8. This comprehensive analysis supports better traffic management and planning by pinpointing peak traffic periods and understanding traffic behavior patterns.

The correlation matrix reveals that there is a moderate positive correlation between traffic volume and the hour of the day, indicating that traffic volume tends to increase with later hours. In contrast, the correlation between traffic volume and the day of the week is weakly negative, suggesting minimal variation in traffic volume across different days. The traffic flow variations throughout the week are illustrated in Fig. 10, which reflects patterns in commuter and commercial traffic.

The traffic volume analysis reveals a clear seasonal pattern with fluctuations that repeat on a weekly basis, suggesting regular peaks and troughs corresponding to weekly traffic

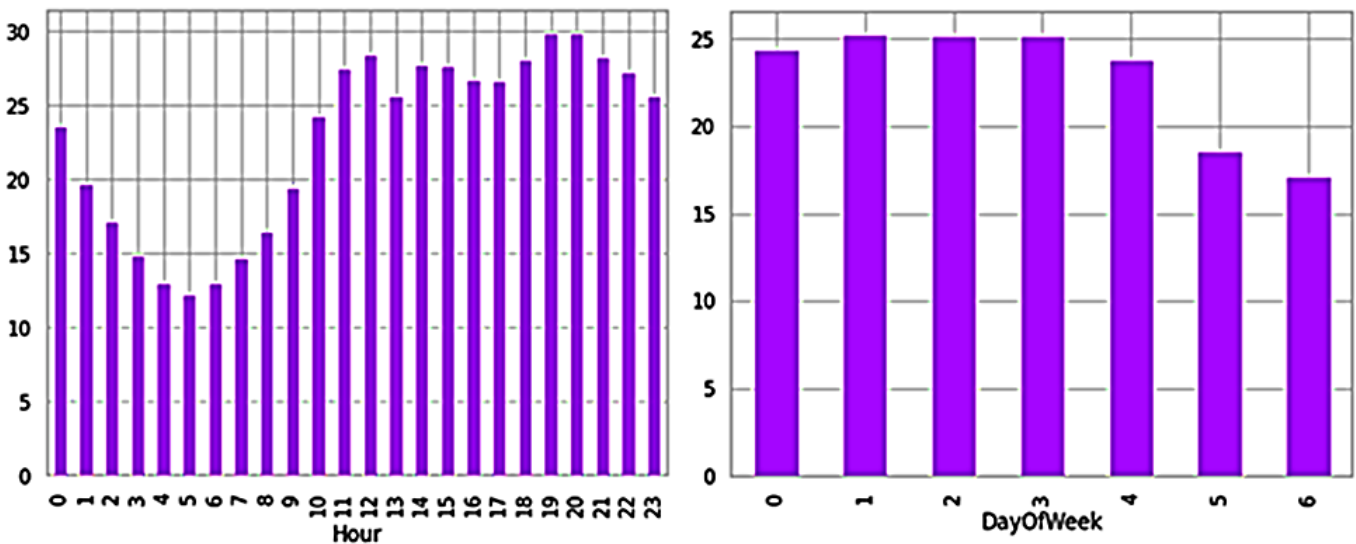


Fig. 7. Peak hour analysis.

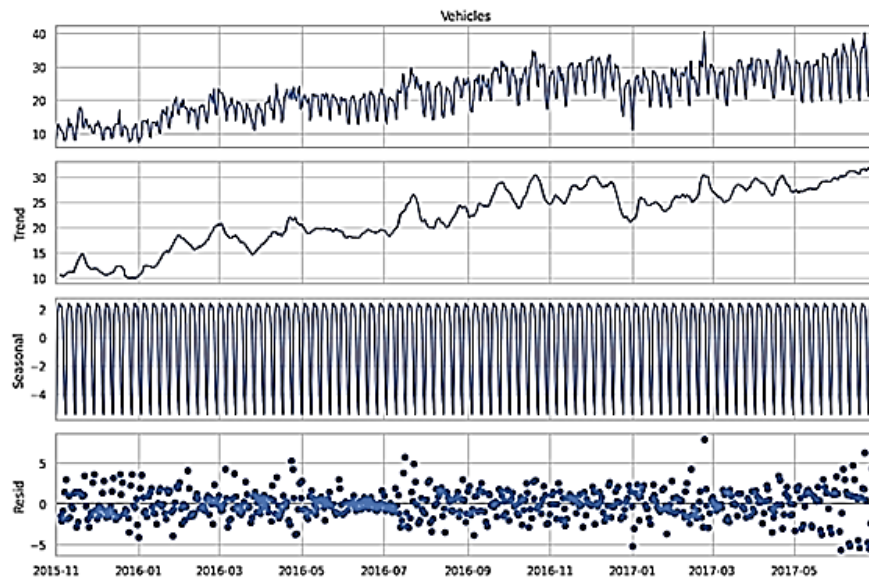


Fig. 8. Time series decomposition.

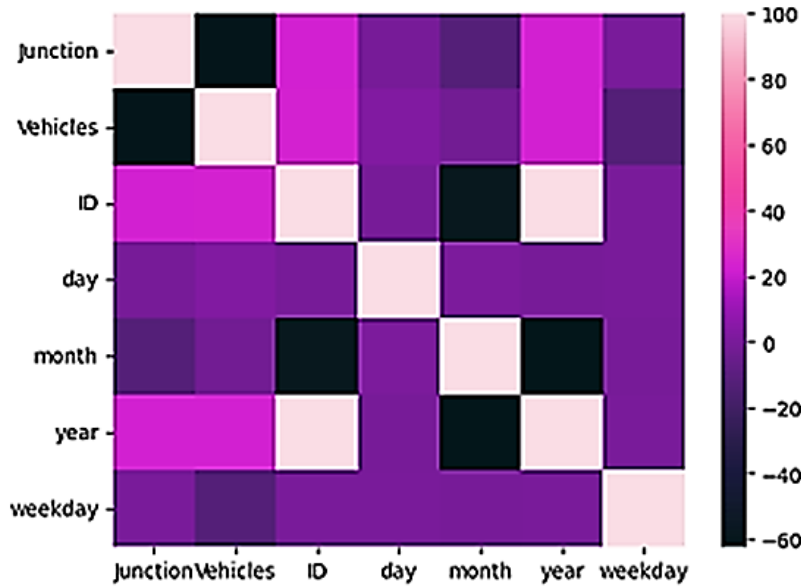


Fig. 9. Correlation matrix.

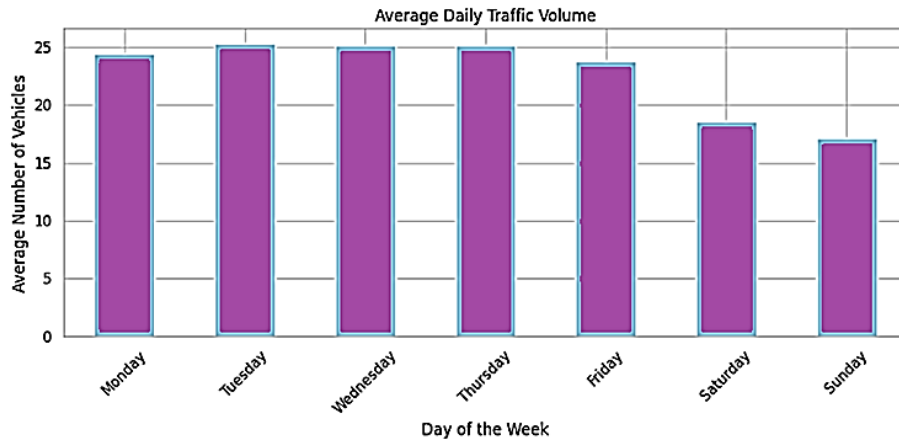


Fig. 10. Average daily traffic rate.

The analysis of holiday effects on traffic volumes, shown in Fig. 11, aims to determine how public holidays impact traffic patterns compared to non-holidays. By marking specific public holidays and comparing the average traffic volumes on these days to those on regular days, the analysis identifies any significant differences in traffic flow.

For long-term trend analysis, monthly traffic volumes were examined to assess any significant changes over an extended period. The monthly average traffic volume data as illustrated in Fig. 12, indicates a general upward trend, suggesting that traffic has been increasing over time. An addition of a trend line to the plot confirmed this long-term upward trajectory.

This trend could reflect urban development, population growth, or other factors influencing traffic patterns. Such insights are valuable for traffic management and infrastructure planning, as they highlight the need for adapting strategies to handle increasing traffic volumes.

The junction comparison analysis in Fig. 13 highlights variations in average traffic volumes across different junctions. By grouping the traffic data by junction and calculating the average number of vehicles for each, the analysis reveals that Junction 1 consistently experiences the highest average traffic volume, significantly surpassing the other junctions.

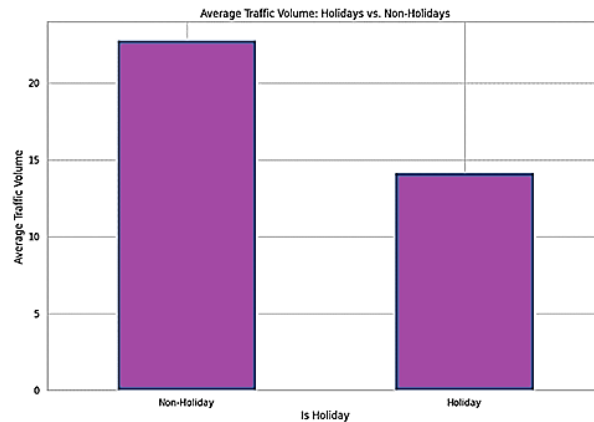


Fig. 11. Holiday effect on traffic volume.

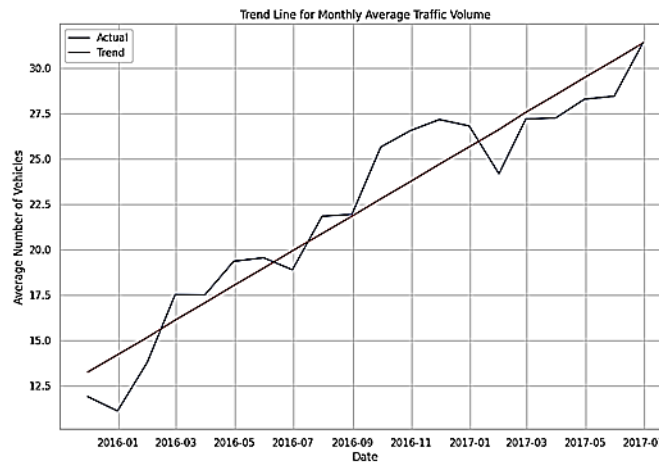


Fig. 12. Monthly average traffic volume over time with a linear trend line.

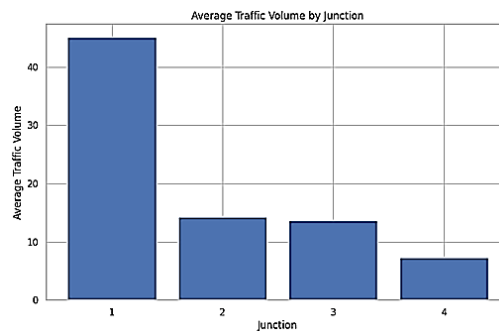


Fig. 13. Average traffic volume by junction.

The analysis of daily traffic volumes through Z-scores has identified specific dates with unusually high traffic levels. Z-scores quantify how far each data point deviates from the average, highlighting days with significantly above-average traffic. These elevated traffic volumes could be attributed to various factors. For instance, there have been special events, like concerts or sports games, that led to increased traffic on these days. Alternatively, temporary disruptions such as road construction or detours have redirected traffic through these areas, causing a spike. Additionally, seasonal patterns or local events also explain the higher traffic volumes observed.

Incorporating the weather data, the scatter plot in Fig. 14 shows that there is no evident correlation between temperature and the number of vehicles, suggesting that temperature has no significant impact on traffic volume.

The analysis of daily traffic volumes included the application of the Augmented Dickey–Fuller (ADF) test to assess the stationarity of the time series data. Stationarity is a critical property for time series analysis, as non-stationary data can lead to misleading results in forecasting models. The ADF test was employed to test the null hypothesis that the traffic

volume time series contains a unit root, which would indicate non-stationarity. The results of the ADF test revealed a significantly negative ADF statistic and a p-value much smaller than conventional significance levels. These findings strongly reject the null hypothesis, indicating that the traffic data is stationary. The Auto-Correlation Function (ACF) plot displays the correlation between the data and its lagged values over time, while the Partial Auto-Correlation Function (PACF) plot shows the direct correlation at specific lags, controlling for the effects of intermediate lags as in Fig. 15.

To complement this stationarity check, a Z-score analysis was performed to identify anomalies in daily traffic volumes. By calculating Z-scores, which indicate how many standard deviations a data point is from the mean, the analysis was able to identify days with significantly higher or lower traffic volumes compared to the average. These anomalies were then visualized on a line graph as in Fig. 16, with red dots marking the days where traffic volumes deviated notably from the norm. This visual representation provided insights into trends, potential seasonality, and outlier events that could be linked to external factors such as road closures, construction projects, or special events.

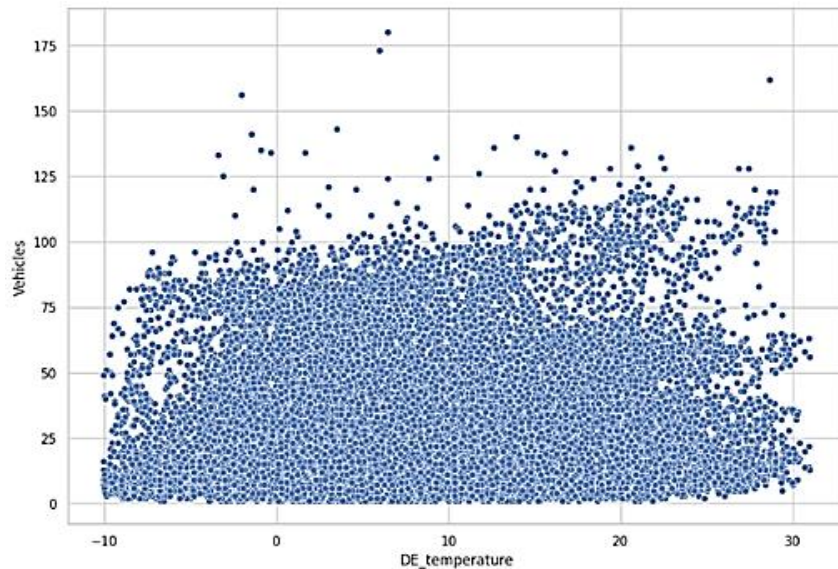


Fig. 14. Scatter plot of temperature vs. number of vehicles.

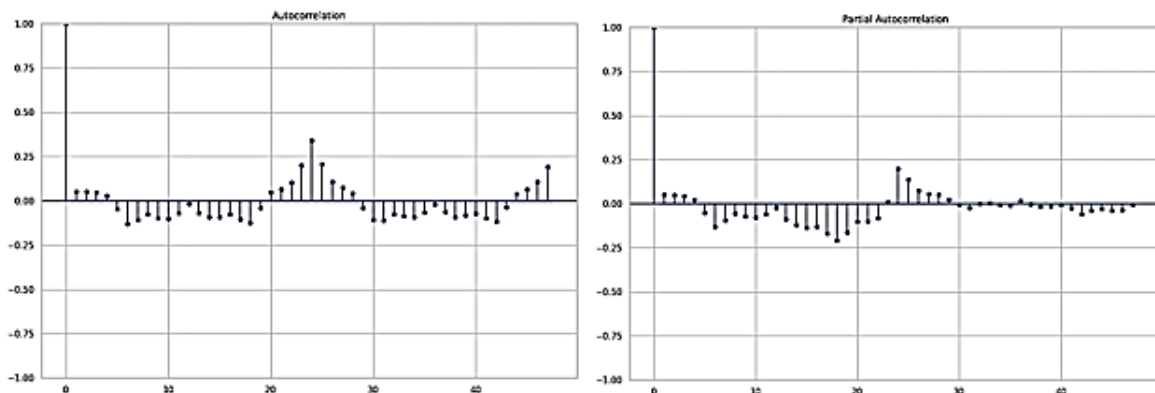


Fig. 15. ACF and PACF plot.

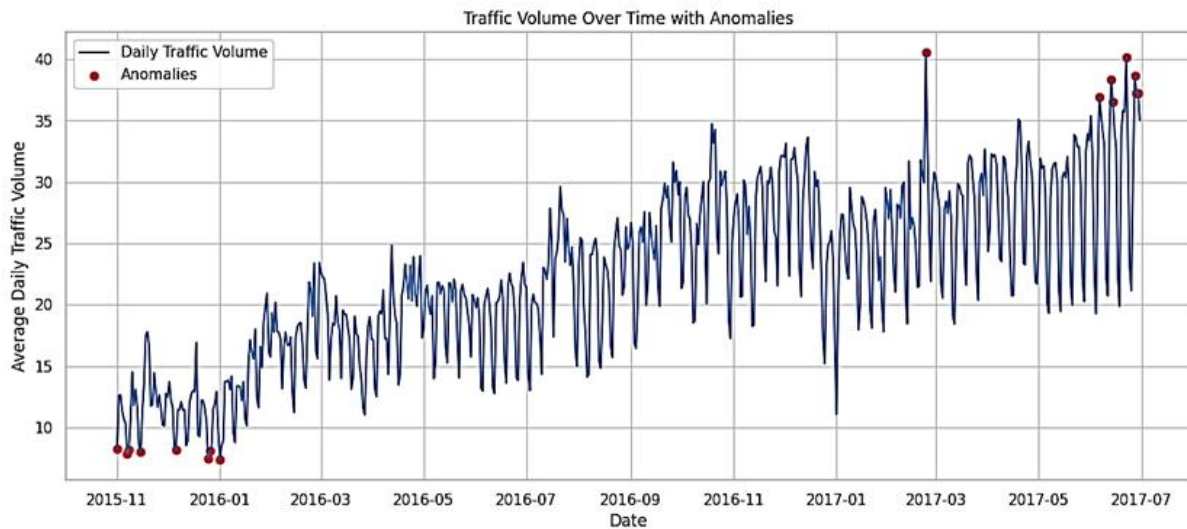


Fig. 16. Traffic volume over time with anomalies.

Additionally, the analysis involved checking for missing data, consistency of data reporting, and potential outliers. Missing timestamps were identified and accounted for, ensuring that the data was complete and accurately represented. The consistency of data reporting was verified by examining the number of records per junction and analyzing the time

intervals between records. This step ensured that data collection was uniform across different junctions and time periods. Outlier detection further refined the analysis by identifying traffic volumes that were unusually high or low as depicted in Fig. 17, which could distort the overall findings if not properly addressed.

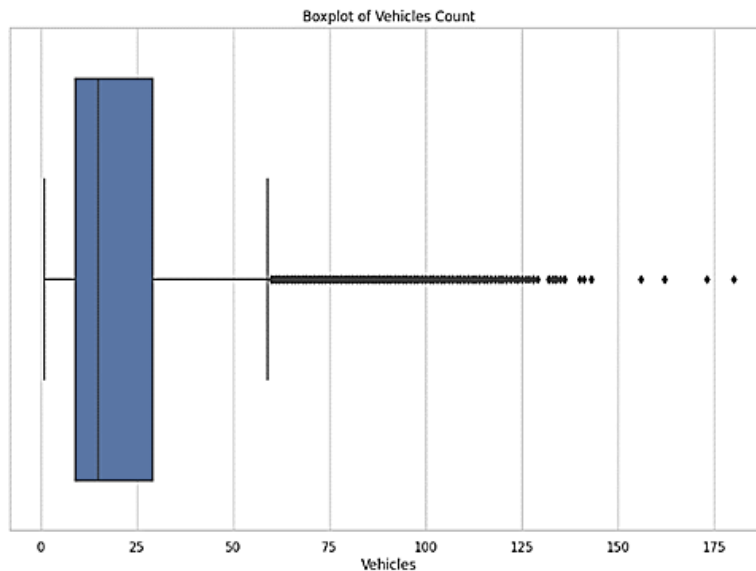


Fig. 17. Box plot of vehicle count.

C. Feature Engineering and Embedding

Feature engineering is a crucial step in improving model performance by creating and transforming features to capture the underlying patterns in traffic congestion data [22]. One of the primary types of features engineered for this purpose is temporal features. These include the hour of the day (extracted from the timestamp) to capture daily traffic variations, the day of the week to distinguish between weekday and weekend traffic patterns, and the month to account for seasonal trends. Additionally, a holiday indicator is used as a binary feature to differentiate between holidays and regular days, which often

exhibit different traffic behaviors. In addition to temporal features, lagged features are introduced, such as the previous hour traffic volume, which helps in incorporating short-term historical trends into the model as represented by Eq. (1).

$$traffic_volume_lag_k = traffic_volume_{(t-k)} \quad (1)$$

where, k denotes the lags in hours, $traffic_volume_{(t-k)}$ denotes the traffic volume at time $t - k$, indicating the value of the traffic volume variable k periods before the current time t . This is particularly useful for predicting current traffic conditions based on recent patterns. Aggregated features like

the daily average traffic volume are also created by averaging traffic data over a day, which helps smooth out short-term fluctuations and captures overall daily trends as illustrated by Eq. (2).

$$daily_avg_traffic = \frac{1}{N} \sum_{i=1}^N traffic_volume_i \quad (2)$$

where N is the total number of observations. Furthermore, interaction features are engineered by combining different factors, such as the interaction between the hour of the day and weather conditions, to capture the combined effect on traffic patterns. Normalization standardizes the features by subtracting the mean and dividing by the standard deviation.

$$scaled_feature = \frac{feature_mean}{std_dev} \quad (3)$$

Feature embedding is particularly useful when dealing with categorical features that have a large number of unique values, such as Junction IDs in traffic data. By converting these categorical variables into dense vectors, feature embedding allows the model to learn complex relationships within the data. Junction IDs are categorical variables representing different traffic junctions. Temporal features are represented by time-based encodings. Using an embedding layer, each unique Junction ID is mapped to a continuous vector in a high-dimensional space as Eq. (4). This allows the model to capture similarities between different junctions.

$$Embedding\ matrix, E \in \mathbb{R}^{V \times d} \quad (4)$$

where, V is the number of unique junctions and d is the dimensionality of the embedding vector. Each junction i is represented by Eq. 5.

$$E_i \in \mathbb{R}^d \quad (5)$$

The embedding matrix E is initialized randomly and is learned during the training process. The embeddings are updated to minimize the loss function, allowing the model to capture relevant patterns in the data as Eq. (6).

$$embedded_vector = E_i \quad (6)$$

D. Proposed Traffic Congestion Prediction Model

The attention-based multilayer GRU model is designed to handle sequential data, such as traffic flow over time, by utilizing both the GRU for capturing temporal dependencies and an attention mechanism to focus on the most relevant time steps. This approach helps in improving the model's predictive performance by selectively concentrating on important historical data.

1) *Attention based multi-layer gated recurrent unit*: The GRU is a variant of Recurrent Neural Networks (RNNs) designed to handle sequential data effectively [23]. It employs update gates and reset gates to manage the flow of information through the network. Fig. 18 illustrates the GRU cell architecture.

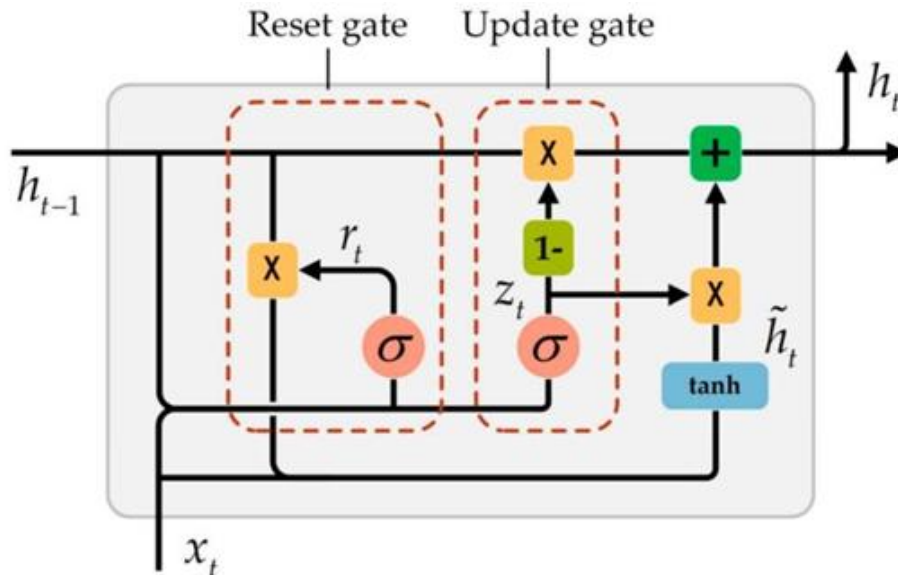


Fig. 18. Cell structure of GRU.

In a GRU cell, a gate controller, represented by z, oversees the operation of both the input and forget gates. When z equals 1, the forget gate is turned off, enabling the input gate to function. On the other hand, when z equals 0, the forget gate is activated and the input gate is turned off. At every time step, the GRU cell maintains the memory from the previous time step (t - 1) while resetting the input for the current step. The operation of the GRU cell is governed by the following equations: Eq. (7) through Eq. (9).

Reset Gate (r_{ti}),

$$r_{ti} = \sigma(W_r \cdot [h_{ti-1}, x_{ti}] + b_r) \quad (7)$$

Update Gate (z_{ti}),

$$z_{ti} = \sigma(W_z \cdot [h_{ti-1}, x_{ti}] + b_z) \quad (8)$$

Candidate Activation (\tilde{h}_{ti}),

$$h_{ti} = (1 - z_{ti}) * h_{ti-1} + z_{ti} * \tilde{h}_{ti} \quad (9)$$

The GRU network is employed to forecast traffic congestion levels at 24 distinct time intervals, spanning from one hour to one day ahead, for model optimization. This GRU model features two hidden layers, with the input layer having 18 nodes and each hidden layer containing 13 nodes, as determined by the two-thirds rule applied to the input layer size and the inclusion of the output layer size. When predicting traffic congestion, extending the input sequence in a GRU network can reduce prediction accuracy because the model tends to equally weight all input variables despite their varying relevance to the forecasted outcomes. To mitigate this issue, an attention mechanism is incorporated, enabling the model to prioritize the most pertinent input variables.

The attention mechanism comprises an encoder that creates an attention vector from the input data and a decoder that generates a hidden state based on the encoder's output [24]. The encoder produces hidden states h_t for each time step t . These hidden states are segmented, and the encoder calculates an attention score $e_{t'}^t$ for each segment's hidden state using the hidden state from the preceding decoder segment. The attention score is calculated as specified in Eq. (10).

$$e_{t'}^t = \text{score}(h_{t'}, h_t) \quad (10)$$

This process creates an attention vector through a Softmax operation on the attention scores as given by Eq. (11).

$$\alpha_{t'}^t = \frac{\exp(e_{t'}^t)}{\sum_k \exp(e_{t'}^k)} \quad (11)$$

The context vector $c_{t'}$ is then computed as a weighted sum of the encoder hidden states as in Eq. (12), where the weights are the attention weights.

$$c_{t'} = \sum_t \alpha_{t'}^t h_t \quad (12)$$

This method ensures that the encoder concentrates on input variables that are closely related to the predicted value whenever the decoder generates an output. The decoder uses the context vector $c_{t'}$ along with its previous hidden state to generate the next hidden state and output as in Eq. (13).

$$h_{t'} = \tanh(W_h[c_{t'}, h_{t'-1}] + b_h) \quad (13)$$

To enhance the accuracy of traffic congestion predictions, the attention mechanism is designed to focus on highly correlated input variables. The size of the attention window is set to 96 for this specific model configuration. The attention based GRU model architecture is illustrated by Fig. 19.

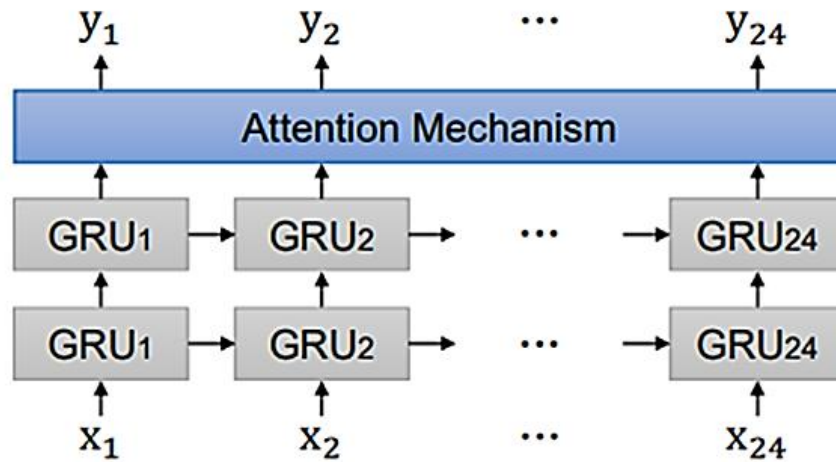


Fig. 19. Attention-based GRU model architecture.

Thus, the attention-based multi-layer GRU captures and processes the temporal dependencies in traffic data, the dense layer integrates and refines these features, and the output layer generates the final traffic prediction based on the transformed data.

E. Hardware and Software Setup

The experimental setup included an NVIDIA GeForce GTX 1080Ti GPU, an Intel Core i7 processor, 32GB of RAM,

and utilized python and the Keras library with TensorFlow as the backend. Keras' intuitive interface, combined with the computational power of Google Colab, enabled efficient model training with GPU support. The dataset was split for training and testing to ensure robust evaluation. Table III outlines the hyperparameters chosen for the training phase, which played a critical role in fine-tuning the model's performance on the traffic prediction dataset, ensuring both accuracy and rapid convergence.

TABLE III. HYPERPARAMETER SPECIFICATION

Hyperparameters	Values
Loss function	Mean squared error
Activation function	Sigmoid
Batch size	32
Epochs	150
Optimizer	Adam
Learning rate	0.001

IV. RESULTS AND DISCUSSION

The model's performance was assessed by means of various metrics, as detailed in Table IV. Table V shows the model

performance assessment using evaluation metrics for traffic congestion prediction.

TABLE IV. EVALUATION METRICS

Metric	Equation
Mean Squared Error (MSE)	$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$
Mean Absolute Error (MAE)	$MAE = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i $
Coefficient of Determination (R ²)	$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$
Mean Absolute Percentage Error (MAPE)	$MAPE = \frac{1}{n} \sum_{i=1}^n \left \frac{y_i - \hat{y}_i}{y_i} \right \times 100$

n is the number of observations, *y_i* is the actual value, \hat{y}_i is the predicted value

TABLE V. PERFORMANCE ASSESSMENT USING EVALUATION METRICS

Evaluation metrics	Values
MSE	0.9678
MAE	0.4322
R ²	0.8686
MAPE	6%

The evaluation metrics demonstrate that the model excels in predicting traffic congestion with impressive performance. The MSE of 0.9678 indicates that the model generates predictions with minimal squared errors, reflecting a high degree of accuracy in capturing the nuances of traffic patterns. The MAE of 0.4322 underscores the model's strong predictive capability, with average deviations being relatively low and manageable. The R² of 0.8686 reveals that the model accounts for approximately 87% of the variability in traffic congestion,

showcasing its effectiveness in explaining the observed data. Additionally, the MAPE of 6% demonstrates that the model's predictions are, on average, within 6% of the actual values, highlighting its robustness and reliability. Overall, these results confirm that the model delivers highly accurate and reliable predictions for traffic congestion, marking it as an exceptional tool for traffic forecasting. Fig. 20 illustrates the predicted and actual values of the suggested attention based multilayer GRU model for traffic congestion prediction.

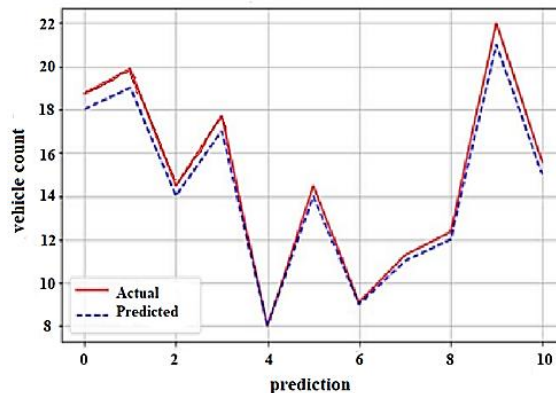


Fig. 20. Predicted vs. actual values of proposed model.

The attention-based multilayer GRU model showed considerable efficiency in predicting traffic congestion. The model demonstrated robust prediction accuracy and reliability, evidenced by an MSE of 0.9678, an MAE of 0.4322, and an R² value of 0.8686. A MAPE of 6% further demonstrates the model's resilience while processing real traffic data. The

integration of the attention mechanism within the multilayer GRU architecture enables the model to concentrate on the most pertinent temporal patterns and features in traffic data, enhancing prediction accuracy and ensuring effective capture of both short-term fluctuations and long-term dependencies in congestion patterns. The attention-based multilayer GRU

model is an exceptionally excellent method for predicting traffic congestion. An attention mechanism in a multilayer GRU aids in the prediction of traffic congestion by allowing the model to concentrate on the most important features and pertinent temporal patterns in the data. The attention mechanism, in contrast to conventional GRU models, gives significant time steps, ensuring that significant congestion-related occurrences or patterns are given priority during prediction. This enhances the model's ability to capture long-term dependencies while reducing the influence of irrelevant or noisy inputs. The multilayer GRU utilizes the attention mechanism to enhance prediction accuracy and interpretability, making it especially suitable for the intricate and variable nature of traffic congestion prediction.

V. CONCLUSION

The proposed attention-based multilayer GRU model offers a substantial improvement over traditional traffic management methods by addressing the dynamic and complex nature of traffic congestion. The model's ability to capture temporal dependencies and intricate traffic patterns through attention mechanisms enables more accurate predictions of traffic conditions and congestion levels. The model achieved a notable improvement in accuracy, with MAE and MSE values of 0.4322 and 0.9678, respectively. This enhanced predictive capability facilitates timely and efficient traffic management interventions, reducing travel times, minimizing fuel consumption, and lowering emissions. The effectiveness of the model in various urban scenarios demonstrates its potential to significantly improve overall traffic flow and urban mobility. Future research could explore integrating real-time data sources and extending the model's application to different traffic management systems to further enhance its effectiveness. Future studies should investigate transfer learning to adapt models for areas with scarce historical data and reduce dependence on computational resources. More thorough and useful solutions for traffic management systems will be ensured by integrating external factors like weather, road construction, and special events, as well as by creating reliable models for unexpected disruptions.

ACKNOWLEDGMENT

I would like to express my sincere gratitude to all those who contributed to the completion of this research paper. I extend my heartfelt thanks to my supervisor, my family, my colleagues and fellow researchers for their encouragement and understanding during the demanding phases of this work.

REFERENCES

- [1] Akhtar, M., & Moridpour, S. (2021). A review of traffic congestion prediction using artificial intelligence. *Journal of Advanced Transportation*, 2021(1), 8878011.
- [2] Ranjan, N., Bhandari, S., Zhao, H. P., Kim, H., & Khan, P. (2020). City-wide traffic congestion prediction based on CNN, LSTM and transpose CNN. *Ieee Access*, 8, 81606-81620.
- [3] Li, T., Ni, A., Zhang, C., Xiao, G., & Gao, L. (2020). Short - term traffic congestion prediction with Conv-BiLSTM considering spatio - temporal features. *IET Intelligent Transport Systems*, 14(14), 1978-1986.
- [4] Li, T., Ni, A., Zhang, C., Xiao, G., & Gao, L. (2020). Short - term traffic congestion prediction with Conv-BiLSTM considering spatio - temporal features. *IET Intelligent Transport Systems*, 14(14), 1978-1986.
- [5] Gollapalli, M., Musleh, D., Ibrahim, N., Khan, M. A., Abbas, S., Atta, A., ... & Omer, A. (2022). A Neuro-Fuzzy Approach to Road Traffic Congestion Prediction. *Computers, Materials & Continua*, 73(1).
- [6] Li, L., Lin, H., Wan, J., Ma, Z., & Wang, H. (2020). MF-TCPV: a machine learning and fuzzy comprehensive evaluation-based framework for traffic congestion prediction and visualization. *IEEE Access*, 8, 227113-227125.
- [7] Li, L., Dai, F., Huang, B., Wang, S., Dou, W., & Fu, X. (2024). AST3DRNet: Attention-Based Spatio-Temporal 3D Residual Neural Networks for Traffic Congestion Prediction. *Sensors*, 24(4), 1261.
- [8] Tsalikidis, N., Mystakidis, A., Koukaras, P., Ivaškevičius, M., Morkūnaitė, L., Ioannidis, D., ... & Tzouvaras, D. (2024). Urban traffic congestion prediction: a multi-step approach utilizing sensor data and weather information. *Smart Cities*, 7(1), 233-253.
- [9] Jiang, W., Han, J., Liu, H., Tao, T., Tan, N., & Xiong, H. (2024, August). Interpretable cascading mixture-of-experts for urban traffic congestion prediction. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 5206-5217).
- [10] Hao, M. J., & Hsieh, B. Y. (2024). Greenshields Model-Based Fuzzy System for Predicting Traffic Congestion on Highways. *IEEE Access*.
- [11] Zhang, T., Wang, J., Pang, T., Pang, Y., Wang, P., & Wang, W. (2024). A deep marked graph process model for citywide traffic congestion forecasting. *Computer - Aided Civil and Infrastructure Engineering*, 39(8), 1180-1196.
- [12] S Jasim, M., Zaghden, N., & Salim Bouhleb, M. (2024). Improving Detection and Prediction of Traffic Congestion in VANETs: An Examination of Machine Learning. *International Journal of Computing and Digital Systems*, 15(1), 947-960.
- [13] Arabiat, A., & Altayeb, M. (2024). Assessing the effectiveness of data mining tools in classifying and predicting road traffic congestion. *Indonesian Journal of Electrical Engineering and Computer Science*, 34(2), 1295-1303.
- [14] Chahal, A., Gulia, P., Gill, N. S., & Priyadarshini, I. (2023). A hybrid univariate traffic congestion prediction model for IOT-enabled smart city. *Information*, 14(5), 268.
- [15] Jin, G., Liu, L., Li, F., & Huang, J. (2023, June). Spatio-temporal graph neural point process for traffic congestion event prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 12, pp. 14268-14276).
- [16] Pan, Z., Sharma, A., Hu, J. Y. C., Liu, Z., Li, A., Liu, H., ... & Geng, T. (2023, June). Ising-traffic: Using ising machine learning to predict traffic congestion under uncertainty. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 8, pp. 9354-9363).
- [17] Zhang, T., Xu, J., Cong, S., Qu, C., & Zhao, W. (2023). A hybrid method of traffic congestion prediction and control. *IEEE Access*, 11, 36471-36491.
- [18] Wang, X., Zeng, R., Zou, F., Liao, L., & Huang, F. (2023). STTF: An efficient transformer model for traffic congestion prediction. *International Journal of Computational Intelligence Systems*, 16(1), 2.
- [19] <https://www.kaggle.com/datasets/fedesoriano/traffic-prediction-dataset>.
- [20] <https://www.kaggle.com/datasets/alioraji/weather-data-nov-2015>
- [21] Jawad, Y. K., & Nitulescu, M. (2024). Improving Driving Style in Connected Vehicles via Predicting Road Surface, Traffic, and Driving Style. *Applied Sciences*, 14(9), 3905.
- [22] Liu, Y., Lyu, C., Liu, X., & Liu, Z. (2020). Automatic feature engineering for bus passenger flow prediction based on modular convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, 22(4), 2349-2358.
- [23] Mahjoub, S., Chrifi-Alaoui, L., Marhic, B., & Delahoche, L. (2022). Predicting energy consumption using LSTM, multi-layer GRU and drop-GRU neural networks. *Sensors*, 22(11), 4062.
- [24] Zou, X., Zhao, J., Zhao, D., Sun, B., He, Y., & Fuentes, S. (2021). Air quality prediction based on a spatiotemporal attention mechanism. *Mobile Information Systems*, 2021(1), 6630944.

Robust Joint Detection of Coronary Artery Plaque and Stenosis in Angiography Using Enhanced DCNN-GAN

M. Jayasree^{1,*}, L. Koteswara Rao²

Department of ECE, Koneru Lakshmaiah Education Foundation, Aziz Nagar, Hyderabad, 500075, Telangana, India^{1,2}

Abstract—Timely detection and diagnosis of coronary artery segment plaque and stenosis in X-ray angiography is of great significance, however, the image quality variation, noise, and artifacts in the original image cause definitive difficulties to the current algorithms. These problems pose a challenge to meaningful analysis via traditional approaches, which compromises the efficiency of detection algorithms. To overcome these drawbacks, the current study presents a new integrated deep learning technique that integrates Deep Convolutional Neural Network (DCNN) with Generative Adversarial Network (GAN) in dual conditional detection. Detailed feature learning extracted from X-ray angiography images are performed through DCNN where it considers vascular structure and automatic pathologic regions detection. The use of GANs is to further enrich the dataset with synthetic images, distortions, and visual noise, which will make the model more immune to various conditions of images. Both approaches combined help in better classification of normal and pathological areas and less sensitiveness to quality of the obtained images. The proposed method therefore has shown an improvement of the diagnostic accuracy as a solid foundation for clinical decision making in cardiovascular systems. The efficacy of the suggested approach has been demonstrated by the following evaluation metrics: 97.9% F1 score, 98.7% accuracy, 98.2% precision, and 98% recall. The results prove higher sensitivity and accuracy of the plaque and stenosis identification comparing to the traditional methods, which confirms the efficiency of using the proposed DCNN-GAN method for considering the real-world fluctuations in the medical imaging. It reveals a decisive advancement in the ability to use algorithms for cardiovascular assessment by providing better results in difficult imaging environments.

Keywords—DCNN-GAN; angiography; coronary artery plaque; stenosis; joint conditional detection

I. INTRODUCTION

Correct detection of plaque formation together with stenosis severity enables the prevention of major cardiovascular health risks which lead to heart attacks and strokes. Timely interventions made possible by early diagnosis help minimize morbidity and mortality rates together with enhancing total patient outcomes. The accumulation of fat deposits in the arteries and the ensuing restriction of these essential blood channels are recognized as coronary artery plaque and stenosis, which are most important issues with cardiovascular fitness [1]. Plaque development starts a complex chain of activities inside artery partitions and is commonly made from ldl cholesterol, cell particles, calcium, and fibrin. These deposits may eventually

solidify and impede blood flow, which might reduce the amount of oxygen-wealthy blood that reaches the coronary heart muscle [2]. Simultaneously, this blockage is made worse by means of stenosis, or the narrowing of the arteries, which is regularly brought on via plaque build-up or the thickening of the artery partitions. The important results of those disorders for the health and properly-being of patients are highlighted via the huge upward thrust within the chance of cardiovascular occasions, together with heart attacks and strokes. X-ray angiography, which gives specific renderings of arterial structures, is one of the clinical imaging modalities this is most regularly utilized inside the clinical diagnosis of coronary artery plaque and stenosis [3]. Timely treatments and individualized treatment techniques are established on early diagnosis and specific evaluation of plaque load and stenosis severity. In order to improve diagnostic accuracy and prognostic capacities, modern research efforts are focused on building state-of-the-art imaging algorithms and computer models, inclusive of DL techniques and Bayesian frameworks [4]. Clinicians may additionally better manage and decrease the dangers related with plaque formation and stenosis with the aid of the use of these novel strategies to better understand and pick out coronary artery sickness, with a purpose to subsequently improve patient results and great of existence [5].

X-ray angiography image offer scientific personnel with a complete view of the artery machine, they may be vital for the identification and therapy of cardiovascular ailments [6]. With the usage of X-rays, this imaging technique highlights the anatomy and operation of blood arteries with the aid of taking real-time photos of them after a contrast agent injection [7]. X-ray angiography helps come across anomalies such as blockages, constriction, and aneurysms that would impair blood circulation to essential organs, mainly the coronary heart and brain, by way of carefully mapping the direction of blood float through arteries and veins [8]. Because those images can precisely pick out the location and diploma of artery blockages, they are helpful in helping to guide interventional treatments along with angioplasty and stent implantation. Moreover, X-ray angiography facilitates medical doctors make properly-knowledgeable judgments on endured affected person care by using allowing them to music the route of the ailment and compare the effectiveness of treatments [9]. The endured significance of X-ray angiography in modern medicine is proven via the truth that, despite the development of non-invasive imaging technology consisting of CT and MRI, this take a look at continues to be important for the prompt and particular

*Corresponding Author: M. Jayasree

detection of acute cardiovascular crises and complex vascular issues.

Image analysis has gone through a revolution due to the fact that, the Deep Convolutional Neural Network, which automatically develop hierarchical representations from uncooked pixel records. Because this magnificence of neural networks is so desirable at figuring out small info and patterns in pictures, it is especially beneficial for tasks like segmentation, category, and object popularity [10]. Multiple convolutional filter layers are used by DCNNs to methodically extract regularly summary records from input photographs. Every layer has the ability to understand wonderful patterns, consisting of edges, textures, and complex systems, which enables the community discover minute versions that are essential for precise image interpretation. The generator and discriminator neural networks in an aggressive game framework make up a Generative Adversarial Network. While the discriminator learns to differentiate among produced and real facts, the generator learns to create synthetic data (together with pix and sounds) that mimic actual samples from a training dataset. GANs reach a pleasant stability thru iterative education: the discriminator becomes better at figuring out authenticity, at the same time as the generator receives more sensible output [11]. The outputs produced by means of this adverse learning process are highly sensible and can be unsuitable for real information. GANs have located programs in a wide range of disciplines, which includes pc vision, herbal language processing, and biomedical imaging. They have revolutionized jobs like picture synthesis, statistics augmentation, and anomaly detection [12].

The selection of DCNN-GAN model occurred because its image quality handling capabilities together with improved diagnostic verification and enhanced feature extraction satisfied the project requirements. The model's generation abilities increases dataset size while making the program resistant to artifacts and noise thus making it appropriate for plaque and stenosis detection in coronary arteries. The suggested study demonstrates a way to higher pick out coronary artery stenosis and plaque in X-ray angiography images by using utilising the complementing traits of GAN and DCNN. An energy of DCNNs is complicated function extraction from clinical snap shots, that's crucial for accurately identifying unwell illnesses. The DCNN element, skilled on annotated datasets, correctly identifies areas as both regular or suggestive of plaque and stenosis. In addition, GANs enhance the dataset by means of producing synthetic pictures that intently resemble real angiography scans. This improves the schooling information and makes the DCNN more resilient to changes in ailment presentation and photo excellent. Moreover, anomaly identity is made viable by way of the hostile education of GANs, which may also display subtle signs and symptoms of contamination development which might be neglected by way of conventional diagnostic strategies. The challenge intends to improve early intervention techniques, enhance diagnostic accuracy, and subsequently resource in more efficient medical selection-making in cardiovascular care by way of combining many technologies right into a single pipeline. Some of the important contributions of the proposed look at are:

- This method improves detection of the coronary artery narrowing and plaque by combining the GANs with DCNNs.
- The synthetic images created and added into the set by the GANs increase the range of inputs to be expected the DCNN model, for various patients' conditions and image differences.
- The diagnosis is made early and is accurate, two factors that help in early intervention by physicians and positive patient wellness.
- The method enhances the ability of medical practitioners to read X-ray angiography images thus enhancing diagnosis in clinical practices.
- The approach incorporated deep learning to solve difficult detection issues and improves medical imaging of the procedures hence improving clinical decision making.

The suggested work's section is organized as follows: Section III has the problem statement, while Section II contains associated initiatives. The methodology of the article is covered in Section IV, along with the suggested work, pre-processing, and execution. Section V presents the findings and discussion, while Section VI offers suggestions for more research.

II. RELATED WORKS

Rodrigues et al. [13] suggests using a two-step DL system to identify stenosis in X-ray coronary angiography pictures in a largely automated manner. The approach uses two separate convolutional neural network architectures to automatically detect and classify the angle of view and calculate the boundaries of the areas of interest in frames when stenosis is evident. To improve the system's performance, approaches including data augmentation and transfer learning are applied. The findings indicate that the LCA and RCA had 0.97 accuracy and 0.68/0.73 recall, respectively, in categorizing the LCA/RCA angle view and regions of interest. These results pave the way for an entirely robotic approach for determining the degree of stenosis using X-ray angiographies, and they compare favorably with earlier results achieved using analogous methodologies.

Ovalle-Magallanes et al. [14] offers a novel technique that uses transfer learning to automatically detect coronary artery stenosis in XCA images by employing a CNN that has already been trained. A common heart condition called coronary artery disease is brought on by abnormal construction of the coronary arteries. It ranks among the leading causes of death globally. XCA is the most standard imaging technique for diagnosing stenosis. The technique selects the best cut and fine-tuned layers using a network-cut and fine-tuning methodology after 20 alternative configurations. Three methodologies (actual data, purely fake data, and artificial and real data) were used to fine-tune the networks. The 10,000 photos in the synthetic dataset were created using a generative model. The findings demonstrated that pre-trained CNNs such as VGG16, ResNet50, and Inception-v3 performed better in stenosis identification than referencing CNNs.

Pang et al. [15] suggests using an object detection network-based technique called Stenosis-DetNet to automatically identify coronary artery stenosis in X-ray images. To optimize temporal information and produce precise detection results, the approach makes use of an order consistency alignment module and a series feature fusion module. The sequence feature fusion module merges all candidate box features, whilst the order consistency alignment module enhances preliminary results by merging a coronary artery displacement information and image characteristics of surrounding pictures. 166 X-ray picture sequences were utilized in the experiment for testing and training. Stenosis-DetNet outperformed the other three approaches in terms of precision and sensitivity, coming up at 94.87% and 82.22% higher, respectively. The suggested approach outperformed the approaches in suppressing false positive and false negative findings of stenosis identification in sequence angiography pictures.

Gil-Rios et al. [16] provides a strategy that overcomes several classification approaches and DL methodologies to automatically detect myocardial stenosis in X-ray coronary pictures. The approach selects features using the Univariate analysis Marginal Distribution Algorithm and compares metaheuristics statistically to investigate the computational cost of the search space. The approach is evaluated on two an X-ray image dataset containing coronary angiograms and compared with six other approaches currently in use. It is appropriate for clinical usage based on the accuracy rate of 0.89 and 0.88, the Jaccard Index of 0.80 and 0.79, and the average computing time of about 0.02 seconds, as demonstrated by the findings. The precision and Jaccard Index assessment metrics are used to assess the efficacy of the procedure.

Stralen et al. [17] recommended a study on coronary artery stenosis (CAD), a serious global health issue for which automatic diagnosis of the condition on X-ray images is essential. Atherosclerotic plaques and stenosis are the disease's causes; these conditions increase the workload of the heart and raise the risk of heart failure. In clinical practice, automated stenosis detection might be utilized as a second reader or for triage purposes. Deep neural networks are used to assess whether stenosis can be detected in X-ray coronary angiography pictures. Employing clinical angiography data from 438 patients, three potential object detectors were trained and evaluated. EfficientDet demonstrated a mean average accuracy of 0.67 in stenosis detection, supporting the notion that attention processes enhance convolutional neural networks' capabilities for medical imaging.

Ovalle-Magallanes et al. [18] enhances the identification of stenosis in X-ray coronary angioplasty using quantum computing. A quantum network is used to improve the performance of a classical network that has already been trained in a hybrid transfer-learning paradigm. Normalization features undergo processing in the quantum network after the classical data have been processed afterwards into a hypersphere utilizing a hyperbolic tangent function. A SoftMax function is used to obtain class probabilities. The data is divided into several circuits inside the quantum network using a distributed variational quantum circuit, which speeds up training without sacrificing detection performance. A small dataset of 250 image

patches from X-ray coronary angiography is used to assess the procedure. In terms of accuracy, recall, and -score, the hybrid classical-quantum network fared much better than the classical network, attaining 91.8033%, 94.9153%, and 91.8033%, respectively.

Han et al. [19] said that the diagnosis and treatment of coronary artery disease depend on the ability to recognize coronary artery stenosis in XRA images. Unfortunately, most methods suffer from poor spatiotemporal-temporal information use. To gather spatiotemporal features at the suggested level for an innovative stenosis detection method, a transformer-based component is provided. The proposal-shifted spatio-temporal tokenization approach gathers region-of-interest characteristics that the Transformer-based feature aggregation network utilizes to acquire knowledge a faraway spatio-temporal context for final constriction prediction, hence enhancing the ROI features. A remarkable score of 90.88% was attained, outperforming the results of 15 other detection techniques, as examinations on 233 XRA sequences, both qualitative and quantitative, validated the approach's effectiveness. This illustrates how well the technique can detect stenosis from XRA pictures.

Algarni et al. [20] suggested an ASCARIS model that conducts classification using the Attention-based Nested U-Net, optimizes maximum principal curvature to improve contrast, and eliminates noise pixels using a modified wiener filter. To improve segmentation accuracy, angle estimation is applied. Classifying X-ray pictures into normal and pathological classes is accomplished by extracting double characteristics from the segmented image using an architecture based on VGG-16. Using the simulation tool MATLAB R2020a, the model's performance was assessed and compared with previous methods in terms of segmentation accuracy, PSNR, Hausdorff distance, revised contrast to noise ratio, accuracy, sensitivity, specificity, mean square error, dice coefficient, Jaccard similarity, and ROC curve. The findings demonstrate that the suggested model works better than current methods, resulting in an optimum categorization of CAD. The technique enhances vascular anatomy and eliminates background artifacts.

To partially automate the process of detecting stenosis from X-ray coronary angiography pictures, a DL system has been presented. The system recognizes and categorizes the angle of view and areas of interest using two different convolutional neural network designs. To improve the system's performance, approaches including data augmentation and transfer learning are applied. The findings indicate that the LCA and RCA had 0.97 accuracy and 0.68/0.73 recall, respectively, in categorizing the LCA/RCA angle view and regions of interest. The technique selects the best cut and fine-tuned layers using a network-cut and fine-tuning methodology after 20 alternative configurations. For automated identification of coronary artery stenosis on X-ray images, the object detection network-based Stenosis-DetNet approach is suggested. The method uses a hybrid transfer-learning paradigm and quantum computing to improve stenosis detection in X-ray coronary angiography. The previous research had problems with the changes in angiography image quality is one of its drawbacks. The performance of the classifier may be impacted by irregular image quality, noise, or artifacts, which might impair the efficiency of the pre-processing and feature extraction processes.

III. PROBLEM STATEMENT

The application of XCA visuals to study arterial stenosis continues to provide significant challenges, partly because of shortcomings found in the currently used approaches. The primary problem is the potentially poor quality of the images, which may contain imperfections and noise that improve preprocessing and feature extraction functions while degrading the effectiveness of the classifier [13] [20]. Furthermore, the current methods, such as the one suggested by Han et al. [19], do not maximize the utilization of dynamical data, leading to suboptimal stenosis identification. Another significant problem is the failure to reliably classify both the LCA/RCA, as demonstrated through additional investigations, with consistent recall frequencies. Furthermore, even though quantum technology is used in combined transfer-learning and other DL-related applications. Additionally, despite the seeming promise of recent developments in DL, including a combined transfer-learning system using quantum technology [18], these technologies may be limited by the small datasets and computational demands, which may harm their application in practical. The lack of reliable methods to repeatedly manage the picture quality and patients' anatomical differences is a further significant problem, which adds to the complexity of accurately evaluating stenosis in various populations [17]. Current models face difficulties because they fail to handle variable image quality while dealing with inconsistent accuracy between different populations and demanding heavy computational processing. The inadequate management of noise and insufficient control of artifacts alongside limited dataset availability produce unreliable and inconsistent accuracy. In real-world medical settings the deployment of these methods

remains impractical because of challenges related to interpretability, manual preprocessing requirements and their limited adaptability. Because of the shortcomings of the previous method, new intelligent and flexible algorithms would be needed for the automated identification of stenosis in coronary arteries utilizing XCA pictures.

IV. JOINT CONDITIONAL DETECTION OF CORONARY ARTERY PLAQUE AND STENOSIS USING DCNN-GAN

Early detection and diagnosis of cardiovascular illness is crucial for the identification of stenosis and coronary artery plaque in X-ray angiography pictures. Conventional techniques frequently encounter image quality changes, including noise and artifacts, which might impair the detection algorithms' accuracy. To address these issues, providing a unique technique in this paper that combines GAN with DCNN. GAN are used in the augmentation of the dataset and DCNNs are used to extract features from X-ray angiography images, which enables in-depth examination of the vascular architecture and any anomalies that might be signs of stenosis and coronary artery plaque. Creating artificial images that mimic varying degrees of noise and visual artifacts. This improves the model's resilience to a range of image circumstances and speeds up the training process. With the help of the integrated DCNN-GAN architecture, joint conditional detection is made easier. In this process, the system uses learnt characteristics to classify regions of interest as either normal or pathological. Reducing the influence of changes in image quality on detection performance, this method seeks to greatly improve diagnostic accuracy and reliability and open the door to more efficient clinical decision-making in cardiovascular medicine.

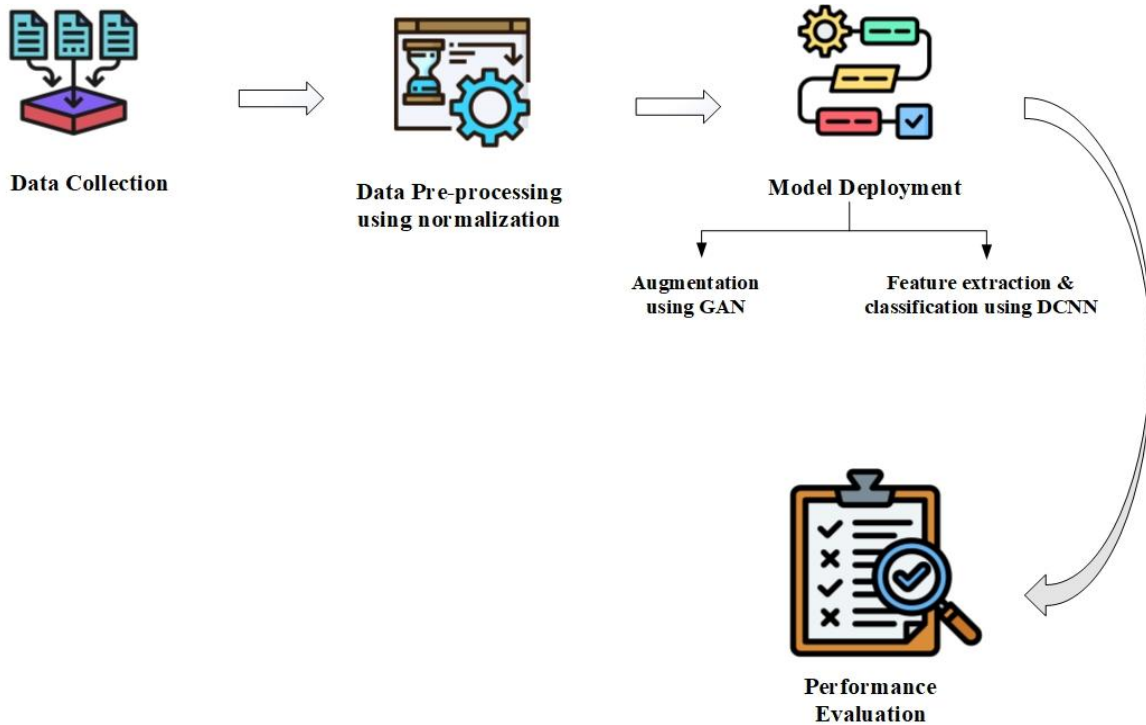


Fig. 1. Flow diagram of the proposed study.

The procedure of utilizing the DCNN-GAN approach to identify coronary artery plaque and stenosis in X-ray angiography images is shown in the Fig. 1. The first step in the procedure is data collection, which involves compiling an extensive collection of varied and high-quality angiogram images. The images are then enhanced and standardized for training using normalization during the Data Pre-processing stage. The core of the technique is DCNN-GAN, which combines Deep DCNN for in-depth extraction and classification of features with GAN for artificial picture generation to enrich the dataset and mimic various visual circumstances, hence boosting model resilience. Finally, performance evaluation to validate its effectiveness and reliability. The cycle process ensures continuous model improvement and robustness in real-world clinical scenarios while addressing the drawbacks of earlier methods for handling variations in picture quality.

A. Data Collection

The dataset Coronal Slices shows the MRI images that depict coronal slices taken from the torso's successive anteroposterior locations. A collection of 5000 X-ray angiography images was obtained from multiple clinical environments to deliver diverse patient population demographics for enhancing model training effectiveness and generalization ability. With its ability to provide a thorough vision of blood vessels and any anomalies inside the coronary arteries, X-ray angiography images are essential diagnostic tools in cardiovascular medicine. These images are produced by way of injecting a substance that contrasts into the bloodstream and acquiring X-rays whilst the agent flows through the heart and coronary arteries. X-ray angiography, that is commonly used to pick out illnesses such as coronary artery sickness, offers scientific experts high-decision images in actual time that display regions of stenosis (narrowing of the arteries), plaque formation, and other cardiovascular problems. When it comes to correctly interpreting those images and making selections about remedy, such as implanting stents or present process pass surgical procedure, their excellent and readability are crucial. In order to reduce radiation exposure and maximize photograph decision, present day X-ray angiography structures use modern generation, ensuring affected person safety and powerful prognosis. The goal of studies using X-ray angiography datasets is to create automatic strategies to improve detection sensitivity and performance. These techniques, which consist of DL algorithms like DCNN-GAN, will assist tailored treatment plans and early analysis in cardiovascular care [21].

B. Data Pre-Processing

The technique of converting unprocessed information right into a clear and on hand shape for research is referred to as statistics guidance. Making sure the data is prepared for ML designs, involves actions including resolving values that are missing, casting off outliers, normalizing or standardizing the records, and encoding specific variables.

1) *Normalization*: Data normalization is the technique of changing information in order that analytics and DL algorithms might also use it in a constant manner. It is typically used to convert raw statistics right into a layout higher suitable for DL techniques including logistic regression, neural networks, and linear regression. Normalizing facts in DL can help with data

simplicity for records this is numerical as well as express. The Min-Max Scaling technique converts statistics into a range among 0 and 1 by way of dividing the information by the difference between the best and least values, after which subtracting the minimum fee from each information factor. The normalization approach is useful in preventing the exceptions from severely affecting the data while working with exceptions.

$$m_o = \frac{m - m_{min}}{m_{max} - m_{min}} \quad (1)$$

In Eq. (1), m_{max} represents a feature's highest value, m_{min} its minimum value, and m_o its normalization value.

2) *Noise removal*: The process of reducing noise in images involves using filters in comparison to eliminate unwanted stochastic changes, or noise, which obscures key elements. A filter called Gaussian blur, which is a jointly-existing approach, aims to apply a function called Gaussian in addition to averaging pixels inside a certain area (to reduce noise). The following Eq. (2) defines the Gaussian blur [22]:

$$G(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (2)$$

Whereas the Gaussian function is represented as $G(x, y)$, the pixel coordinates are represented by x and y , while the standard variation σ explains the level of uniformity. The filter's job is to reveal, by convolution, the Gaussian kernel with the image. This minimizes the high-frequency elements, or noise, while maintaining almost all of the boundaries.

C. GAN

Text, audio, and image data samples may be synthesized using generative models called GANs. Combining training a generator and discriminator neural networks at the same time is the basic idea behind GANs. While the discriminator learns how to discern between actual and phony data, the generator learns how to create synthetic data samples. Through adversarial training, where the generator tries to fool the ML algorithm and the discriminator tries to discern between genuine and fake samples, GANs are trained to generate realistic, high-quality data.

While the discriminator D seeks to discern between actual samples (from the true data distribution $P_{data}(x)$ and fraudulent samples created by G , the generator G seeks to make realistic data samples from random noise z (taken from a previous distribution $p_z(z)$). With θ_g as the generator's parameters, the generator network learns to map the input noise, z , to the data space, $G(z; \theta_g)$. Conversely, an input x is mapped by the discriminator, whose parameters are θ_d , to a probability $D(x; \theta_d)$ that x is a genuine sample.

A GAN must optimize two loss functions during training: one for the generator and one for the discriminator. While the generator is taught to trick the discriminator into believing bogus samples to be real, the discriminator is trained to optimize the chance of accurately categorizing actual and fake samples. The value function $V(G, D)$ may be used to structure this as a minimax game.

$$\min_G \max_D V(G, D) =$$

$$E_{x \sim P_{data}(x)}[\log D(x)] + E_{z \sim P_z(z)}[\log(1 - D(G(z)))] \quad (3)$$

In Eq. (3), D is stimulated to produce high probability for genuine samples by $E_{x \sim P_{data}(x)}[\log D(x)]$, which stands for the expected value of the discriminator's output logarithm for real data. On the other hand, $E_{z \sim P_z(z)}[\log(1 - D(G(z)))]$ stands for the anticipated value of the logarithm of one less the output of the discriminator for fictitious data, which incentivizes D to produce low probability for fictitious samples produced by G .

D and G are updated in turn throughout training. The discriminator is enhanced to more accurately distinguish true

from false data, while the generator is updated to offer accurate data that can trick the discriminator. In an ideal scenario, this adversarial process results in the generator generating extremely realistic samples that are identical to genuine data over time, reaching a Nash equilibrium where neither the discriminator nor the generator can operate better without altering the other. Because GANs can learn complicated data distributions and produce high-quality synthetic data, they are frequently employed in many different applications, such as image production, style transfer, and data augmentation.

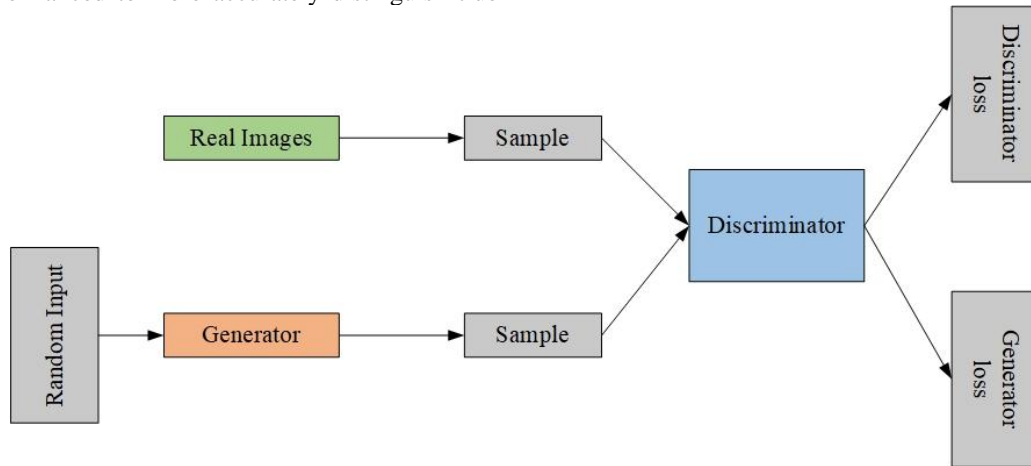


Fig. 2. Architecture of GAN.

The design and operation of a GAN are depicted in the Fig. 2. The generator and discriminator are its two primary parts. The generator, which converts random noise into a created image, is the first step in the process. Using the training dataset, this generator network learns to produce images that look authentic. The discriminator is then given both the produced and actual images from the dataset. It is the discriminator's job to determine if these images are authentic or not. It produces a likelihood that indicates if each image is artificially created or real. During training, both the generator and the discriminator are in competition with one another: the discriminator wants to distinguish between real and fake pictures with accuracy, while the generator wants to produce images that seem exactly like real ones to trick the discriminator. The discriminator sharpens its capacity to spot phony images, while the generator refines its output to trick the discriminator even more. Until the generator generates very realistic images that the discriminator can no longer accurately separate from real images, adversarial training will be conducted. Because of this dynamic process, GANs are able to produce practical statistics, which makes them effective tools for quite a few programs like information augmentation, photo synthesis, and the creation of original content.

D. DCNN

The circle of relatives of DL models called DCNN has tested magnificent overall performance in photo and video recognition programs due to its ability to mechanically generate hierarchical feature representations from unprocessed input records. Common layers seen in a DCNN design are convolutional neural networks, pooling, and fully connected layers. Crucial to the process are the convolutional layers, which use filters (kernels)

to convolve across the input image and create feature maps that emphasize certain elements like edges or textures. The convolution procedure may be mathematically represented in Eq. (4),

$$(X * W)(i, j) = \sum_m \sum_n X(i + m, j + n)W(m, n) \quad (4)$$

Output feature map, given by input X and a filter W . Taking small portions of the input into consideration, this process captures spatial hierarchies. Activation functions like as ReLU, which are defined as in Eq. (5),

$$ReLU(x) = \max(0, x) \quad (5)$$

are applied after the convolutional layers to induce non-linearity.

Pooling layers, often max pooling, substantially minimize the spatial dimensions of the feature maps while achieving spatial invariance and minimizing computation costs. The down sampled output is represented by Y , and the max pooling operation may be expressed mathematically in Eq. (6),

$$Y(i, j) = \max\{X(p, q) | p, q \in \text{poolinhregion}\} \quad (6)$$

More intricate and abstract information are learnt as the network becomes deeper, leading to fully linked layers that combine these features to provide predictions. A SoftMax function for classification problems is produced by flattening and feeding the output of the last convolutional or pooling layer into one or more fully connected layers.

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_k e^{z_k}} \quad (7)$$

Eq. (7) is the formula for the SoftMax function, where z is the SoftMax layer's input vector and $\sigma(z)_j$ denotes the probability of the j -th class.

Backpropagation and optimization methods like stochastic gradient descent (SGD) are used during DCNN training in order to minimize a loss function, in this case, the cross-entropy loss for classification.

$$H(p, q) = -\sum_i q_i \log(p_i) \quad (8)$$

Eq. (8) is the definition of the cross-entropy loss for a true distribution q and a forecasted probability distribution p . Across a wide range of contemporary computer vision applications, from object identification and segmentation to facial recognition and autonomous driving, DCNN are essential for their automated and efficient extraction of pertinent information from high-dimensional input.

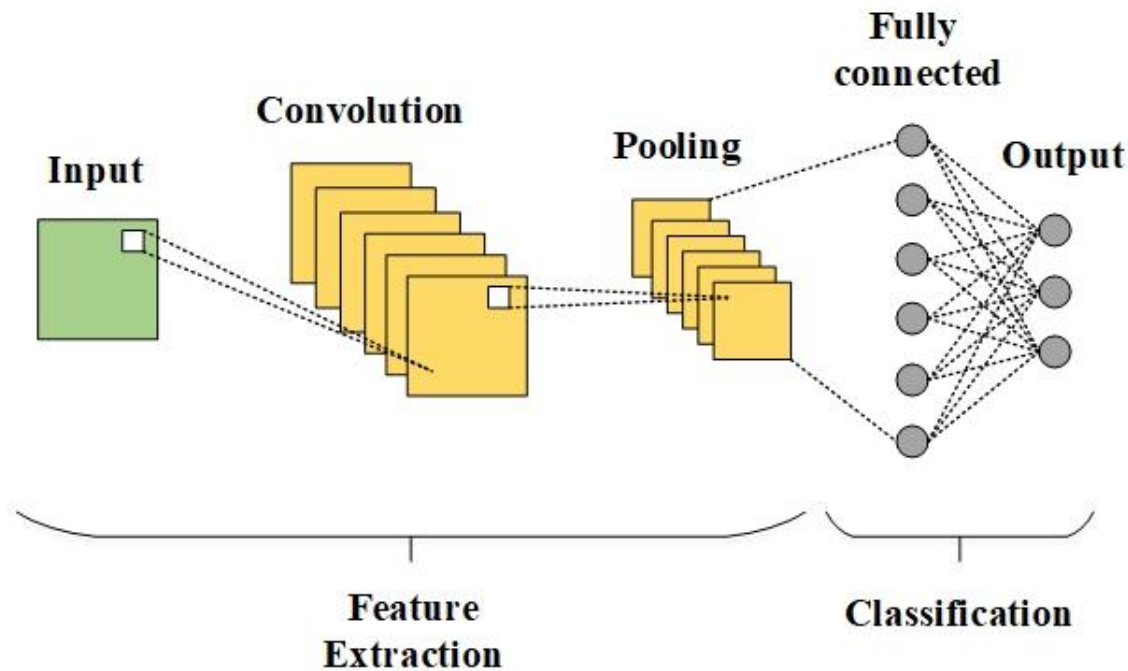


Fig. 3. Architecture of DCNN.

The two main phases of a DCNN architecture feature extraction and classification are shown in the Fig. 3. Several convolutional operations are performed on the input image during the feature extraction stage. These operations allow the filters to identify different features, such as edges and textures, and produce feature maps. After that, these feature maps are run via pooling layers, which achieve spatial invariance and lower computational burden by lowering the spatial dimensions of the feature maps by choosing the largest value from an area. During the category level, the final result of the remaining pooling layer is fed into fully linked layers after being compressed into a vector with one measurement. These layers combine the gathered features to generate the final categorization. The last layer typically makes use of a SoftMax activation function to generate an opportunity distribution over all possible lessons. DCNNs' design allows them to effectively apprehend and classify complex styles in the input records.

It is crucial for the activate detection and treatment of cardiovascular disorders to identify coronary artery plaque and stenosis in X-ray angiography images. Variabilities in picture excellent, which include noise and artifacts, pose serious hurdles to traditional tactics and may obstruct accurate detection. Suggesting a unique method that mixes GAN with DCNN to resolve this. Because of its skill in extracting complicated traits

from medical photos, DNN can identify diseased areas and examine vascular architecture in high-quality element. The DCNN successfully learns to categorize areas as regular or suggestive of plaque and stenosis through training on annotated datasets. In addition, GANs produce synthetic images that imitate actual angiography scans, including exclusive levels of noise and artifacts to decorate the education set. This ensures that the model is resistant to the many photo circumstances that get up in medical practice, further to improving the robustness of the DCNN. Joint conditional detection is supported by way of the integrated DCNN-GAN architecture, which complements diagnostic accuracy via permitting the gadget to regulate to and compare a variety of photo traits. Better affected person consequences are expected because of this strategy's primary upgrades to automatic cardiovascular diagnostics efficacy and reliability. Grad-CAM produces visual explanations which define fundamental areas that impact the model's prediction process. Visual interpretability enables healthcare professionals to both verify and trust the automated diagnosis process resulting in model acceptance for plaque and stenosis recognition tasks.

V. RESULT AND DISCUSSION

The outcomes of the proposed DCNN-GAN framework reveal widespread gains in the popularity of plaque and coronary

artery stenosis in X-ray angiography photographs. It has shown to a hit to mix GAN for statistics augmentation with DCNN for function extraction and classification. Using an NVIDIA RTX 3090 GPU allowed the proposed DCNN-GAN model to finish its training process within 12 hours. The model's optimized design achieves accelerated convergence while maintaining high diagnostic precision through more efficient operations thus allowing use in real-time clinical environments. The model can control image high-quality variations, consisting of noise and artifacts, that are frequently seen in scientific situations. As a result, the gadget has a huge potential to exactly discover and classify regions of interest, generating correct and dependable diagnostic consequences. These encouraging findings imply that the DCNN-GAN architecture can considerably enhance diagnostic overall performance, facilitating early cardiovascular sickness prognosis and intervention and improving patient care in popular in clinical exercise. The proposed work is implemented using python.

A. Experimental Outcome

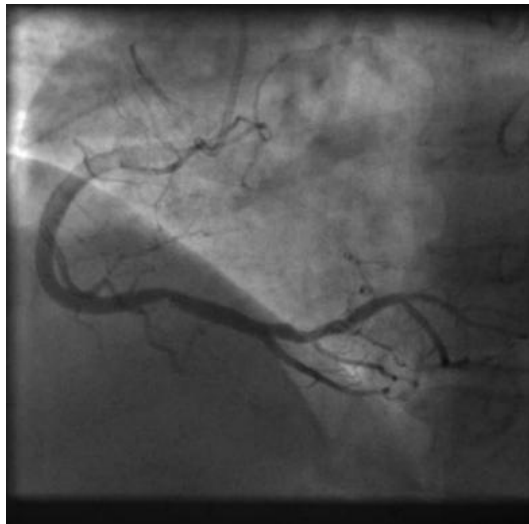


Fig. 4. Coronary angiographic image with visible stenosis.

The Fig. 4 demonstrates the coronary angiogram which is a medical examination of coronary arteries. In this particular image, one of the coronary arteries has a condition called stenosis which narrowed. This narrowing can reduce the blood supply to the heart muscle and cause discomfort such as chest pain.

TABLE I. PERFORMANCE EVALUATION

Category	Accuracy	Sensitivity (Recall)	Specificity	Precision
3-CAT Prediction	0.98	0.97	0.96	0.98
2-CAT Prediction	0.96	0.95	0.94	0.96
3R-CAT Prediction	0.94	0.93	0.92	0.94
2R-CAT Prediction	0.92	0.91	0.90	0.92
Random Guessing	0.50	0.50	0.50	0.50

The Table I summarizes the performance of the DCNN-GAN model in predicting coronary artery plaque and stenosis across different categories. The 3-CAT Prediction (3 categories) achieves the highest accuracy (98%), sensitivity (97%), and precision (98%). The 2-CAT Prediction (2 categories) also performs well with 96% accuracy and 95% sensitivity. The 3R-CAT and 2R-CAT Predictions, slightly reduced versions, show strong but lower performance. Random guessing, included for comparison, shows much lower metrics, prominence the model's effectiveness.

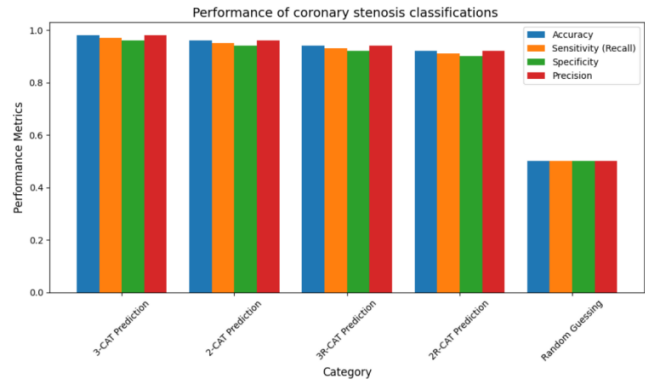


Fig. 5. Performance of coronary stenosis classifications.

The Fig. 5 shows how various models of classification of coronary stenosis performed. The x-axis shows the different models, and the y-axis represents the performance metrics: sensitivity also referred to as recall, specificity and precision. In general, the 3-CAT (3 categories) such as normal, early and advanced prediction model demonstrates the highest accuracy as compared to the other models and also, the highest sensitivity as well as specificity. But the precision rate of 2-CAT (2 categories) prediction model is the highest among all the models. Random guessing makes the worst performance manifestation across each specified parameter.

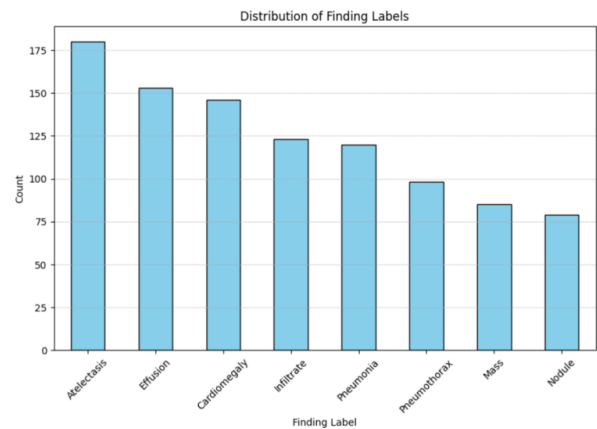


Fig. 6. Distribution of finding labels.

The Fig. 6 represents the frequency of the different finding labels in the dataset and the figure reveals finding labels such as "Atelectasis", "Edema", "Pleural effusion", "Pneumonia", "Consolidation", and "Nodules" in patients with the listed pathologic conditions but this says that "Consolidation", "Nodules" among the listed pathologic conditions are rarely found.

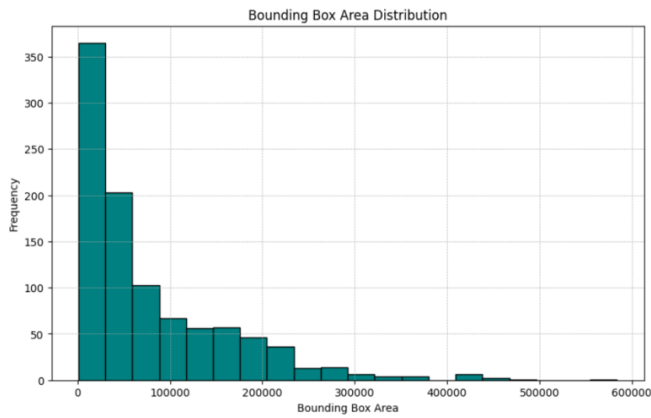


Fig. 7. Histogram.

The Fig. 7 indicates a histogram representing the distribution of bounding field areas. The x-axis suggests the location values, and the y-axis represents the frequency of occurrences. The histogram reveals that most bounding packing containers have regions between zero and 100,000 rectangular pixels, with some outliers having areas above 500,000 square pixels.

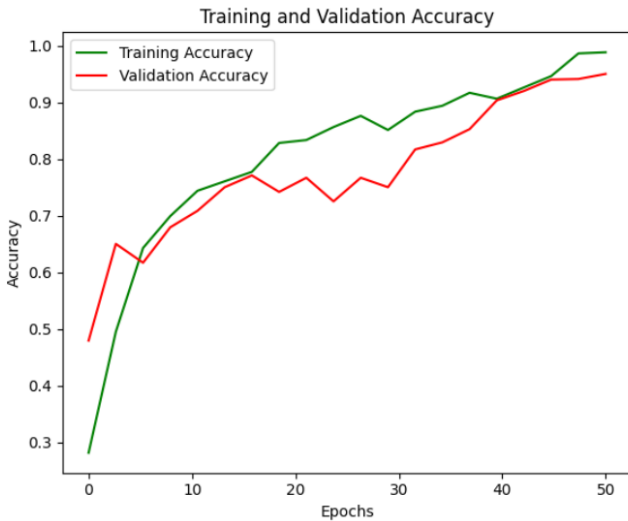


Fig. 8. Training and testing accuracy.

The Fig. 8 shows the Training and Validation Accuracy of the DCNN-GAN version's performance across one hundred epochs. Both accuracies rise quickly, suggesting that the model learned rapidly during the early epochs. Around the 20th epoch, schooling accuracy strategies one hundred%, demonstrating the model's right suit to the schooling set. Additionally, testing accuracy increases rapidly but exhibits slight fluctuations, suggesting a small amount of overfitting that eventually stabilizes. Both accuracies plateau and live comparatively regular after the twentieth epoch, with checking out accuracy carefully trailing accuracy, indicating robust generalization to formerly unknown facts. Effective regularization is seen by the small space between the 2 traces, which avoids considerable overfitting. This overall fashion suggests that the model may additionally acquire statistics and generalization from the preliminary dataset with great accuracy in each the validation and education tiers.



Fig. 9. Training and testing loss.

The Training and Testing Loss, shown in Fig. 9. Of the DCNN-GAN version's loss values across 60 epochs. Both losses start high at first, that's indicative of the version's early gaining knowledge of section. As a result of the version's adeptness at getting to know from the schooling information, the training loss drops off speedy, stabilizing around the 20th epoch and staying low. On the alternative hand, the testing loss has a greater complex pattern, first lowering after which experiencing a surge around the 20th epoch before stabilizing. This version indicates that the model had a few overfitting issues and made changes as training went on. The version correctly reduced errors on the training and testing datasets while, after the thirtieth epoch, both losses converge and hold low levels. When a version efficiently lowers mistakes and maintains stability, it is stated to have sturdy generalization overall performance and may produce accurate and dependable predictions whilst implemented to sparkling, untested data.

B. Performance Metrics

To assess a DL model's overall performance in figuring out artery plaques and constriction in X-ray angiography snap shots, essential factors to do not forget are precision, consider, precision, and F1 score. By calculating the percentage of correct high quality and accurate terrible forecasts among all forecasts, accuracy evaluates the version's ordinary correctness. Through expressing the ratio of proper positives to the overall of actual and fake positives, precision suggests how dependable the superb predictions are. Recall, that's often referred to as sensitivity, quantifies the degree to which a model is capable of as it should be pick out actual high-quality scenarios. The ratio of genuine positives to the entire of proper positives and false negatives is used to calculate it. The F1 score strikes a stability between the 2 variables to produce a single statistic that debts for false positives in each case and false negatives: the harmonic average of accuracy and recollect. Collectively, these metrics provide a comprehensive assessment of the version's efficacy, showcasing its accuracy and electricity in diagnosing cardiovascular conditions.

1) *Accuracy*: Accuracy is a measure of the way regularly a ML model predicts a result successfully. Accuracy can be calculated through dividing the entire number of estimates by means of the quantity of accurate forecasts.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

2) *Precision*: Precision is a metric that quantifies how regularly a ML model efficiently predicts the nice magnificence. The amount of particular superb forecasts (genuine positives) divided via the whole range of favorable predictions (false and actual fine) that the model properly anticipated can be used to calculate accuracy.

$$Precision = \frac{TP}{TP+FP} \quad (10)$$

3) *Recall*: Recall is a statistic used to describe how often a ML model correctly identifies positive instances, or real positives, out of all the genuine positive examples in the dataset. By dividing the total number of positive instances by the number of true positives, recall may be calculated. The latter includes both true positives (patients who are successfully found) and false negative results (missed cases).

$$Recall = \frac{TP}{TP+FN} \quad (11)$$

4) *F1 score*: The F1 score, often called the F-measure, is the harmonic mean of the accuracy and recall of a classification model. Because both measures have the same weight in the score, the F1 measure appropriately depicts the reliability of a model.

$$F1\ score = 2 * \frac{Precision \times Recall}{Precision + Recall} \quad (12)$$

In Eq. (9), Eq. (10), Eq. (11), and Eq. (12), TP and TN represent true positive and true negative. Whereas, FP and FN represent as False negative and False positive.

TABLE II. PERFORMANCE METRICS

Metrics	Efficiency
Accuracy	98.7%
Precision	98.2%
Recall	98%
F1 score	97.9%

The DCNN-GAN model exhibits remarkable performance metrics in Table II with respect to the detection of heart plaque and stenosis in X-ray angiography pictures. The model predicts outcomes with a 98.7% accuracy rate, which is quite good. With an accuracy of 98.2%, the version detects genuine positives with few false positives. Because the version is so exact at figuring out authentic positives, its excessive take into account charge of 98% ensures that just a few genuine positives are neglected. The model's potential to manipulate inaccurate consequences and false negatives by using balancing accuracy and don't forget is confirmed by way of its F1 rating of 97.9%. Together, those measures show the version's extremely good outcomes and its capacity to substantially boom diagnostic accuracy and

dependability for the identity of cardiovascular illness in scientific settings.

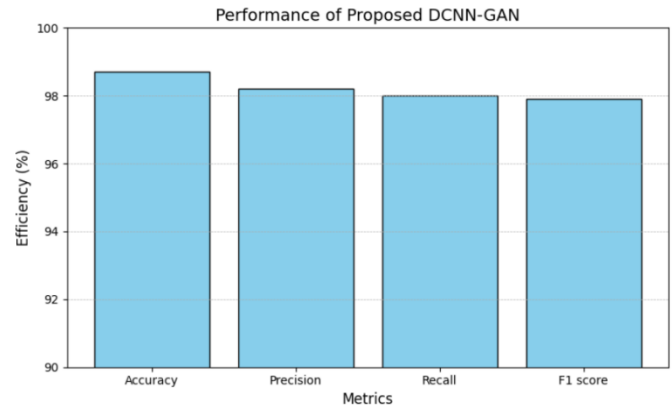


Fig. 10. Performance efficiency of the proposed model.

Fig. 10 shows the effectiveness metrics for a version that recognizes artery plaques and stenosis in X-ray angiography photographs. The 4 number one performance metrics displayed inside the graph are accuracy, precision, consider, and F1 score. The simulation is the maximum accurate forecaster normal, with a 98.7% accuracy fee. The accuracy of 98.2% suggests that the model well recognizes actual positives with a low range of false positives, demonstrating the reliability of the advantageous predictions. The model's keep in mind, which stands at 98%, is quite lower and indicates its accuracy in figuring out actual wonderful instances. The model's capacity to deal with fake positives and false negatives is validated by the F1 rating, which stands at 97.9% and moves a compromise between accuracy and bear in mind. These measures, taken together, highlight the model's strong and remarkable performance and in efficiently figuring out and categorizing cardiovascular illnesses from X-ray angiography images.

TABLE III. PERFORMANCE EVALUATION

Methods	Accuracy	Precision	Recall	F1 score
RCNN	78.3%	78.1%	77.5%	77%
RNN-LSTM	88.4%	88.1%	88%	78.5%
3D-CNN	90%	89.5%	89.2%	88.8%
SVM	82.8%	82.3%	82%	81.5%
Proposed method	98.7%	98.2%	98%	97.9%

The performance parameters of several techniques for detection of coronary artery plaque and stenosis in X-ray angiography pics are shown in Table III. The proposed DCNN-GAN approach performs appreciably higher than the opposite methods. Its famous excellent basic accuracy with an accuracy charge of 98.7%. Its effective ability to discover certainly ideal facts is meditated in a take into account fee of 98%, and an accuracy of 98.2% indicating a very good prediction with a relatively reliable F1 score of 97.9% indicating consumption fake positives and negatives are dealt with correctly with a stability of remember and accuracy. In contrast, 78.3% accuracy is done through RCNN, 88.4% by using RNN-LSTM, 90.0%, and 82.8% by using SVM, all of which aren't reached by using

the proposed method a low F1 score, accuracy, and keep in mind, which confirmed that the proposed technique performs well in those parameters. This suggests that the DCNN-GAN system appreciably improves the sensitivity and efficiency of cardiac diagnosis in addition to presenting extra correct and reliable insights.

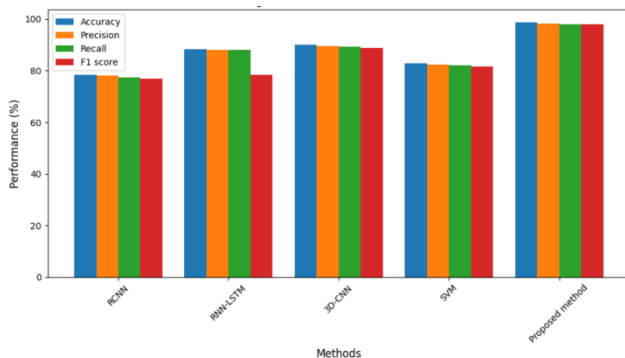


Fig. 11. Performance comparison.

The Fig. 11 indicates 4 measures of the efficacy of the diverse strategies in detecting pulmonary fibrosis and fibrosis. The proposed DCNN-GAN approach performs significantly higher than the other techniques. It achieves the very best universal values of round 98.7% accuracy, 98.2% accuracy, 98% recollect and 97.9% F1 score. All measures hover approximately 78%, with RCNN being the worst in evaluation. RNN-LSTM does better but still fall short, its metrics are around 88%. SVM scores approximately 82%, whereas the 3D-CNN approach performs competitively with nearly 90% for all measures. When compared to conventional approaches, the suggested method's evident superiority shows how robustly and correctly it can detect cardiovascular problems, greatly increasing diagnostic effectiveness. This demonstrates how the suggested approach may enhance clinical results in the identification and treatment of cardiovascular disease.

C. Discussion

Previous research on X-ray angiography-based coronary artery plaque and stenosis identification has encountered problems with noise, artifacts, and image quality, all of which harm the effectiveness of classifiers and accuracy in diagnosing [23]. These difficulties frequently cause the extraction of features and methods for preprocessing to be less successful, producing less dependable findings. These drawbacks are addressed by the suggested DCNN-GAN methodology, which incorporates deep learning methods to improve data robustness and framework dependability. While the DCNN product is extraordinary in correct function extraction, the GAN product offers artificial pix with varying quantities of noise distortion. This combined method gives the approach's potential to deal with optical differences so its use in diverse situations in medication is extremely good [24]. The technique retains outstanding results regardless of low-quality inputs since it was trained on a wide dataset of artificial and high-quality pictures. Subsequent research has to cognizance on increasing the adaptability of the DCNN-GAN machine to exclusive scientific settings and imaging modalities. Analysis of multidimensional data integration in essential modalities (e.g., combination of X-

rays with CT or MRI) and improvement of methods that can potentially modulate noises in energetically may also be vital to boom manner readability and doctor recognition and self-belief. Adoption and improvement of the machine may be based on partnerships with healthcare centers to provide greater range of statistics, promote computerized cardiovascular ailment, and beautify results for patients.

VI. CONCLUSION AND FUTURE WORKS

This proposed DCNN-GAN framework seems to perform very well in identifying the statuses of coronary artery plaque in X-ray angiography images, solving the problem of previously used approaches. Since, DCNN is optimal for function extraction and GAN for data augmentation the model is able to handle fluctuations in image quality, presence of noise and artifacts which is usual in medical images. The performance measurements, including the test accuracy that is 98.7%, precision 98.2% and recall 98%, corroborates the model productivity in providing accurate diagnostic results to further improve cardiovascular diseases diagnosis and early diagnosis. Besides enhancing the ability of doctors to diagnose their patients more accurately, this approach enhances the reliability and dependability of automatic systems being used in clinical practice.

Further, there is significant possibility for the development of DCNN-GAN model for other imaging modality like CT-MRI and by the application of the more comprehensive multiple dimension data for diagnosis. More studies could be conducted on the novel methods of dealing with dynamic noise levels and image distortions in real-time that would make the model more flexible for implementation in different clinical settings. Further cooperation with healthcare centers would be important in order to gather more varied and exhaustive information about patients, to let the model perform better throughout different individuals and diseases. Also, improved understanding of processes that enable real-time decision making and diagnostic accuracy through further enhancing of the model can contribute to widespread adoption of automated cardiovascular diagnostics into practice, which will benefit patient outcomes and decrease the time lapse before diagnosis. Implementation of such systems would dramatically change management of cardiovascular diseases with an ability to intervene at an early stage and potentially address heart diseases on a much bigger scale.

REFERENCES

- [1] A. Abdulmanafi, L. Duong, R. Ibrahim, and N. Dahdah, "A deep learning-based model for characterization of atherosclerotic plaque in coronary arteries using optical coherence tomography images," *Medical Physics*, vol. 48, no. 7, pp. 3511–3524, 2021.
- [2] X. Liu, J. Du, J. Yang, P. Xiong, J. Liu, and F. Lin, "Coronary artery fibrous plaque detection based on multi-scale convolutional neural networks," *Journal of Signal Processing Systems*, vol. 92, no. 3, pp. 325–333, 2020.
- [3] T. Masuda et al., "Deep learning with convolutional neural network for estimation of the characterisation of coronary plaques: Validation using IB-IVUS," *Radiography*, vol. 28, no. 1, pp. 61–67, 2022.
- [4] H. Cho et al., "Intravascular ultrasound-based deep learning for plaque characterization in coronary artery disease," *Atherosclerosis*, vol. 324, pp. 69–75, 2021.
- [5] L. Wang et al., "PlaqueNet: deep learning enabled coronary artery plaque segmentation from coronary computed tomography angiography," *Visual Computing for Industry, Biomedicine, and Art*, vol. 7, no. 1, p. 6, 2024.

- [6] S. Park et al., "A novel deep learning model for a computed tomography diagnosis of coronary plaque erosion," *Scientific Reports*, vol. 13, no. 1, p. 22992, 2023.
- [7] A. M. Fischer et al., "Accuracy of an artificial intelligence deep learning algorithm implementing a recurrent neural network with long short-term memory for the automated detection of calcified plaques from coronary computed tomography angiography," *Journal of thoracic imaging*, vol. 35, pp. S49–S57, 2020.
- [8] X. Jin et al., "Automatic coronary plaque detection, classification, and stenosis grading using deep learning and radiomics on computed tomography angiography images: a multi-center multi-vendor study," *European Radiology*, vol. 32, no. 8, pp. 5276–5286, 2022.
- [9] H. Shibutani et al., "Automated classification of coronary atherosclerotic plaque in optical frequency domain imaging based on deep learning," *Atherosclerosis*, vol. 328, pp. 100–105, 2021.
- [10] A. R. Ildayhid et al., "Coronary artery stenosis and high-risk plaque assessed with an unsupervised fully automated deep learning technique," *JACC: Advances*, p. 100861, 2024.
- [11] N. D. Schnellbacher et al., "Machine-learning-based clinical plaque detection using a synthetic plaque lesion model for coronary CTA," in *Medical Imaging 2021: Computer-Aided Diagnosis*, SPIE, 2021, pp. 654–660.
- [12] N. Gessert et al., "Automatic plaque detection in IVOCT pullbacks using convolutional neural networks," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 426–434, 2018.
- [13] D. L. Rodrigues, M. N. Menezes, F. J. Pinto, and A. L. Oliveira, "Automated detection of coronary artery stenosis in X-ray angiography using deep neural networks," *arXiv preprint arXiv:2103.02969*, 2021.
- [14] E. Ovalle-Magallanes, J. G. Avina-Cervantes, I. Cruz-Aceves, and J. Ruiz-Pinales, "Transfer learning for stenosis detection in X-ray coronary angiography," *Mathematics*, vol. 8, no. 9, p. 1510, 2020.
- [15] K. Pang, D. Ai, H. Fang, J. Fan, H. Song, and J. Yang, "Stenosis-DetNet: Sequence consistency-based stenosis detection for X-ray coronary angiography," *Computerized Medical Imaging and Graphics*, vol. 89, p. 101900, 2021.
- [16] M.-A. Gil-Rios et al., "Automatic feature selection for stenosis detection in x-ray coronary angiograms," *Mathematics*, vol. 9, no. 19, p. 2471, 2021.
- [17] P. V. Stralen, D. L. Rodrigues, A. L. Oliveira, M. N. Menezes, and F. J. Pinto, "Stenosis detection in X-ray coronary angiography with deep neural networks leveraged by attention mechanisms," in *Proceedings of the 9th International Conference on Bioinformatics Research and Applications*, 2022, pp. 123–128.
- [18] E. Ovalle-Magallanes, J. G. Avina-Cervantes, I. Cruz-Aceves, and J. Ruiz-Pinales, "Hybrid classical-quantum Convolutional Neural Network for stenosis detection in X-ray coronary angiography," *Expert Systems with Applications*, vol. 189, p. 116112, 2022.
- [19] T. Han et al., "Coronary artery stenosis detection via proposal-shifted spatial-temporal transformer in X-ray angiography," *Computers in Biology and Medicine*, vol. 153, p. 106546, 2023.
- [20] M. Algarni, A. Al-Rezqi, F. Saeed, A. Alsaedi, and F. Ghabban, "Multi-constraints based deep learning model for automated segmentation and diagnosis of coronary artery disease in X-ray angiographic images," *PeerJ Computer Science*, vol. 8, p. e993, 2022.
- [21] "NIH Chest X-rays." Accessed: Jun. 18, 2024. [Online]. Available: <https://www.kaggle.com/datasets/nih-chest-xrays/data>
- [22] T. G. Devi, N. Patil, S. Rai, and C. S. Philipose, "Gaussian blurring technique for detecting and classifying acute lymphoblastic leukemia cancer cells from microscopic biopsy images," *Life*, vol. 13, no. 2, p. 348, 2023.
- [23] S. Li and Y. Fan, "Coronary Artery Segmentation in X-ray Angiography Based on Deep Learning Approach," in *2024 43rd Chinese Control Conference (CCC)*, IEEE, 2024, pp. 7345–7350.
- [24] F. Denzinger et al., "Deep learning algorithms for coronary artery plaque characterisation from CCTA scans," in *Bildverarbeitung für die Medizin 2020: Algorithmen-Systeme-Anwendungen. Proceedings des Workshops vom 15. bis 17. März 2020 in Berlin*, Springer, 2020, pp. 193–198.

Design and Research of Accounting Automation Management System Based on Swarm Intelligence Algorithm and Deep Learning

Dan Gui¹, Wei Ma², Wanfei Chen^{3*}

Financial Assets Department,

State Grid Henan Electric Power Company Information and Communication Branch, Zhengzhou 450000, China^{1,3}

Finance Department, State Grid Henan Electric Power Company, Zhengzhou 450000, China²

Abstract—In the current research, the application verification of traditional algorithms in actual accounting management is insufficient, and deep learning data processing capabilities need to be fully optimized in complex accounting scenarios. Given the challenges of efficiency and accuracy faced by the current accounting industry in the context of big data, this study creatively combines the swarm intelligence algorithm and deep learning technology to design and implement an efficient and accurate accounting automation management system. The research aims to investigate the potential of swarm intelligence algorithms and deep learning techniques in developing an automated accounting management system, with a focus on improving efficiency, accuracy, and scalability. Key research questions include exploring the optimal configuration of swarm intelligence algorithms for accounting tasks and assessing the performance of deep learning models in automating various accounting processes. Through experimental verification, the system is tested with the financial data of a large enterprise for three consecutive years. The results show that the system can significantly shorten the time of financial statement generation by 65%, reduce the error rate to less than 0.5%, and increase the accuracy of abnormal data recognition by as much as 90%. These data not only reflect the significant improvement of the efficiency and accuracy of the system but also prove its great potential in early warning of financial risk, providing intelligent and automated solutions for the accounting industry.

Keywords—Swarm intelligence algorithm; deep learning; accounting; automation management

I. INTRODUCTION

In today's wave of digital transformation, the accounting industry faces unprecedented challenges and opportunities [1, 2]. Traditional accounting processes, including data entry, account reconciliation, financial statement generation, etc., often rely on manual operations, which are time-consuming, labor-intensive, and prone to errors [3]. With the rapid development of big data, artificial intelligence, and other technologies, the emergence of accounting automation management systems provides new ideas and possibilities for solving these problems. Among them, the integration and application of swarm intelligence algorithms and deep learning technology are becoming a research hotspot in accounting automation management, opening up a new path for realizing the intelligence and automation of accounting work [4, 5].

Swarm intelligence algorithms, such as the ant colony algorithm and particle swarm optimization algorithm, are optimization algorithms that imitate the behavior of biological swarms in nature and have strong global search ability and robustness [6, 7]. In accounting automation management, swarm intelligence algorithms can effectively deal with complex decision-making problems, such as financial forecasting, cost control, etc., and find the optimal or approximately optimal solution by simulating the collaboration and competition of swarms [8]. Deep learning, as an essential branch of artificial intelligence, can automatically learn features from massive financial data through its powerful data processing and pattern recognition capabilities, realize automated account classification, anomaly detection, and other functions, and significantly improve the efficiency and accuracy of accounting work [9, 10].

However, there are still many challenges in applying swarm intelligence algorithms and deep learning technology to design accounting automation management systems [11]. How to design a reasonable algorithm model to adapt to the complexity and diversity of accounting data; How to ensure the stability and robustness of the algorithm and avoid decision-making errors caused by data fluctuations; How to realize the effective use of data on the premise of protecting data privacy is an urgent problem to be solved [12, 13]. In addition, developing and applying an accounting automation management system also involves compatibility with existing accounting software, user interface design, system security, and other issues, which require interdisciplinary knowledge and skills [14]. While there has been significant progress in the development of accounting automation systems, there are still several challenges and limitations that remain unaddressed. For instance, existing systems often lack the flexibility and scalability to handle complex accounting tasks and large datasets efficiently. Additionally, many systems rely on traditional rule-based approaches, which may not be well-suited for the dynamic and unpredictable nature of accounting data. In this paper, we propose a novel accounting automation management system based on swarm intelligence algorithms and deep learning techniques that aims to address these limitations and fill the existing gap in the field. Our system leverages the strengths of swarm intelligence for optimization and deep learning for pattern recognition, enabling it to handle complex accounting tasks with high accuracy and efficiency.

The purpose of this study is to deeply explore the key technologies and methods of accounting automation management system design based on swarm intelligence algorithm and deep learning and propose an efficient, intelligent, and safe accounting automation management solution through theoretical analysis and empirical research to provide theoretical guidance and practical reference for the digital transformation of the accounting industry. This research will focus on the following contents: First, analyze the application potential and limitations of swarm intelligence algorithm and deep learning in accounting automation management; The second is to design and implement the prototype of an accounting automation management system based on swarm intelligence algorithm and deep learning, and evaluate its performance and effect; The third is to put forward the algorithm optimization strategy according to the characteristics of accounting data to improve the intelligence level of the system; The fourth is to discuss the challenges and countermeasures of accounting automation management system in practical application, and provide direction for future research and practice.

The design and research of accounting automation management systems based on swarm intelligence algorithms and deep learning is not only an essential direction of technological innovation in the accounting field but also a key force in promoting the digital transformation of the accounting industry. Through this study, we expect to provide new ideas and methods for developing and applying accounting automation management systems, promote the intelligence and automation process of the accounting industry, and provide more powerful and intelligent tools for enterprise financial management and decision support. In the following part of this article, we will delve into the design and implementation of an accounting automation management system based on bee colony intelligence algorithms and deep learning technology. In Section II, we will first introduce the theoretical background and the research related to the research on accounting automation management strategy based on swarm intelligence algorithm, and then describe in detail the system architecture and method we propose in Section III. In Section IV, we will present the results of the experimental evaluation, demonstrating the effectiveness and efficiency of our system. Finally, in Section V, we will discuss the significance of our findings and suggest directions for future research. The paper is concluded in Section VI.

II. RESEARCH ON ACCOUNTING AUTOMATION MANAGEMENT STRATEGY BASED ON SWARM INTELLIGENCE ALGORITHM

A. Ant Colony Algorithm

Social insects in nature, such as ants, show strong adaptability and flexibility and cooperate to complete tasks such

as foraging and nesting by releasing up to 20 kinds of pheromones. These pheromones provide a means of navigation and communication for ants with limited vision, especially playing a key role in finding pathways back to the nest and dividing labor and cooperating [15]. Inspired by the social behavior of ants, scientists have developed ant colony algorithms, which provide new ideas for solving complex problems by simulating the action mechanism of pheromones. Years of studies have shown that ants can secrete a substance that affects the environment, promotes mutual communication or perceives environmental changes, namely pheromones. This discovery has deepened our understanding of ants' environmental interaction mechanisms [16, 17].

Imagine an ant facing two paths around obstacles. Because there is no former pheromone to guide it, it chooses randomly. During walking, ants release pheromones, which provide a selection basis for subsequent ants. Subsequently, ants are more inclined to choose the path with high pheromone concentration and supplement pheromone at the same time, and the pheromone concentration will naturally decay with time [18, 19]. Through the continuous selection and pheromone update of colony ants, pheromone space is formed to guide ants in making decisions.

The ant colony algorithm is based on an abstract model, which is shown in Fig. 1. Ants communicate and obtain environmental information through pheromones and perform tasks independently. The model assumptions include the following: environmental changes have feedback on ants; Pheromone is a medium for ants to communicate and obtain environmental information; Except for pheromones, behaviors are independent of each other and are not directly affected by other ants. The primary challenge of applying an ant colony algorithm is to transform the problem into an understandable form of an ant colony; that is, the problem needs to be properly described. After the description, an algorithm model based on the pheromone decision mechanism is developed. As the basis of ant communication and organization, the pheromone must simulate the use of the pheromone in the algorithm to guide ants' exploration path, just like real ants foraging and obstacle avoidance [20]. Renewal of pheromones directly controls ant colony behavior.

In applying the ant colony algorithm, heuristic information is defined as the basis of ant decision-making, which usually reflects prior knowledge. For example, in the traveling salesperson problem, the heuristic information is the reciprocal of the path distance. The longer the path, the smaller the information value, which affects the ant transition probability. Its calculation formula is shown in Eq. (1). In network problems, heuristic information correlates delay, bandwidth, and packet loss rate to guide the algorithm in finding the path that conforms to the network characteristics.

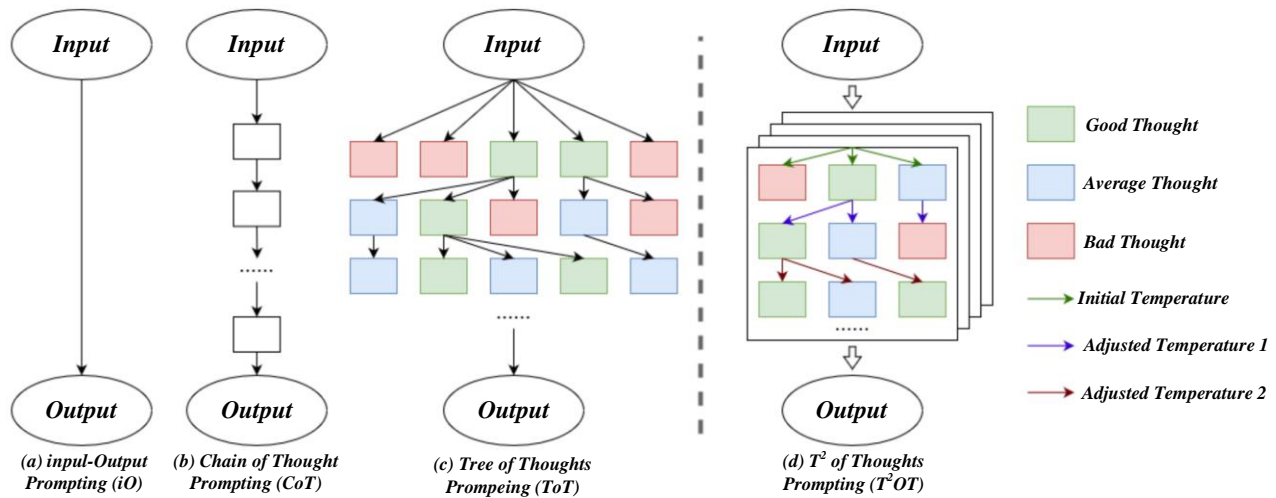


Fig. 1. Ant colony algorithm model.

$$p_{ij}^k = \begin{cases} \frac{[\tau_{ij}]^\alpha [\eta_{ij}]^\beta}{\sum_{j \in allowed_k} [\tau_{ij}]^\alpha [\eta_{ij}]^\beta}, j \in allowed_k \\ 0, otherwise \end{cases} \quad \Delta\tau_{ij}^k = \begin{cases} Q / L_k & Ki \text{ to } Kj \\ 0, & other \end{cases} \quad (5)$$

In Eq. (1), η_{ij} represents the heuristic information of the link (i, j) , τ_{ij} represents the pheromone concentration of the link (i, j) , and $allowed_k$ represents the set of nodes that ant k can transfer to in the next step, α and β have the influence coefficients representing the pheromone concentration τ_{ij} and the heuristic information η_{ij} of network link (i, j) , respectively. At the initial time, the pheromone concentration between each path is the same, and $\eta_{ij}(t)$ of the molecule is the heuristic information from node i to node j . Its expression is $\eta_{ij}(t) = 1/d_{ij}(t)$, where d_{ij} is the Euclidean distance between node i and node j , and the expression is Eq. (2):

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (2)$$

(x_i, y_i) is a heuristic factor representing the relative importance of pheromone concentration; (x_j, y_j) is the visibility heuristic factor. The ant determines the next position by calculating the path transition probability and applying the wheel gambling method. After completing a round of searching, the path length is evaluated, and the shortest path is identified. Subsequently, the pheromone is updated according to the principle of "volatilization-release," as shown in Eq. (3) and Eq. (4).

$$\tau_{ij}(t+1) = (1-\rho) * \tau_{ij}(t) + \Delta\tau_{ij}^k \quad (3)$$

$$\tau^k = \sum_{k=1}^m \Delta\tau_{ij}^k \quad (4)$$

Where ρ is the pheromone volatilization factor, and the value range is $\rho \in [0, 1]$; m represents the number of ant colonies. Equation (5) represents the amount of pheromone left by all ants in the current search process, where the pheromone left by each ant in the current iteration process can be expressed as:

Among them, Q is a constant, which represents the total amount of pheromone that can be released in one search process of ant pheromone; L_k represents the total length of the path traveled by ant k in this cycle, and K_i to K_j represents a range sequence.

B. Drosophila Algorithm

Set the population size (popsize) and the maximum number of iterations (maxgen). Among them, $(X_{axis}; Y_{axis})$ represents the two-dimensional coordinates of each individual in the fruit fly community, LR represents the position range of the fruit fly population, and the mathematical expression of the initial position is as follows (6)-(7):

$$X_{axis} = rand(LR) \quad (6)$$

$$Y_{axis} = rand(LR) \quad (7)$$

$rand$ is a function used to generate random numbers. When individuals in the community search for food, their flight direction and distance are random. The following expression represents the new position when the fruit fly i flies to the next moment, as shown in Eq. (8)-(9):

$$X_i = X_{axis} + rand(FR) \quad (8)$$

$$Y_i = Y_{axis} + rand(FR) \quad (9)$$

FR denotes the range of a single flight, and $axis$ refers to the straight line in the coordinate system. Because the specific location of the food source is unknown, first use the following formula to calculate the distance $Dist_i$ of the individual fruit fly from the origin:

$$Dist_i = \sqrt{X_i^2 + Y_i^2} \quad (10)$$

Then, the taste concentration determination value S_i is calculated by the following Eq. (11):

$$S_i = 1 / Dist_i \quad (11)$$

Fitness refers to the quality or quality measure of a solution. The taste concentration value $Smell_i$ of each individual in the current population is expressed by the following Eq. (12):

$$Smell_i = fitness(S_i) \quad (12)$$

C. Ant Colony Algorithm for Path Preprocessing of Drosophila Algorithm

When the traditional ant colony algorithm searches the path, the node traversal probability is equal, leading to a blind search in the initial stage [21]. In order to solve this problem, the Drosophila algorithm is introduced to pre-plan the path, which is transformed into the pheromone required by the ant colony algorithm, and the ant colony algorithm is guided to avoid blind search, reduce node traversal, and shorten the running time [22]. After pre-planning, the ant colony algorithm optimizes the search on this basis.

According to the requirements, FOA (Fruit Fly Optimization Algorithm) is first used for path preprocessing, and then the preprocessed path is converted into a pheromone, which is imported into the ant colony algorithm (ACO). The calculation formula is shown in Eq. (13).

$$\tau(i, j) = \tau^s(i, j) + \Delta\tau^B(i, j) \quad (13)$$

Where $\tau(i, j)$ represents the pheromone concentration from node i to node j , $\tau^s(i, j)$ is the original pheromone from node i to node j , and $\tau^B(i, j)$ refers to the pheromone increment from node i to node j converting the results of FOA search.

The ant uses the roulette method and selects the next node according to the formula. After the algorithm is finished, the pheromone is updated, and a portion is added and volatilized. After each generation of ants iterates, the paths are compared, and the shortest path is determined when the set number of iterations is reached.

III. DESIGN OF ACCOUNTING AUTOMATION MANAGEMENT SYSTEM BASED ON SWARM INTELLIGENCE ALGORITHM AND DEEP LEARNING

A. Deep Learning Neural Network Technology

In the study, the performance and effectiveness of the accounting automation management system were evaluated in depth, and the comprehensive functionality, data accuracy and operational efficiency of the accounting automation management system were compared with popular accounting and financial software solutions in the market (such as QuickBooks, Xero, SAP, Oracle Financials and Kingdee, which may include other software) in order to fully verify the performance of the system in the real financial environment. To achieve this assessment, the accounting automation management system was systematically integrated with the above-mentioned software and extensive testing and data analysis were implemented. In terms of text processing of accounting sheets, special attention is paid to text records

containing complex information such as equipment numbers, timestamps, exception reports, etc., and through preprocessing (including text segmentation, data cleaning, format standardization) and feature extraction (using bag-of-word model BoW, word frequency-inverse document frequency TF-IDF, N-gram model and word embedding technology, etc., the word embedding technology effectively solves the problems of dimensional disaster, data sparsity and semantic loss caused by high-dimensional data by mapping words or phrases to real low-dimensional vectors [25]), which successfully extracted critical information, which was critical for subsequent root cause analysis and ticket recommendations [23, 24]. In addition, the autoencoder (AE) deep learning technology is introduced to compress the high-dimensional input data into the low-dimensional feature vector space by using its powerful feature extraction ability [26], and the applications of denoising autoencoder (to enhance the robustness of the model), sparse autoencoder (to improve the compression ratio and learn the data structure), and variational autoencoder (to maximize the probability of data sample union, optimize the model parameters and hidden variables a posteriori, and endow the model generation ability and latent distribution simulation ability) are explored.

Neural network attention is introduced into the design of an accounting automation management system, which makes the network automatically focus on crucial information and improves its performance and interpretability. It is divided into three categories: soft attention (global), intricate attention (local), and self-attention (internal attention) [27, 28]. Soft attention calculates the weights based on similarity, utilizing all the input information, but it is computationally heavy. Intricate attention focuses on local areas, reduces computation, and conforms to visual characteristics but may ignore critical information. Self-attention considers inter-input and inter-output relationships and enhances long-distance dependency and structure capture. Principle of attention mechanism: When generating output, the model selectively focuses on relevant parts of the input according to the context, assigns different weights, optimizes information coding, and improves accuracy and robustness. The steps include calculating weights, such as dot product, additivity, self-attention, etc.; Apply weights to the weighted average or splice the inputs; Update the output, direct output, or further process.

In order to optimize the feature extraction capability, this study applies convolution operation to graph structure to realize graph data feature learning and classification. Unlike CNN in regular grid processing, GCN (Graph Convolution Networks) is good at unstructured graph data. The advantages are that it deals with complex graph structures and has good interpretability and visualization. The adjacency matrix expresses the graph structure, vectorizes node and edge features, and updates the features by convolution. Through the operation of the adjacency matrix and node feature matrix, feature aggregation and transfer are realized, and local features are extracted by convolution, similar to CNN. GCN also contains pooling operations for dimension reduction and feature abstraction of graph data.

B. Model Building

Automated accounting analysis aims to quickly and accurately identify and solve system failures and improve

system stability and reliability. Analysis methods are divided into traditional and machine learning categories. Traditional methods rely on manual system performance analysis or use tools such as FMEA to identify anomalies. However, their ability to process large-scale data and quickly locate anomalies is limited and error-prone [29, 30]. The machine learning method builds a model by analyzing the system structure and historical data and automatically identifying anomalies' root causes. It has the advantages of processing massive data, quickly analyzing, and automatically adjusting the model to improve accuracy and real-time response, significantly superior to traditional manual methods.

The model is shown in Fig. 2. First, the preprocessed node data is vectorized for model training. The words in the node are split into characters, and a 70-dimensional hot encoding represents each character. The encoding table contains 26 letters, ten digits, 33 unique characters, and line breaks. Each node is transformed into a matrix of 70. Excessively long characters are truncated, and more characters should be filled with zero vectors. The encoding order starts from the last character, which is convenient for the weight association of fully connected layers. The vectorized node data is input into the model training. The model contains nine layers of neural network: 6 one-dimensional convolutional layers and three fully connected layers. A maximum pooling layer follows the first, second, and last convolutional layers. In the model, the kernel size of the first two convolution layers is 7, the core size of the last four layers is 3, each layer has 256 kernels, and the size of the pooling layer is 3. The node data outputs a 1008×256 feature map through the first convolutional layer. After the first pooling layer, the feature map size is 336×256 .

Text feature extraction only uses work order data, ignoring the relationship between nodes. GCN model can handle graph structure, learn nodes and relationship features, and be used for root cause analysis. GCN training requires node features and an adjacency matrix. The feature extraction method is the same as before, and the top 200 high-frequency words are selected. The adjacency matrix represents the node relationship, and the weight is the number of occurrences of dependent paths in the historical work order description. GCN consists of two graph convolutional layers and a fully connected layer. Node features and adjacency matrix input convolutional layer, activated by ReLU, and fully connected layer and SoftMax output node features. The node features and graph features extracted from the text are used to match the abnormal root causes of the work

order. Preprocess the exception description of the current work order, convert it into a character-level vectorized representation, and obtain the work order feature vector. By calculating the comprehensive similarity between the work order and the node, the node with the highest similarity is selected as the abnormal node. The root cause analysis is output by classification, and the similarity between the work order and each node is given, and the probability of abnormal nodes is calculated accordingly.

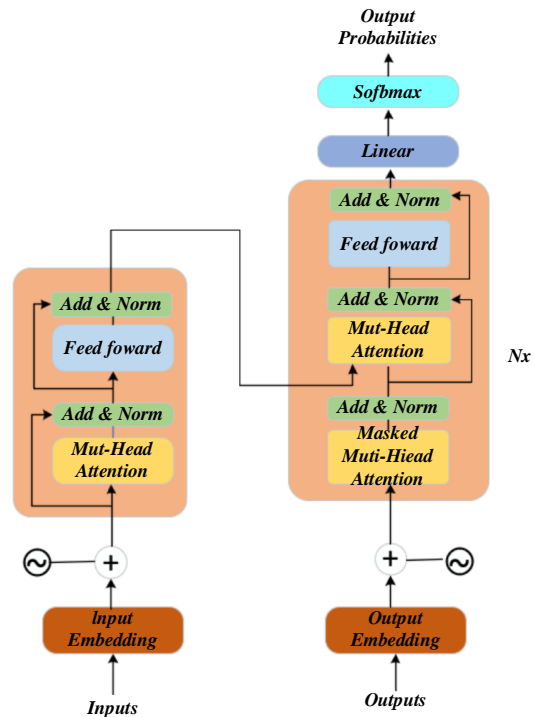


Fig. 2. Input data processing model.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

Fig. 3 compares the algorithm's performance at a confidence level of 90%. The model based on swarm intelligence algorithm and deep learning performs outstandingly in two-thirds of the dimension of ROC, thanks to its utilization of second-order difference information, which improves search efficiency. The algorithm also has advantages in the extreme dimensions of ES and entropy. The model continues to show competitiveness at higher confidence levels (95%, 97.5%, 99%).

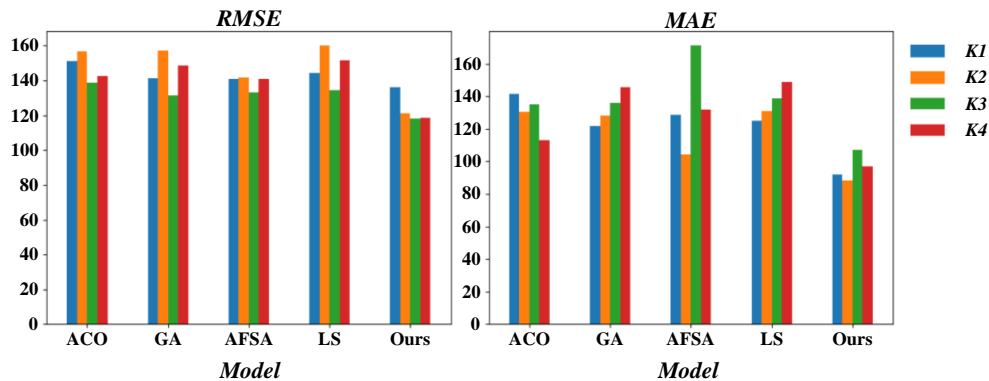


Fig. 3. Model results.

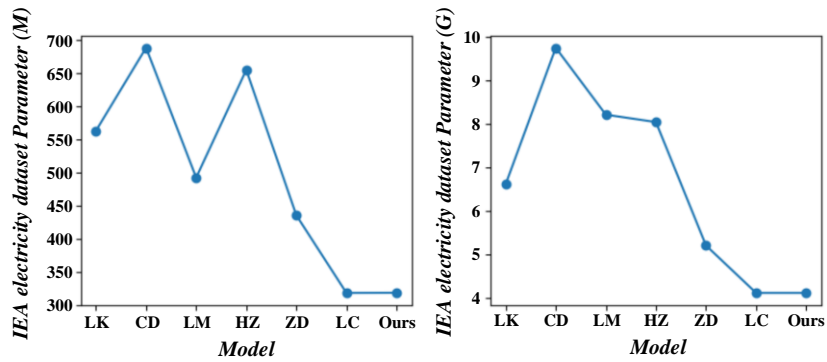


Fig. 4. Results at different confidence levels.

In Fig. 4, the HV value based on the swarm intelligence algorithm and deep learning model is the highest, and the second-order differential evolution operator effectively broadens the frontier of the Pareto solution set and enhances the extensibility of understanding. Table I shows that GBDT is the best among traditional methods but not as good as deep learning. CharCNN-based TicketRCA is better than CNN and RNN models. Due to the lack of a timing series of work order data keywords, models such as TextCNN are limited. CharCNN character-level representation is better. After the combination of GCN, TicketRCA root cause analysis is significantly improved, proving that the graph model is adequate. Although TicketMining and NetSieve are better than traditional machine learning, they are not as good as TicketRCA, indicating that deep learning can better extract text features.

In Fig. 5, Four strategies construct the anchor node table: the top 100 is selected by degree sorting, the top 100 by PageRank sorting ($d = 0.9$), and 100 is randomly selected. The joint

strategy combines the top three in a ratio of 4: 4: 2 (degree sorting 40, PageRank 40, random 20). Experiments show that the accuracy of degree ranking and PageRank strategy is similar. The stochastic strategy also performs similarly, and the joint strategy has the highest accuracy, indicating that the combination of centrality and random noise is beneficial to improve the node classification performance.

In the ablation experiment, the joint strategy is used to construct the anchor node table and generate the subgraph sequence. Fig. 6 shows that the complete model samples ten anchor nodes and distances, three neighbors and relationships, one self-node, and a self-ring edge. The ablation model does not sample anchor nodes, neighbors, self-nodes, and self-ring edges. The complete model is better than the Transformer, which samples the whole graph. In ablation, unsampled anchor nodes have the most significant influence, followed by neighbors, and self-nodes and self-ring edges have less influence.

TABLE I. MODEL COMPARISON OF RESULTS

	Accuracy	Precision	Recall	F1-score
SVM	0.3311	0.2717	0.3311	0.2706
DecisionTree	0.3872	0.2519	0.3872	0.2904
RandomForest	0.3971	0.3014	0.3971	0.3179
GBDT	0.5214	0.4488	0.5126	0.4642

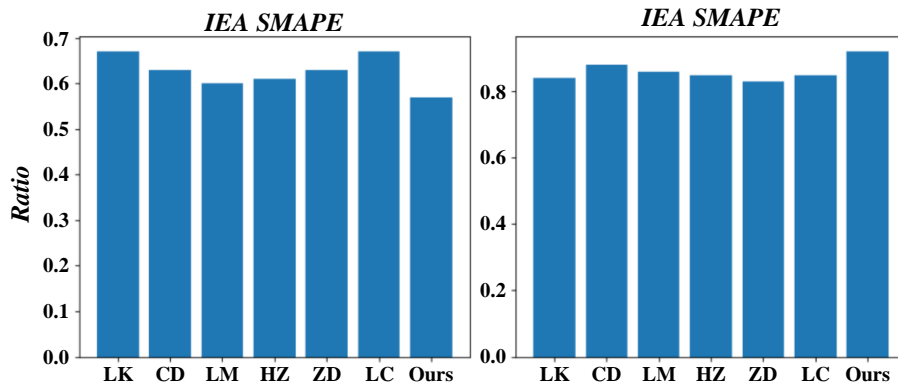


Fig. 5. Comparison of node classification accuracy of anchor node sampling methods based on different strategies.

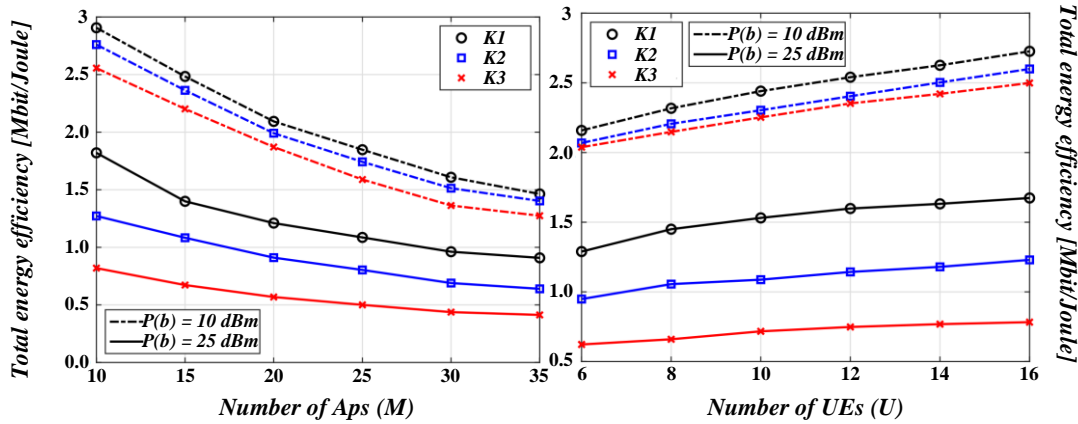


Fig. 6. Comparison of node classification accuracy based on different sampling objects.

The hyperparameter analysis is shown in Fig. 7. The data set performs best when the anchor node table size is 100 and 10 anchor nodes are sampled. Increasing the number of anchor nodes usually improves accuracy, but performance degrades after more than 10-20. The performance is optimal when the anchor node table size is 100. In order to improve system performance, more anchor nodes are needed to cover a wide range of knowledge fields due to the vast number of entities and relationships. When using large and complex data sets, it is necessary to build a larger anchor node table, sample more anchor nodes, and select appropriate strategies to ensure representativeness and improve model performance and robustness.

Fig. 8 shows the impact of the number of convolution kernels on predictive analysis. Using a data set with rich relationship types, its 237 relationships are denser. The results support the conjecture. more convolution kernels and high-dimensional relational embeddings improve performance. Since 128-dimensional embeddings cannot fully describe features, high-dimensional embeddings require broad convolution support.

Fig. 9 shows that on the sparse graph, the experimental results decrease by 5% and 9%, respectively, highlighting the improvement of the method in prediction accuracy. Removing both mechanisms significantly decreased accuracy. The embedding quality may only depend on the relationship

embedding of small graphs near the target node. One-hop relationship is more important than neighbor nodes. Rich relationship types generate more combinations and provide nodes with more expressive subgraph features. Two-dimensional convolution and Transformer encoder fuse relationships and node embeddings. Therefore, compared with FB15k-237, which has 20 times more types of relationships, the performance degradation of WN18RR is more evident without these two modules.

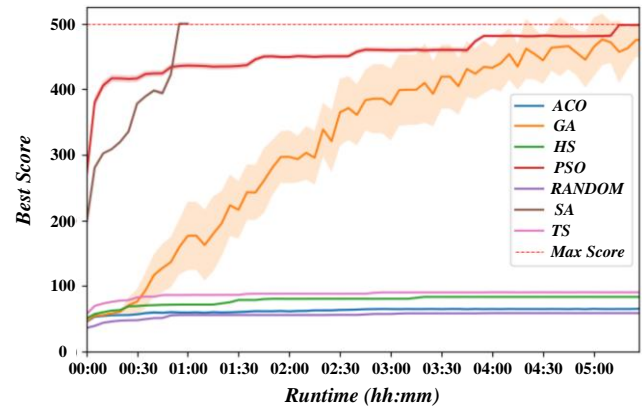


Fig. 7. Hyperparameter analysis.

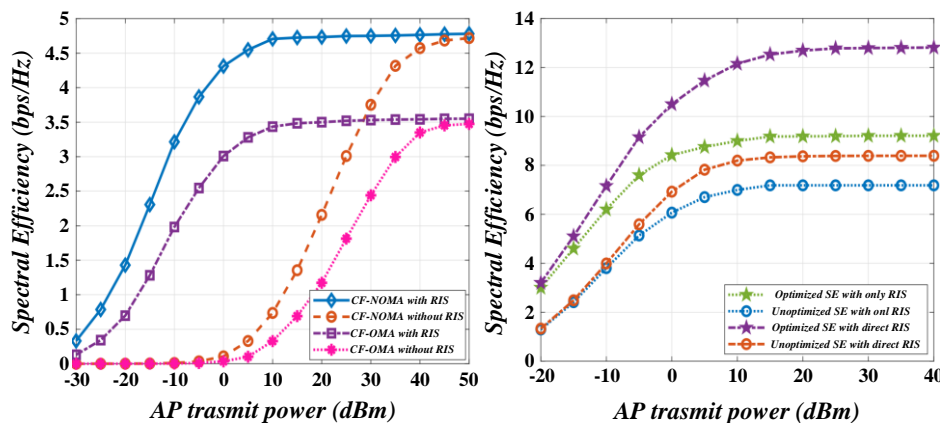


Fig. 8. Effect of the number of convolution kernels on predictive analysis.

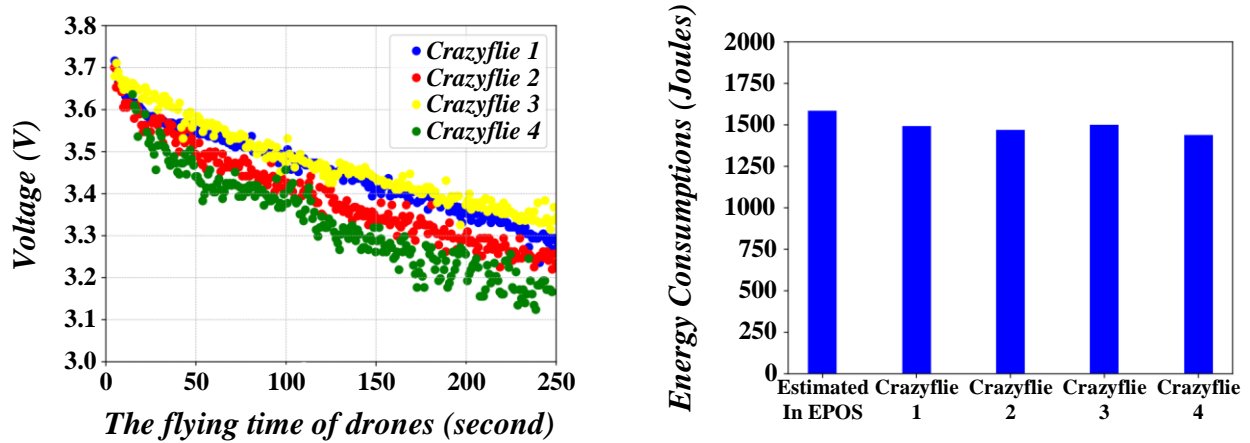


Fig. 9. Experimental results of sparse graph.

Fig. 10 shows that the AUC values of all models exceed 0.5, indicating that the prediction is better than a random guess and has data fitting ability for M&A prediction. The linear regression AUC value is high, but the F1 score is lower than that of decision tree regression, reflecting its foundation and low robustness. The logistic regression AUC reached 0.65, which was the best performance, thanks to the non-linear fit. The decision tree regression AUC was only 0.551, which was poorly fitted. AdaBoost boosts AUC but is limited by weak classifier performance. Through high-dimensional mapping and the kernel trick, the AUC of SVM is 0.014 higher than AdaBoost's. The overall effect of the machine learning method is not good, mainly due to insufficient capture of M&A data distribution information. Baseline performed the worst among the deep learning models, with an AUC of about 0.5. LSTM is significantly improved, AUC super logistic regression 0.04, good at long sequence data, capturing data connections. The model proposed in this paper has an AUC of 0.721, the best

ACC and F1Score, which are 0.071 and 0.031 higher than logistic regression and LSTM, respectively, indicating that it can effectively fit the prediction data and achieve accurate prediction.

Fig. 11 shows the loss curve of the ablation experimental model, which intuitively reflects the prediction accuracy. The AUC value of AttDNN was 0.721, with the largest area of the ROC curve, followed closely by LSTM, with a difference of 3.1 percentage points in AUC values. Logistic regression AUC is higher and marked yellow. The AUC of linear regression, decision tree, AdaBoost, and SVM are low, the prediction effect is poor, and the ROC curves almost coincide. As the basic algorithm of deep learning, the Baseline lacks a regularization layer and attention mechanism training. Its ROC curve is close to the (FPR = 1, TPR = 1) connection of random guess, and the prediction effect is equivalent to random.

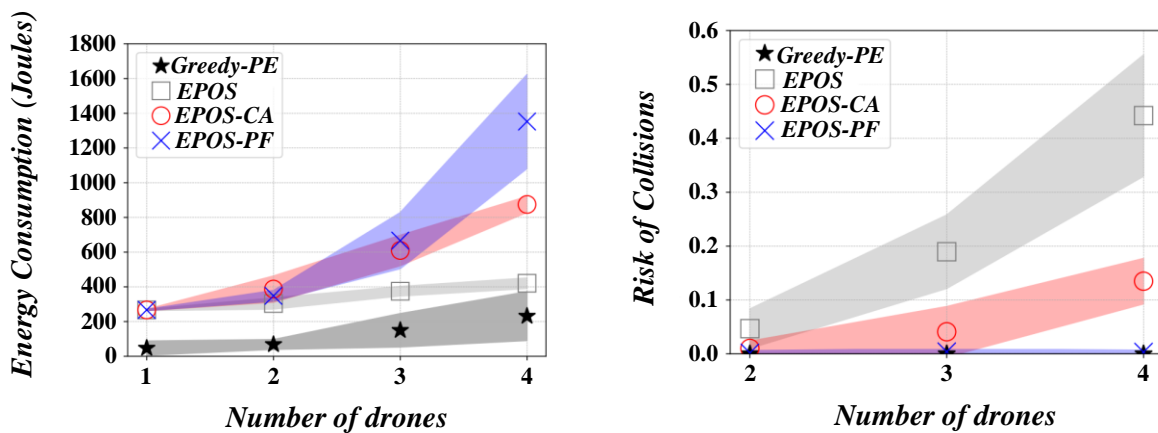


Fig. 10. Results of ablation experiment.

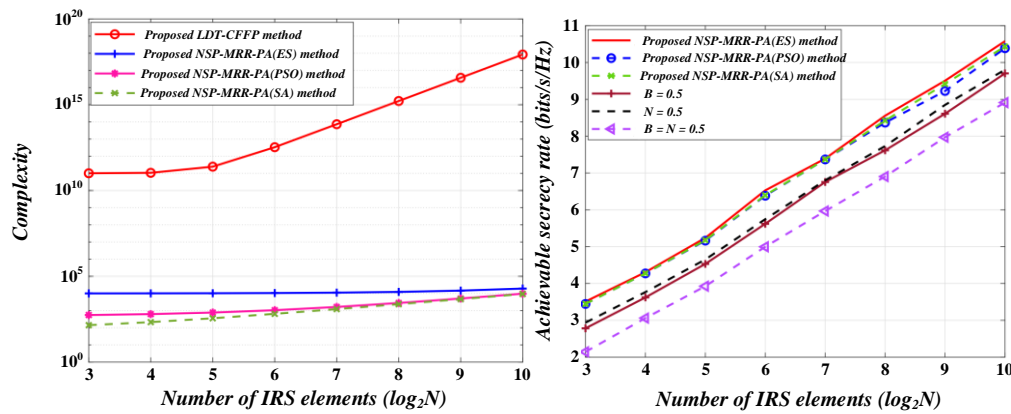


Fig. 11. Loss curve of ablation experimental model.

V. DISCUSSION

Our results demonstrate the effectiveness and efficiency of our proposed system in handling complex accounting tasks and large datasets. However, there are still several challenges and limitations that need to be addressed in future work. For instance, while our system has shown promising performance in controlled environments, its robustness and scalability in real-world scenarios remain to be tested. Additionally, there is a need for further research on the integration of our system with existing accounting software and platforms to facilitate seamless adoption and use. In conclusion, our work represents a significant step forward in the field of accounting automation, and we believe that it provides a solid foundation for future research and development efforts.

VI. CONCLUSION

In the context of the digital transformation of the accounting industry, this study creatively combines a swarm intelligence algorithm with deep learning technology to design and develop an intelligent accounting automation management system that aims to address the limitations of traditional accounting management in big data processing.

By applying swarm intelligence algorithms and deep learning technology to accounting automation management, this study significantly improves the efficiency and accuracy of accounting work and enhances the ability to detect early warnings of financial risk. Experimental data show that the system's efficiency in processing financial statements has increased by 65%, mainly due to the optimization path of the swarm intelligence algorithm and the intelligent analysis ability of deep learning, which significantly shortens the data processing cycle.

This study provides an intelligent and automated solution for the accounting industry, and the accuracy rate is greatly improved. The error rate is reduced to less than 0.5%, indicating that the system can effectively identify and reduce human errors when processing complex financial data, significantly improving the accuracy of data processing.

Through continuous training of deep learning models, the system's recognition accuracy of abnormal data has reached

90%, providing strong technical support for financial risk early warning and effectively reducing potential financial risks.

This study designed an accounting automation management system based on swarm intelligence algorithms and deep learning, which is expected to improve accounting processes' efficiency and accuracy significantly. However, this study has limitations: the system was only tested in a controlled environment, and the experimental dataset cannot cover all accounting task scenarios. Subsequent research requires more extensive testing in real-world scenarios, using more diverse datasets and further exploring integration with existing accounting software platforms to develop complex swarm intelligence algorithms and deep learning models to enhance system performance.

REFERENCES

- [1] J. E. Gerken et al., "Geometric deep learning and equivariant neural networks," *Artificial Intelligence Review*, vol. 56, no. 12, pp. 14605-14662, 2023.
- [2] M. S. Hashish, H. M. Hasanien, Z. Ullah, A. Alkhuayli, and A. O. Badr, "Giant Trevally Optimization Approach for Probabilistic Optimal Power Flow of Power Systems Including Renewable Energy Systems Uncertainty," *Sustainability*, vol. 15, no. 18, 2023.
- [3] S. Ferilli, E. Bernasconi, D. Di Pietro, and D. Redavid, "A Graph DB-Based Solution for Semantic Technologies in the Future Internet," *Future Internet*, vol. 15, no. 10, 2023.
- [4] C. Cavallaro, C. Crespi, V. Cutello, M. Pavone, and F. Zito, "Group Dynamics in Memory-Enhanced Ant Colonies: The Influence of Colony Division on a Maze Navigation Problem," *Algorithms*, vol. 17, no. 2, 2024.
- [5] C. L. Galimberti, L. Furieri, L. Xu, and G. Ferrari-Trecate, "Hamiltonian Deep Neural Networks Guaranteeing Nonvanishing Gradients by Design," *Ieee Transactions on Automatic Control*, vol. 68, no. 5, pp. 3155-3162, 2023.
- [6] J. Yu, X. You, and S. Liu, "A heterogeneous guided ant colony algorithm based on space explosion and long-short memory," *Applied Soft Computing*, vol. 113, 2021.
- [7] F. Zhang, Z. Gao, J. Huang, P. Zhen, H.-B. Chen, and J. Yan, "HFOD: A hardware-friendly quantization method for object detection on embedded FPGAs," *IEICE Electronics Express*, vol. 19, no. 8, 2022.
- [8] X. Wang, Z. Fu, and X. Li, "A Graph Deep Learning-Based Fault Detection and Positioning Method for Internet Communication Networks," *IEEE Access*, vol. 11, pp. 102261-102270, 2023.
- [9] H. Chen and H. Eldardiry, "Graph Time-series Modeling in Deep Learning: A Survey," *Acm Transactions on Knowledge Discovery from Data*, vol. 18, no. 5, 2024.

- [10] L. V. Jospin, H. Laga, F. Boussaid, W. Buntine, and M. Bennamoun, "Hands-On Bayesian Neural Networks-A Tutorial for Deep Learning Users," *IEEE Computational Intelligence Magazine*, vol. 17, no. 2, pp. 29-48, 2022.
- [11] H. Liu, B. Hu, and Y. Cao, "HDMA-CGAN: Advancing Image Style Transfer with Deep Learning," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 38, no. 09, 2024.
- [12] J. Fu, Z. Yang, M. Liu, H. Zhang, and Y. Zhang, "Highly-efficient design method for coding metasurfaces based on deep learning," *Optics Communications*, vol. 529, 2023.
- [13] Emilio Abad-Segura, Alfonso Infante-Moro, Mariana-Daniela González-Zamar, and Eloy López-Meneses, "Influential factors for a secure perception of accounting management with blockchain technology," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 10, no. 2, pp. 100264, 2024.
- [14] Erik S. Boyle, "How do auditors' use of industry norms differentially impact management evaluations of audit quality under principles-based and rules-based accounting standards?" *Journal of International Accounting, Auditing and Taxation*, vol. 54, pp. 100598, 2024.
- [15] Elsa Pedroso and Carlos F. Gomes, "Disentangling the effects of top management on management accounting systems utilization," *International Journal of Accounting Information Systems*, vol. 53, pp. 100678, 2024.
- [16] Fangjuan Qiu, Nan Hu, Peng Liang, and Kevin Dow, "Measuring management accounting practices using textual analysis," *Management Accounting Research*, vol. 58, pp. 100818, 2023.
- [17] Linda Hui Shi, Kristin Brandl, Jing Song, and Shaoming Zou, "Global account management: Knowledge resources and capabilities for relationship management," *International Business Review*, vol. 33, no. 5, pp. 102315, 2024.
- [18] Haiyan Sun, "Construction of integration path of management accounting and financial accounting based on big data analysis," *Optik*, vol. 272, pp. 170321, 2023.
- [19] Esra Gülmez, Halil Ibrahim Koruca, Mehmet Emin Aydin, and Kemal Burak Urganci, "Heuristic and swarm intelligence algorithms for work-life balance problem," *Computers & Industrial Engineering*, vol. 187, pp. 109857, 2024.
- [20] Gang Hu, Feiyang Huang, Kang Chen, and Guo Wei, "MNEARO: A meta swarm intelligence optimization algorithm for engineering applications," *Computer Methods in Applied Mechanics and Engineering*, vol. 419, pp. 116664, 2024.
- [21] Su Hu and Hua Yin, "Research on the optimum synchronous network search data extraction based on swarm intelligence algorithm," *Future Generation Computer Systems*, vol. 125, pp. 151-155, 2021.
- [22] Lifu Ding, Youkai Cui, Gangfeng Yan, Yaojia Huang, and Zhen Fan, "Distributed energy management of multi-area integrated energy system based on multi-agent deep reinforcement learning," *International Journal of Electrical Power & Energy Systems*, vol. 157, pp. 109867, 2024.
- [23] Lukáš Klein, Ivan Zelinka, and David Seidl, "Optimizing parameters in swarm intelligence using reinforcement learning: An application of Proximal Policy Optimization to the iSOMA algorithm," *Swarm and Evolutionary Computation*, vol. 85, pp. 101487, 2024.
- [24] Li Sheng Kong, Muhammed Basheer Jasser, Samuel-Soma M. Ajibade, and Ali Wagdy Mohamed, "A systematic review on software reliability prediction via swarm intelligence algorithms," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 7, pp. 102132, 2024.
- [25] Xianfang Liu, "Mathematical scheduling model of complex industrial process combining swarm intelligence algorithm and swarm dimension reduction technology," *Results in Engineering*, vol. 21, pp. 101796, 2024.
- [26] Qirat Nizamani et al., "Nature-inspired swarm intelligence algorithms for optimal distributed generation allocation: A comprehensive review for minimizing power losses in distribution networks," *Alexandria Engineering Journal*, vol. 105, pp. 692-723, 2024.
- [27] Haoxin Wang and Libao Shi, "A multi-direction guided mutation-driven stable swarm intelligence algorithm with translation and rotation invariance for global optimization," *Applied Soft Computing*, vol. 159, pp. 111614, 2024.
- [28] Ahmad Alferidi, Mohammed Alsolami, Badr Lami, and Sami Ben Slama, "Design and implementation of an indoor environment management system using a deep reinforcement learning approach," *Ain Shams Engineering Journal*, vol. 14, no. 11, pp. 102534, 2023.
- [29] Frank Bodendorf and Jörg Franke, "Synthesis of activity-based costing and deep learning to support cost management: A case study in the automotive industry," *Computers & Industrial Engineering*, vol. 196, pp. 110449, 2024.
- [30] Jiaxin Chen, Xiaolin Tang, and Kai Yang, "A unified benchmark for deep reinforcement learning-based energy management: Novel training ideas with the unweighted reward," *Energy*, vol. 307, pp. 132687, 2024.

Enhancing Road Safety: A Multi-Modal Drowsiness Detection System for Drivers

Guirrou Hamza¹, Mohamed Zeriab Es-Sadek², Youssef Taher³

ENSAM, Mohammed V University Rabat, Morocco^{1,2}

Centre of Guidance and Planning of Education Rabat, Morocco³

Abstract—Driver drowsiness is a major contributing factor in road accidents, emphasizing the need for enhanced detection measures to improve car safety. This paper describes a multi-modal fatigue detection system that uses data from an internal camera, a front camera, and vehicle factors to reliably assess driver alertness. The technology outperforms traditional methods in terms of detection accuracy by utilizing powerful machine learning algorithms. Simulation and real-world tests show considerable improvements in reliability and performance. This integrated strategy offers a promising alternative for reducing the dangers associated with driver weariness and improving overall traffic safety.

Keywords—Component; fatigue detection; drowsiness monitoring; ADAS

I. INTRODUCTION

The rising number of traffic accidents due to driver drowsiness poses a significant threat to worldwide vehicle safety. Drowsiness decreases reaction times, alertness, and decision-making abilities, potentially leading to serious consequences. According to World Health Organization study, drowsy drivers cause up to 30% of road accidents [1]. Despite advances in car safety systems, detecting and reducing driver drowsiness remains a major concern. Existing systems frequently rely on a single data source, such as driver monitoring cameras or vehicle behaviour analysis, which may not provide a complete picture of the driver's state.

This study overcomes this limitation by creating a multi-modal sleepiness detection system that combines data from an interior camera, a front camera, and a variety of vehicle factors. The inside camera catches the driver's facial expressions and eye movements, which provide direct indications of weariness. The front camera detects the vehicle's position relative to road markings and other cars, providing contextual information about the driving environment. Furthermore, vehicle data, such as steering patterns, speed variations, and lane deviations, contribute to a thorough evaluation of driver behavior and potential sleepiness signs.

By combining these disparate data streams, the proposed system intends to improve the accuracy and reliability of sleepiness detection, thus enhancing overall traffic safety. This paper describes how the integrated system was designed, implemented, and evaluated. It begins with an overview of sleepiness labeling and monitoring technologies, followed by data collection and specification in accordance with norms and regulations. Next, the system architecture and feature extraction algorithms are given. The report also explains the detection

algorithms used to process data and generate alerts, as well as the system integration and real-time operation techniques.

The system's performance is confirmed by simulation and real-world testing, which show considerable gains in detection accuracy over typical single-modality systems. The findings highlight the system's potential to improve automobile safety and enable better driving experiences. Finally, the consequences of these findings for vehicle safety are examined, and ideas for further research are suggested.

II. INSTRUMENTATION AND DATA COLLECTION

A. Drowsiness Labeling

Drowsiness is classified in a variety of ways, including subjective and objective measures. The Karolinska Sleepiness Scale (KSS) is the most commonly used tool. Participants rated their drowsiness on a scale of 1 (very awake) to 9 (extremely drowsy, fighting sleep). The KSS is known for its simplicity and rigorous validation, making it a dependable tool in both research and therapeutic settings. Another popular subjective measure is the Stanford Sleepiness Scale (SSS), which asks people to score their sleepiness on a scale of 1 (feeling active and vital) to 7 (no longer fighting sleep, sleep onset imminent) [2].

Other scales are the HFC Drowsiness Scale, Epworth Sleepiness Scale (ESS), Johns Drowsiness Scale (JDS), Observer Rating of Drowsiness (ORD), and Subjective Drowsiness Rating (SDR). The HFC sleepiness Scale, ORD, and SDR use external labelling methods, in which an observer assesses the subject's sleepiness based on observable behaviors and physical indications. In contrast, the ESS is a self-administered questionnaire that assesses an individual's proclivity to fall asleep in a variety of scenarios, providing a total score indicative of general daytime drowsiness. The JDS is unique in that it relies on physiological signs, such as eye movements, blink rate, or brain activity, to objectively quantify drowsiness.

The KSS was chosen as the major instrument for sleepiness labelling in this study because of its ease of use and solid validation record. The KSS allows participants to easily and reliably self-assess their state of drowsiness, giving a strong and trustworthy indicator to support the study's aims.

B. Drowsiness Monitoring

Driver sleepiness detection strategies include a wide range of physiological [3], behavioral [4], and vehicle-based approaches [5], each with differing levels of intrusion and accuracy. Physiological approaches such as

electroencephalography (EEG), electrocardiography (ECG), electromyography (EMG), and electrooculography (EOG) are highly accurate because they detect early signs of drowsiness. In this study, ECG data will be collected with self-reported participant assessments to provide precise baseline measures of sleepiness levels in a controlled environment. This combination provides an excellent paradigm for evaluating drowsiness using both subjective and objective inputs.

The fundamental goal of this research is to combine behavioral and vehicle-based detection algorithms to increase overall accuracy and solve edge circumstances that may be difficult for a single modality. Eye movements, facial emotions, and head position are among the indications used in behavioral analysis. Although these procedures are less physically intrusive than physiological measures, their relationship with monitoring raises issues about psychological intrusiveness. Typically, behavioural detection uses cameras and deep learning algorithms to diagnose sleepiness states based on data including blink duration, blink frequency, percentage of eyelid closure (PERCLOS), yawning, and head posture. Eye movement-based tests are especially effective because of their high association with tiredness.

Vehicle-based approaches are used in addition to behavioural detection to capture driving information such as lane position, steering wheel movements, acceleration patterns, and pedal usage. Steering wheel movement and lane-keeping performance are commonly investigated measurements, with conflicting results about their relative accuracy in diagnosing drowsiness. Vehicle-based procedures are most effective in locations with clear road markings and favorable weather conditions, but they are often less dependable than physiological or behavioral measures when used alone.

The combination of behavioral and vehicle-based methods takes advantage of the strengths of both approaches. Behavioural methods provide extensive, real-time insights into the driver's state by continuously monitoring facial and ocular traits, whereas vehicle-based methods provide a practical, non-intrusive way of evaluating driving performance. By combining these modalities, the proposed method improves the resilience and diversity of sleepiness detection while balancing accuracy, intrusiveness, and practicality. This hybrid technique is intended to provide a comprehensive solution fit for real-world applications, effectively tackling a wide range of scenarios and edge cases.

C. Data Collection

To improve data collecting for sleepiness detection, a comprehensive technique was used to gain a holistic picture of driver alertness. ECG was used to monitor participants' heart

rate and heart rate variability, providing important information about their physiological status. Participants used the KSS to self-report their sleepiness levels, allowing subjective fatigue ratings to be correlated with physiological data.

Front and interior cameras were carefully placed to capture the vehicle's position in the lane, as well as facial expressions, eye movements, and head posture, allowing for thorough behavioral analysis. Furthermore, vehicle data were tracked via the Controller Area Network (CAN) technology, which captured crucial driving metrics like steering wheel movements, lane deviations, and pedal usage. The obtained data is utilized to train the model, validate it, and calculate the system's performance.

D. Norms and Regulations

To maintain safety and dependability, drowsiness and distraction detection systems in the automotive sector must meet high standards. Euro NCAP (European New Car Assessment Program) [6] is a major regulatory body that provides rigorous methods for evaluating the performance of advanced driver assistance systems (ADAS). Euro NCAP evaluates these systems on their ability to detect and reduce risks associated with driver fatigue and distraction, which plays an important part in establishing vehicle safety ratings. These studies include extensive assessments of the system's response to real-time sleepiness signs, accuracy in detecting distractions, and overall reliability under varied driving scenarios.

Furthermore, the European Union's General Safety Regulation 2 (GSR2) mandates all new vehicles to have driver monitoring devices that can identify both tiredness and distraction [7]. GSR2 requires that these systems meet high precision, reliability, and user data protection standards in order to improve road safety.

We created software and system requirements in accordance with the automobile safety standard ISO 26262 [8] and the applicable regulation. Furthermore, we measured performance and validated it in accordance with the defined requirements and applicable norms.

III. OPERATIONAL DECISION MODEL

A. System Architecture

The suggested multi-modal sleepiness detection system is intended to use the strengths of several data sources to deliver a complete assessment of driver weariness. The system architecture is separated into four major components: data collecting, processing, analysis, and alarm production. Each component is critical to guaranteeing the accuracy and reliability of the sleepiness detection procedure. Fig. 1 depicts the major components of our multi-model driver drowsiness detection system (DDAS):

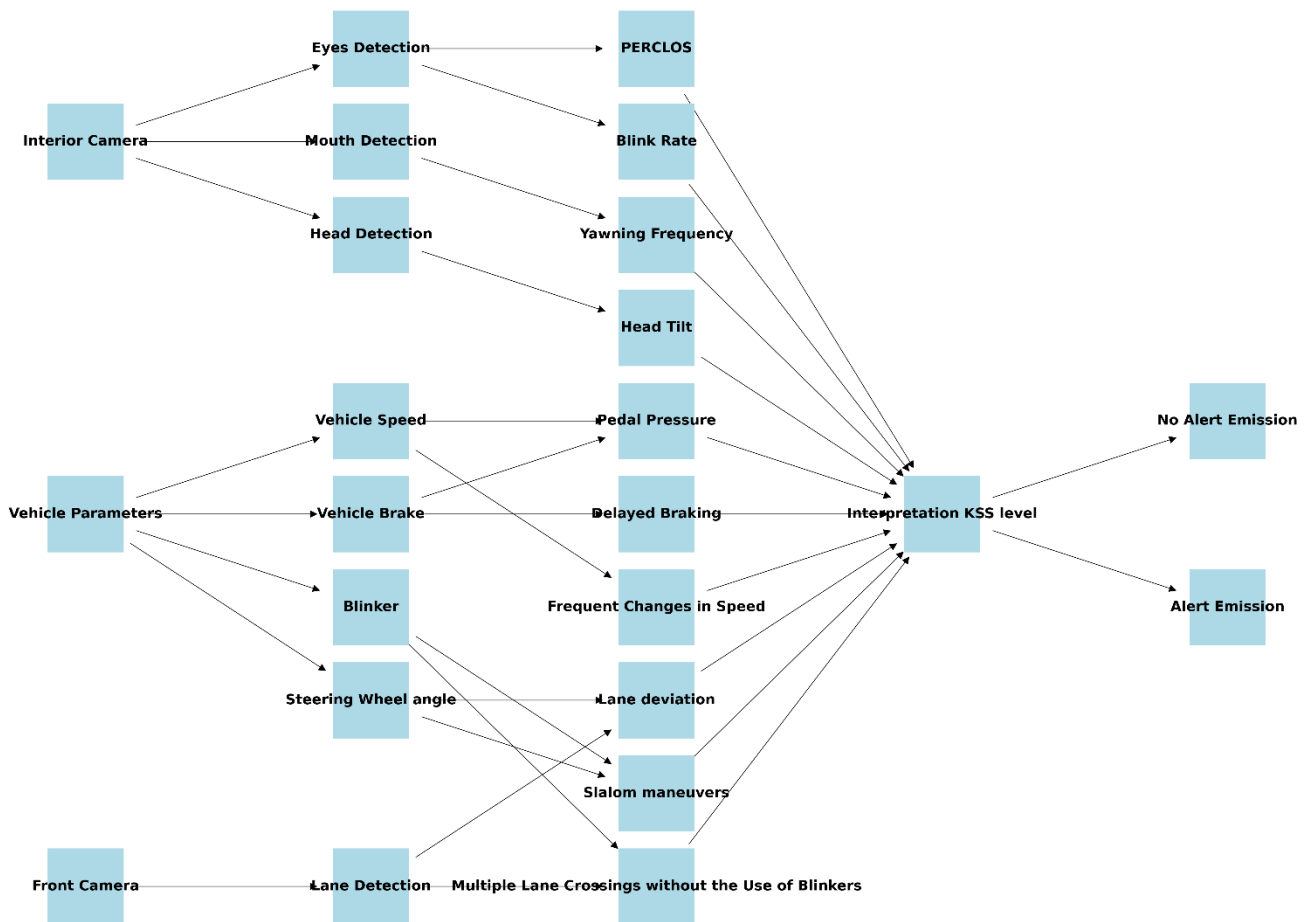


Fig. 1. Multi-modal driver drowsiness detection system framework.

The data acquisition device uses three basic sources: an internal camera, a front camera, and vehicle parameters. The interior camera is set up to capture the driver's facial expressions and eye movements. The front camera, positioned behind the rearview mirror and facing the road, captures real-time imagery of the road ahead. Vehicle parameters are obtained from the CAN network.

Data processing entails extracting and identifying unique properties from each data source. The internal camera identifies face landmarks such as eyes, mouth, and head orientation. These traits are identified and tracked using methods such as facial recognition and feature point detection. The front camera identifies lane markers and the vehicle's relative distance to the lane, while line detection algorithms track lane deviations and headway distance. Meanwhile, the vehicle parameters subsystem gathers information on steering wheel movements, speed variations, brake pedal pressure, and acceleration patterns.

The data analysis unit examines the retrieved features to determine the driver's state. Blink frequency, duration of eye closure (PERCLOS), and yawning frequency are all examples of fatigue indicators in interior camera footage. High blink rates and prolonged eye closures are clear signs of tiredness. The data

from the forward-facing camera is utilized to calculate lane deviation frequency, time-to-lane crossing, and headway. Frequent lane drifting, crossing lines, without indicating and maintaining a low headway distance can all indicate a lack of attention. Vehicle parameters are examined to identify slalom motions, speed abnormalities, and erratic braking. Sudden steering wheel movements, irregular speeds, and abrupt brakes all indicate driver weariness or inattention. To process these features and identify drowsiness-related patterns, an advanced machine learning technique called support vector machine (SVM) is used. These models are trained using annotated datasets to distinguish between normal driving behavior and indicators of weariness.

When drowsiness is identified, the system issues relevant alerts to the driver. These alerts include both voice notifications and visual cautions on the dashboard. These warnings are intended to catch the driver's attention and encourage them to take a rest.

B. Features Extraction and Decision Making

1) *Lane detection and deviation:* The front camera identifies lane deviations, indicating irregular driving behavior.

The video stream is preprocessed to determine the road's region of interest (ROI), then edge detection is performed using the Canny edge detector:

$$G(x, y) = \sqrt{G_x^2 + G_y^2}$$

where, $G(x, y)$ is the gradient magnitude at pixel (x, y) , and G_x, G_y are horizontal and vertical gradients, respectively. The identified edges are converted into line segments using the Hough Transform:

$$\rho = x \cos \theta + y \sin \theta$$

where (x, y) are edge points in the image, ρ is the perpendicular distance from the origin to the line, and θ is the line's angle.

Lane position deviations are measured as the Standard Deviation of Lane Position (SDLP):

$$SDLP = \sqrt{\left(\frac{1}{n} \sum_{i=1}^n (\rho_i - \bar{\rho})^2\right)}$$

where ρ_i is the lateral position of the vehicle at time i , and $\bar{\rho}$ is the average lane position across the observation window [9].

2) *PERCLOS (Percentage of Eye Closure)*: PERCLOS quantifies the proportion of time that the eyes are closed during an observation session:

$$PERCLOS = \frac{N_{closed}}{N_{total}}$$

Where:

The term N_{closed} = Number refers to the number of frames with an Eye Aspect Ratio (EAR) below the threshold, which indicates closed eyes.

N_{total} is the total number of frames captured during the observation time.

The Eye Aspect Ratio (EAR) for each frame is determined as:

$$EAR = \frac{dist(p_2, p_6) + dist(p_3, p_5)}{2 \cdot dist(p_1, p_4)}$$

Where:

p_1, p_2, \dots, p_6 = Coordinates for eye landmarks [10].

3) *Blink rate*: Blink rate is the number of blinks per minute, calculated as:

$$Blink Rate = \frac{N_{blinks}}{T} \times 60$$

Where:

N_{blinks} = Number of detected blinks during the observation period.

T = Duration of the observation period in seconds [11].

4) *Yawning frequency*: Yawning frequency refers to the number of yawns each minute:

$$Yawning Frequency = \frac{N_{yawns}}{T} \times 60$$

Where:

N_{yawns} is the number of yawns identified throughout the observation time.

T represents the duration of the observation period in seconds. [12]

5) *Head tilt*: Head tilt angle (θ) is measured by measuring the vertical and horizontal distances between specified facial landmarks:

$$\theta = \arctan\left(\frac{dist(p_{nose}, p_{chin})}{dist(p_{eye_corner_left}, p_{eye_corner_right})}\right)$$

Where:

p_{nose} = Landmark for the tip of nose.

p_{chin} = Chin landmark.

$p_{eye_corner_left}$ and $p_{eye_corner_right}$ = Landmarks for the outer corners of the left and right eyes [13].

6) *Slalom*: To identify slaloming using steering wheel angle ($\alpha(t)$), examine the rate of change ($\alpha'(t) = \frac{d\alpha}{dt}$) for sudden adjustments. Calculate the frequency of oscillations (f_{steer}) using the Fourier Transform of $\alpha(t)$.

we calculate the amplitude of oscillations ($A_{steer} = \frac{1}{T} \frac{d\alpha}{dt} \int_{t_0}^{t_0+T} |\alpha(t)| dt$) to reflect the magnitude of steering changes. A composite Slaloming Index (SI) comprises the following metrics:

$$SI = w_1 \cdot f_{steer} + w_1 \cdot A_{steer}$$

High SI values indicate slaloming without using the blinker, indicating erratic steering that could be attributed to fatigue [14].

7) *SVM-Based model*

a) *Feature vector*: The feature vector x combines all necessary parameters for detecting tiredness:

$x = [\text{PERCLOS}, \text{Blink Rate}, \text{Yawning Frequency}, \text{Head Tilt}, \text{Pedal Pressure}, \text{Delayed Braking}, \text{Frequent Change in Speed}, \text{Lane Deviation}, \text{Slalom Maneuver}, \text{Multiple Lane Crossings without Blinker}]$

b) *SVM decision function*: The SVM decision function is defined as follows:

$$f(x) = w \cdot x + b$$

Where:

$W = [W_1, W_1, \dots, W_{10}]$ Weight vector for the features.

b : Bias word.

$f(x)$: A decision score that indicates the possibility of drowsiness.

Classification Rule

The categorization choice depends on the sign of $f(x)$:

$$Y = \sin f(x)$$

Where:

Y = +1 indicates Non-Drowsy.

Y = -1 indicates Drowsy.

c) *Mapping to KSS level*: The decision score $f(x)$ is mapped to the KSS level using the mapping function $g(f(x))$:

$$K = g(f(x))$$

Where:

KSS levels range from 1 (high alertness) to 9 (high drowsiness).

d) *KSS-based decision outcomes*: The system action is based on the KSS level K:

$$\text{State} = \begin{cases} \text{No Alert, if } K < 7 \\ \text{Alert, if } K \geq 7 \end{cases}$$

e) *SVM training objective*: The SVM is trained by minimizing the following objective function:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i$$

subject to:

$$y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0$$

The regularization parameter C is used to balance the trade-off between maximizing margin and minimizing misclassification errors.

ξ_i : Slack variables for misclassified data points [15][16].

C. Operational Phases of DDAS

1) *Initialization of DDAS*: The DDAS is activated under the following conditions: the engine is turned on, the driver is present with the door closed and the seatbelt buckled, and no malfunctions are identified in the monitored parameters.

2) *Learning phase*: The DDAS learning phase occurs once every driving session, assuming the sleepiness function is engaged. This phase begins immediately upon system activation and lasts one minute, during which the system assesses the driver's sleepiness level. If a driver change is detected, the system resets and begins a new learning period. After completing the learning phase, the system moves on to the monitoring phase.

3) *Monitoring conditions*: The learning phase is reset every time the vehicle is started or a driver change is detected. To accurately detect tiredness, the vehicle must travel at a minimum speed of 50 km/h.

4) *Monitoring phase*: Following the learning phase, the DDAS enters the monitoring phase, which continues until the engine is turned off. During this phase, notifications are disabled in certain situations, such as when the vehicle's speed falls below 50 km/h. The technology continuously detects the driver's fatigue level.

IV. VALIDATION AND PERFORMANCE ANALYSIS

A. Simulation Based Result

The proposed DDAS's performance was evaluated through a series of video simulations that analyzed video footage acquired during the data collecting phase to detect sleepiness occurrences based on predetermined thresholds and metrics. To evaluate the system's usefulness and accuracy, a confusion matrix was created, which provided a detailed breakdown of the system's classification.

The video simulations included a diverse collection of lighting conditions, face angles, and subject sleepiness levels. The system's outputs were recorded and compared to the ground truth labels. The system's overall detection accuracy was assessed to be 92%, proving its capacity to discern between drowsy and alert states. In particular, the system properly identified alert states (True Negatives) in 94% of cases while accurately detecting drowsiness (True Positives) in 86%. However, it misidentified alert states as drowsy (False Positives) in 8% of cases and failed to detect drowsiness (False Negatives) in 6%. Fig. 2 displays the confusion matrix based on a huge dataset injected:

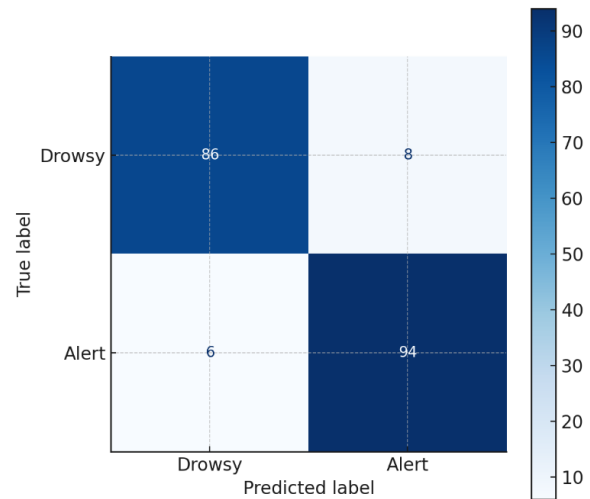


Fig. 1. DDAM confusion matrix.

The system demonstrated great sensitivity in identifying drowsiness, which is crucial for timely intervention and accident avoidance. The relatively low False Negative rate demonstrates its effectiveness in reducing unnoticed drowsiness. However, the False Positive rate, while acceptable, implies that further refinement in feature extraction and threshold tweaking could help eliminate unwanted warnings, hence improving user experience.

The system's performance was further tested under difficult conditions. Due to limited visibility of facial landmarks in low-light conditions, accuracy dropped somewhat to 88%. Under severe angles, accuracy remained stable at 90%, thanks to improved preprocessing and feature normalization. During rapid head movements, there was a modest decrease in True Positive detection, indicating an area for improvement in motion compensation.

B. Real-Condition Testing

Real-world testing in operating settings was carried out to validate the suggested sleepiness detection system. The

technology was placed in a vehicle and tested with drivers doing typical driving tasks. The scenarios included changing lighting (daylight, dusk, and night), different road surroundings (urban and highway), and dynamic driver behaviors. During these experiments, the system tracked and evaluated the driver's facial expressions, blinking patterns, and head movements to detect drowsiness in real time. The real condition testing findings are reported in the confusion Table I below:

TABLE I. REAL TIME DRIVING RESULTS

Predicted \ Actual	Drowsy	Alert
Drowsy	TP: 82%	FP: 12%
Alert	FN: 10%	TN: 88%

The system had an overall detection accuracy of 85%. It recognized drowsiness (True Positives) in 82% of cases and accurately identified alert states (True Negatives) in 88% of cases. However, it misclassified alert states as drowsy (false positives) in 12% of cases and failed to detect tiredness (false negatives) in 10% of cases.

C. Results, Discussion and Future Work

The testing findings show that the suggested DDAS works reliably under both simulation and real-world settings. The system's exceptional sensitivity in detecting drowsiness provides quick intervention, which is crucial for avoiding accidents. Furthermore, the comparatively low False Negative rate demonstrates its usefulness in reducing undetected drowsiness, an important feature of driver safety systems. The system's overall accuracy, especially in difficult settings like low light and severe facial angles, demonstrates its durability and adaptability to real-world applications.

These favorable results suggest that the system meets the safety and performance requirements stipulated in the GSR2 regulatory frameworks and Euro NCAP standards. Compliance with these standards demonstrates the system's ability to greatly improve driver safety and reduce traffic deaths. Its capacity to identify tiredness with high reliability is consistent with the growing emphasis on integrating advanced driver monitoring systems into vehicle safety regulations.

While the system worked effectively, there is still room for improvement. Future work should focus on lowering the False Positive rate in order to improve the user experience and reduce unwanted notifications. Advanced techniques, such as deep learning-based facial analysis, multi-modal data integration, and feature extraction method improvement, may improve system performance. Furthermore, further field testing with a more diversified driver population would help generalize the system's usefulness across different demographics and driving circumstances.

The positive results demonstrate the system's suitability for real-world deployment, assuring compliance with international

safety standards while providing a dependable solution for improving driver safety and contributing to the growth of intelligent vehicle technology.

V. CONCLUSION

This study introduces a multi-modal detection system that combines interior and front cameras with vehicle parameters to improve drowsiness detection accuracy. Using powerful machine learning, the system achieves 92% accuracy in simulations and 85% in real-world tests, consistently diagnosing fatigue under varied settings.

REFERENCES

- [1] S. Saleem, Sneddon, Risk assessment of road traffic accidents related to sleepiness during driving: a systematic review. *EMHJ*. Vol. 28 No. 9 – 2022.
- [2] V.P. Martin, J. Taillard, J. Rubenstein, P. Philip, R. Lopez, J.A. Micoulaud-Franch. What do measurement tools tell us about sleepiness and 6 hypersomnolence in adults? Historical approaches and future 7 prospects. *Médecine du Sommeil*. Volume 19, Issue 4, December 2022, Pages 221-240.
- [3] A. Chowdhury, R. Shankaran, M. Kavakli, M. Haque. Sensor Applications and Physiological Features in Drivers' Drowsiness Detection: A Review. *IEEE Sensors Journal*, february 2018.
- [4] F. Kerkamm, D. Dengler, M. Eichler, D. Materzok-Köppen, L. Belz, F. A. Neumann, B. C. Zyriax, V. Harth, and M. Oldenburg. Measurement Methods of Fatigue, Sleepiness, and Sleep Behaviour Aboard Ships: A Systematic Review. *Public Health* 2022, 19, 120.
- [5] E. Perkins, C. Sitaula, M. Burke, F. Marzbanrad. Challenges of Driver Drowsiness Prediction: The Remaining Steps to Implementation. *IEEE*. 25 November 2022.
- [6] C. Cabutí, J. Jackson. Driver Drowsiness Metrics for DDAW and Euro NCAP. *IDIADA Automotive Technology*. 2 nd May 2023.
- [7] European Commission. General Safety Regulation. *IMCO Committee, European Parliament – 22 February 2022*.
- [8] Functional Safety Standard for Modern Road Vehicles. *ISO 26262: Novembre 2011*.
- [9] TS. Teoh, PP. Em, N. A. A. Aziz. Driver Drowsiness Detection Based on LiDARBased Road Boundary Detection. *IEEE*, 10 December 2024.
- [10] D. Sommer, M. Golz. Evaluation of PERCLOS based Current Fatigue Monitoring Technologies. *IEEE*. 11 November 2010.
- [11] Amna Rahman, Mehreen Sirshar, Aliya Khan. Real Time Drowsiness Detection using Eye Blink Monitoring. *IEEE*, 04 February 2016.
- [12] Z. Jie, M. Mahmoud, Q. Stafford-Fraser, P. Robinson, E. Dias, L. Skrypchuk. Analysis of yawning behaviour in spontaneous expressions of drowsy drivers. *IEEE*, 07 June 2018.
- [13] BLAIR P. GRUBB, DANIEL J. KOSINSKI, K. BOEHM, K. KIP. The Postural Orthostatic Tachycardia Syndrome: A Neurocardiogenic Variant Identified During Head-Up Tilt Table Testing. *THE POSTURAL ORTHOSTATIC TACHYCARDIA SYNDROME*, septembre 1997.
- [14] S. Buma, H. Kajino, J. Cho, T. Takahashi, Shun'ichi Doi. Analysis and Consideration of the Driver Motion according to the Rolling by Slalom Running. *Review of Automotive Engineering*, 2009.
- [15] H. Shuyan, Z. Gangtie. Driver drowsiness detection with eyelid related parameters by Support Vector Machine. Volume 36, Issue 4, May 2009.
- [16] G. Li, W. Chung. Detection of Driver Drowsiness Using Wavelet Analysis of Heart Rate Variability and a Support Vector Machine Classifier. *Sensors*, 2 December 2013.

Decoding Face Attributes: A Modified AlexNet Model with Emphasis on Correlation-Heterogeneity Relationship Between Facial Attributes

Abdelaali Benaiss¹, Otman Maarouf², Rachid El Ayachi³, Mohamed Biniz⁴, Mustapha Oujaoura⁵

Department of Computer Science-Faculty of Sciences and Technics-Laboratory TIAD,
Sultan Moulay Slimane University, BP: 592, Beni Mellal, Morocco^{1,3}

Department of Computer Science-Faculty of Sciences, Agadir Ibn Zohr University, BP 8106, Agadir, Morocco²
Department of Computer Science-Polydisciplinary Faculty-Laboratory LIMATI,

Sultan Moulay Slimane University, BP 523, Beni Mellal, Morocco⁴

Department of Computer Science, Network & Telecom (GIRT)-Laboratory of Informatics Mathematics & Communication Systems (MISCOM)-National School of Applied Sciences (ENSA), Cadi Ayyad University, BP 575, Safi, Morocco⁵

Abstract—Face attribute estimation has several applications in computer vision, biometric systems, face verification /identification and image retrieval. The performance of face attribute estimation has been improved by using machine learning algorithms. In recent years, most algorithms have addressed this problem in multiple binary problem. Specifically, CNN-based approaches, which we can divide them into two classes; shared features and parts-based approaches. In shared features approach, the model uses two types of CNNs: one for feature extraction succeed by another one, for attribute classification. In the parts-based approaches, the approaches split the face image into multiple parts according to the geometric position of each attribute and train a CNN model for each part of the face. However, the shared features approach can handle attributes correlation but ignored attribute heterogeneity and gain in training time. On the other hand, the parts-based approaches can handle attributes heterogeneity but ignore attributes correlation and need more time in the training set compared with a shared feature approach. In this work, we propose a face attribute estimation method, which combined shared features and a parts-based approach into one model. Our model splits the input face image into five parts: whole image part, face part, face upper part, lower part, and nose part. In the same manner, the face attributes are subdivided into five groups according to the geometric position in the face image. We train shared feature model for each part, and we proposed an algorithm for feature selection task followed by AdaBoost algorithm to handle attribute classification task. Through a set of experiments using the LFWA and IITM Face Emotion datasets, we demonstrate that our approach shows higher efficiency of face attribute estimation compared with the state-of-the art methods.

Keywords—Face attribute estimation; biometrics; Convolutional Neural Network (CNN); face verification; computer vision

I. INTRODUCTION

The face recognition systems are the most attractive topics in the biometrics systems because of their convenience, hygiene, and low cost. Specifically, the face as the objective signal can be acquired in a contactless manner without requiring any special equipment. Moreover, due to the multiple advantages provided by the face recognition system, these kinds of systems are the

most used in industry, combined with fingerprint-based systems as personal authentication for smartphones, security gates, payment services, computer human interaction, etc. In addition, the explosive development of Convolutional Neural Networks (CNNs) models, Graphic Units Process (GPU) material and opensource frameworks (e.g, Keras, PyTorch, Caffe, Dlib, etc), the Convolutional Neural Networks (CNNs) have replaced most traditional methods for face recognition problems. To further illustrate the versatility and application of convolutional neural networks (CNNs) in various domains, it is noteworthy to mention recent advancements in related fields. For instance, the work on automatic translation from English to Amazigh using transformer learning [1] demonstrates the adaptability of deep learning models in language processing tasks. Similarly, the recognition of Tifinagh characters using optimized convolutional neural networks [2] highlights the effectiveness of CNNs in character recognition tasks. These studies underscore the broad applicability and potential of CNNs, reinforcing their relevance in face recognition systems as well. To address the great demand for face recognition on face retrieval systems [3], video surveillance environments [4], and criminal investigation protocols [5]. There are multiples studies aimed at improving the performance of face recognition by improving the face attribute estimation.

For improving the face attributes task, the research community provides a number of face databases with their attributes like: CelebA [6] and LFWA [7] datasets, at each dataset, we have a number of face attributes (40/73 attributes) that describe the biological characteristics of the face (e.g ; color, shape and texture) or give information about a subject, weather it is wearing ornaments such as glasses or earrings or not. In addition, the works in study [8] and [9] make an assumption about relationship between face attributes based on the co-occurrence probabilities of two attributes in some databases. To illustrate, the attribute ‘Male’ has a high probability of attributes such as ‘Goatee’, while ‘Female’ has high probability of attributes such as ‘Heavy Makeup’ and ‘Wearing Lipstick’. On the contrary, the lower probability of occurrence for some attributes, the more likely those attributes are to be heterogenous. In the same manner, the works [10], [11], [12] and

[13] show that the relationship between attributes is based on the face parts. For example, 'Black Hair' and 'Blond Hair' are attributes related to 'Hair', which has a position in the upper face part and 'Big Nose' with 'Pointy Nose' are attributes related to 'Nose', which has a position in middle face part, where 'Double Chin' and 'Goatee' are attributes related to the low face part.

In general, face attribute estimation approaches consist of three processes: face detection, feature extraction and attribute classification. Among these processes, feature extraction and attributes classification are the most important, since they have the greatest impact on the estimation accuracy. To deal with the challenging problem of feature extraction and attribute classification, multiple works have been published about this object. In all published works, specifically CNN-based methods, they have a significant impact on feature extraction in terms of accuracy. Some works like in [14] and [10], use the same features for estimating multiple attributes without considering the attribute heterogeneity, and some others are limited to estimating a single attribute [15], or training a separate model for each face attribute without considering the attribute correlation [3]. Although, all previous works have a challenging problem when dealing with non-frontal face images, low image quality, occlusion, and pose variations. The region of interest (ROI) is often suitable and it may cover only a small part of the image, while the face image is dominated by the effects of position pose and viewpoint. Since then, Part-based methods in [5], [10], [11] and [16] have recently become the most popular approaches to dealing with pose variation, occlusion and low image quality.

In this paper, we present a novel method which combined three principal approaches; multi-task, part-based and attributes relationship to achieves better results on face attributes estimation. However, the contributions of this research aim to present;

- 1) Image denoising and online data augmentation with a specific technic, which get the experimental condition close to real-world scenarios.
- 2) Data balanced process to handle the challenge of minority class in databases.
- 3) Face split process combined with attributes subgrouping to handle respectively; head pose and attributes heterogeneity in same task.
- 4) An algorithm to handle feature selection step, followed by Adaboost classifier to addressed the attributes correlation and achieved the final estimation of each attribute.

The remainder of this paper is organized as follows; Section II gives a brief overview of the related work on face attributes estimation. The proposed approach with the baseline networks is outlined in Section III. Section IV displays the different experiments that have been conducted to achieve better results than competing methods. A discussion of the results is addressed in Section V. Finally, we conclude our work in Section VI.

II. RELATED WORK

A. Multi-Tasks Learning Approaches

Multi-task learning (MTL) in facial attribute estimation consists of training a model to achieve multiple attribute

prediction using, in some cases, shared representation approaches. For instance, the work in study [6], proposes two CNNs; LNet and ANet; the first one is pre-trained for face localization and the second one is pre-trained for attribute prediction. Those two CNNs are succeeded by an SVM classifier for attribute classification. Also, the approach in study [16] proposed Deep Multi-task Learning (DMTL) network consists of learning a modified AlexNet with a batch normalization (BN) layer inserted after each Conv Layer for the shared part of model, followed by, a category-specification block for attributes estimation. This method can handle heterogeneous information about attributes. In addition, this shows superior performance compared to state-of-the-art methods on public benchmarks. Moreover, another architecture is designed in study [9], where a ResNet50 Network is adapted as the backbone architecture, they take the first 46 layer as Shared Layers succeed by Task-specific Layers consist of two branches res5C1 and res5C1 corresponding to smile and gender prediction (res5C1and res5C2 are two Residual Blocks of ResNet50). Those two branches are passed by attention block (coined as Att_C) to represent the dependency between smile and gender attributes. Furthermore, two novel approaches have been proposed in study [8], the first one called HyperFace uses AlexNet as backbone of model, while the second one called HyperFace-ResNet is based on ResNet-101. Those two architectures perform well in the face detection, landmarks localization, pose estimation and gender recognition on various public available unconstrained datasets, those approaches show a novel hypothesis about fusing the intermediate layers of the backbone structure. In other words, the introduction of multi-tasks learning approaches in CNN-based models shows better results compared with single-task learning approaches in the terms of attribute correlation.

B. Parts-Based Approaches

In parts-based approaches, the object image and face image are split into small parts and each part is taken as input of the feature extraction process. For example, the work in study [10], proposed Pose Aligned Networks for Deep Attribute Modeling (PANDA), which divide person image into small parts coined as Poselets and each one is passed by a trained CNN. The top-level activations of all CNNs are concatenated to obtain a pose-normalized deep representation and then Linear SVM classifier is trained for attribute classification. In some recent works, the facial attributes are clustered in many groups according to their location in facial parts and relationship in terms of correlation and heterogeneity, before the process of feature extraction and attribute classification. However, in study [5] they split the 40 face attributes into six or nine groups, and they proposed a multi-task deep CNN (MCNN) combined with an auxiliary network (AUX) to achieve the final estimation for each attribute group. Based on the MCNN-AUX model, the study [11] proposed Partially Shared Multi-task CNN (PS-CNN), since they split all the 40 attributes into four attribute groups including Upper, Middle, Lower, and Whole Image. Then the attributes classification of each group can be considered as individual attribute learning task. This method shows the promise results on two databases CelebA [6] and LFWA [7]. In short, parts-based approaches still the better approaches to addressing heterogenous attributes and face parts occlusion.

C. Attributes Grouping

Inspired by the fact that information contained in a face image is geometrically distributed on face parts like; eyes, nose, mouth, etc. For this reason, many recent works are based on this assumption. They have been adding attribute grouping approach as pre-processing step before feature extraction in attribute estimation process. For example, the work in study [8] split 40 attributes of LFWA and CelebA dataset into six subgroups based on attribute color and texture. This approach shows better results compared with other work like PANDA [10] with more than 10% of improvement in term of accuracy detection. Besides, in [5] the attributes have been separated into nine groups; Gender group, Nose group, Eyes group, Face group, AroundHead group, FacialHair group, Cheecks group and Fat group, each one contains a number of attributes according to their facial parts. Furthermore, the Faceness-Net approach proposed in study [12] separate attributes of CelebA dataset into five groups, similarly to [5] and [8]. This work proposes: Hair group, Eye group, Nose group, Mouth group and a Beard group. Each group is represented by branch in the model, those branches was summed into a face label map, which clearly suggests face's location in the image. Moreover, The Faceness-net approach shows better results compared to state-of-the-art methods in term of face detection problems (near 98.05 % AP = average precision on AFW dataset and 92.11 % AP on PASCAL dataset). Therefore, in [6] the subject face in dataset has been split into 14 parts and in the same manner, the 40 face attributes have been separated into 14 subsets. Each part of the face has a subset of attributes based on the visibility value of this attributes in this part of face (visibility value $> \tau$ = visibility threshold). In short, the attribute grouping step handles the attribute correlation and reduces model training time and hyperparameters, but still handles the attribute heterogeneity. On the other hand, in study [17] the group of attributes is subdivided into four groups based on attribute categories (nominal, ordinal, holistic and local). The approach trains four sub-network types according to each subgroup of attributes. This approach can handle attribute heterogeneities compared with the previous works in studies [5], [6], [8], [12], [16], and shown better results on average accuracy reach 93% on CelebA dataset and 86.3% on LFWA dataset. In conclusion, all the previous works prove that the attributes grouping step as pre-processing step can guide the model to achieve expected results, in terms of attribute relationship.

On the whole, the goal of this work is to combine an attribute grouping approach, parts-based approach and multi-tasks learning in order to make a novel model that can handle attribute correlation and heterogeneity at the same time and reach better results compared with state-of-the-art methods.

III. PROPOSED METHOD

In the majority approach's, the first step consists of taking the crop face from image and normalized it based on facial landmarks or splitting it to small parts before attacking CNN structure. In two cases, the crop face has less information about some attributes like; Wear Hat, Bald and 5 O 'Clock Shadow. Since, we propose an approach consists of five CNNs models, one of them, specifically design to extracted the information about Hair, Face background and Neck, at the same time, the

other fours CNNs are specialized in facial details. More details about the proposed approaches are giving in following sections.

A. Face Split and Attributes Grouping

To split face image into three parts, we have used the position of facial keypoints return by MTCNN method presented in [17] (as shown in Fig. 1). The segmentation of face image and face attributes are pre-processing step of input image before it has been processed by CNNs models, those three parts are coined in this work as UP for upper-part, LP for lower-part and NP for nose-part indicated in Fig. 1, respectively, by (A), (B) and (C). (E) represent the face part.

In this approach, we have five CNNs coined as UP-Net (for upper part convolutional neural network), LP-Net (for lower part convolutional neural network), NP-Net (for nose part convolutional neural network), FP-Net (for face bounding box convolutional neural network) and WI-Net (for whole face image convolutional neural network). Each CNN take a specific input image to predicts a subset of specific attributes. Therefore, UP-Net take upper part of face as input and predict their corresponding attributes (attributes with index number #2, #4, #6, #13, #16 and #24), LP-Net take lower part of face as input to predict the attributes with index number (#7, #17, #22, #23, #25, #32 and #37). Where, NP-Net predict the attributes with index number (#8 and #28) according to nose part of face and FP-Net predict the attributes with index number (#5, #19, #20, #21, #26, #27, #30 and #40) which contain the information about face region. Finally, WI-Net receive whole face image as input and given the prediction of attributes with index number (#1, #3, #10, #11, #12, #14, #15, #18, #29, #31, #33, #34, #35, #36, #38 and #39). The label of each index number is described in Table I.

Indeed, certain segments are more effective at predicting a subset of attributes than others. For example, we can expect that upper-part (UP) would contain information about the person being bald, wearing hat or having certain types and color of hair. Therefore, this part of face can still predict attributes related to eyes, eyebrows and hair, at the same time, the lower-part (LP) can predict attributes related to mouth, goatee and moustache. Where, nose-part give information about nose. In short, we join each part with their corresponding attributes.

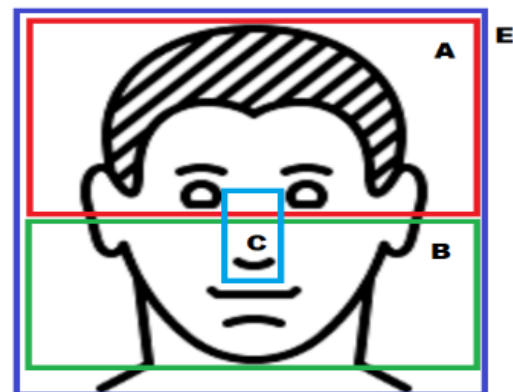


Fig. 1. The face image is divided into four parts; (E) face part, (A) upper part, (B) lower part, (C) nose part. The segmentation of face has been made based on keypoints return by MTCNN method in [17] (Best viewed in color).

TABLE I. FACE ATTRIBUTE LABELS DEFINED IN LFWA [7] DATASET

Index	Attribute	Index	Attribute
#1	5 O 'Clock Shadow	#21	Male
#2	Arched Eyebrows	#22	Mouth Slightly Open
#3	Attractive	#23	Mustache
#4	Bags Under Eyes	#24	Narrow Eyes
#5	Bald	#25	No Beard
#6	Bangs	#26	Oval Face
#7	Big Lips	#27	Pale Skin
#8	Big Nose	#28	Pointy Nose
#9	Black Hair	#29	Receding Hairline
#10	Blond Hair	#30	Rosy Cheeks
#11	Blurry	#31	Sideburns
#12	Brown Hair	#32	Smiling
#13	Bushy Eyebrows	#33	Straight Hair
#14	Chubby	#34	Wavy Hair
#15	Double Chin	#35	Wearing earrings
#16	Eyeglasses	#36	Wearing Hat
#17	Goatee	#37	Wearing Lipstick
#18	Gray Hair	#38	Wearing Necklace
#19	Heavy Makeup	#39	Wearing Necktie
#20	High Cheekbones	#40	Young

B. Attributes Correlation and Heterogeneity

The face verification and image search fields are the first methods has been introduced image attributes as descriptor. They used a subset of 40 binary attributes to describe each face in dataset (see Table I). They later extended the number of attributes with addition ones to achieve 73 binary attributes. In the recent years, more approach's shown attributes dependency and non-dependency [5] to improve accuracy of attributes detection. Further, as shown in the Fig. 2, the attribute with clear color square shown strong positive correlation between their two corresponding attributes respectively, in x-axis and y-axis. On the other hand, attribute with dark color square shown low correlation (heterogeneity). To draw the image in Fig. 2, we have been calculated co-occurrence matrix of each attribute index in LFWA dataset and each square in co-occurrence matrix present the probability to have a specific two attributes in same image. For example, the attribute No_beard (#25) has a strong correlation with Heavy_Makeup (#19), Wearing_Earrings (#35) and Wearing_Lipstick (#37) which has a probability more than 90% (shown by clear color square in Fig. 2) even though the attribute No_beard (#25) has weak correlation with Mustache (#23) which has a probability near 0% (shown by dark color square). In short, the proposed approach is motivated by previous assumptions and it has been introduced in the task specification process which well be detailed in following sections.

C. Network Architecture

Different parts of the face may have different signals for each attribute and sometimes signals coming from one part

cannot infer certain attributes accurately. For example, the information about nose like Big_Nose (#8) (in the Nose Part of face) cannot give information about Wearing_hat (#36) (in the top of head, Whole image part). Therefore, based on these assumptions, we have been explored the advantages of part-based approach to handle the non-dependency of attributes in the shared feature process of our network and we have been explored paire-wise co-occurrence matrix of attributes to handle dependency attributes in task specification process.

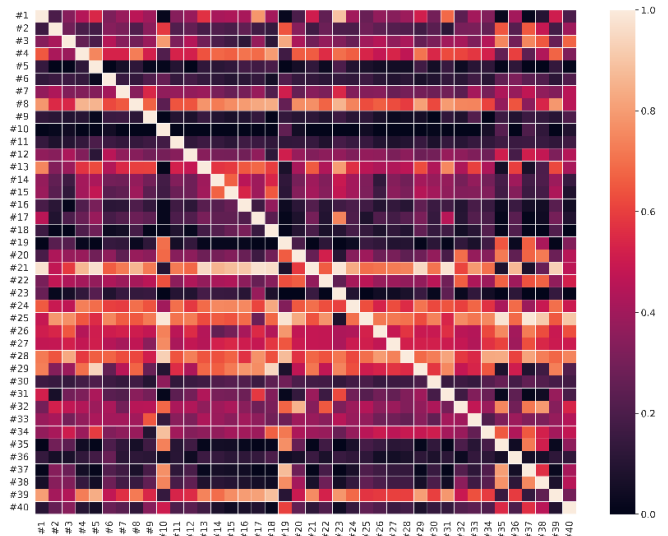


Fig. 2. Pair-wise co-occurrence matrix of the 40 face attributes (see Table I) provided with the LFWA [7] database (Best viewed in color).

Inspired from the works in [8] and [16], we have chosen to work with AlexNet as base structure of our model because it has a good result in the term of accuracy on challenging database and faster than GoogleNet and has a small structure compared with VGG and ResNet models. However, the AlexNet consists of five convolutional layers, three max pooling layer and the three Full connected layers. Based on work in study [18], we have been inserted a batch normalization layer after each convolution layer to avoiding overfitting problem. The original and modified AlexNet are shown in the Fig. 3.

Inserting Batch Normalization Layer to AlexNet model. Regularization stands out as an effective technique to combat overfitting issues. Incorporating Batch Normalization (BN) [14] between convolution layers can enhance model stability and regularization. The BN layer normalizes the output from the preceding activation layer by subtracting the mini-batch mean and dividing by its standard deviation. Essentially, this normalization process adjusts the means and variances of layer inputs by introducing two trainable parameters per layer. Additionally, BN reduces the network's sensitivity to the initialization of individual layers, enabling the use of higher learning rates. In our approach, we set a fixed learning rate of 0.001. The Fig. 3 shown more details about original and modified AlexNet model used in this work. The batch normalization layers add into AlexNet model are motioned in Fig. 3 by blue square. C shown the number of attributes returned by AlexNet block according to input facial part. The numbers denote the kernel size, cardinality and features maps for given layer.

Overall structure. The proposed approach consists of denoising process followed by image splitting process to get five parts (deemed as WI, FP, UP, LP and NP). Those five parts are resized into 224x224 size image whose have been taken as input of five subnets (coined as WI-Net, FP-Net, UP-Net, LP-Net and NP-Net). Each subnet of them has a modified AlexNet described in Fig. 3 as base structure.

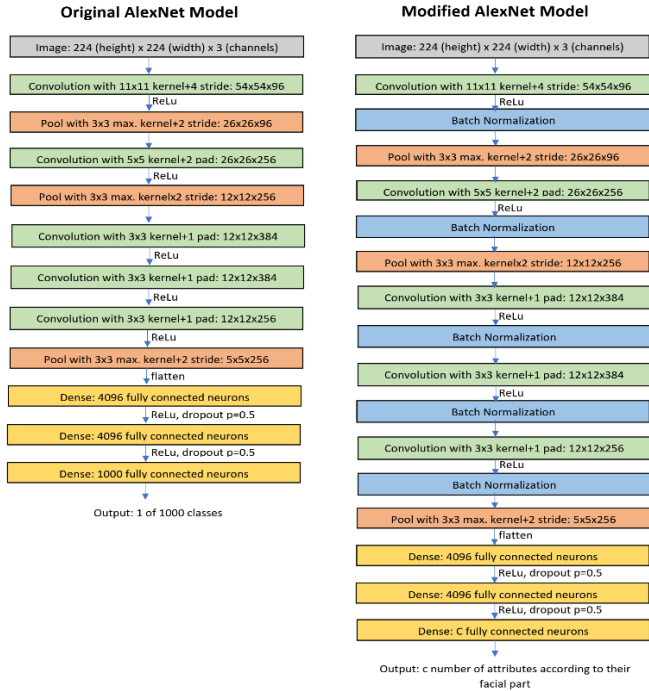


Fig. 3. The original and modified AlexNet used in backbone of our structure (Best viewed in color).

Since, the batch normalization (BN) layers add into AlexNet backbone adjusted the means and variations between the Conv. layers and make the main structure more stable in the learning process. Moreover, at the output of each subnet, we have a specific number of face attributes see Subsection A for more details (17 face attributes for WI, 8 for FP, 6 for UP, 7 for LP and 2 NP). To show the correlation between attributes, we have been proposed a coding algorithm repose on occurrence matrix to handle face attributes correlation see subsection E for more details. The feature selection has been followed by Adaboost classifier to achieve the final estimation for each attribute, see Fig. 4 to have an overview of our proposed method.

D. Simulation of Real-World Scenarios by Data Augmentation

In real-world, there are no universal patterns for facial attributes across all individuals and all datasets present in the literature are limited and don't accurately mirror real-world scenarios based on this assumption, a more realistic attributes estimation system should be trained on dataset prepared with data augmentation process to achieve real-world conditions.

Data augmentation, as discussed in study [19] and study [20], serves as a regularization technique by introducing synthetic images to the neural network, simulating more realistic conditions and viewpoints. This approach helps mitigate

overfitting issues stemming from limited datasets. Various transformations can be applied to create additional modified images, including translation, zooming, brightness adjustments, and more. These subtle variations enable the model to generalize better to unseen data and enhance its robustness when exposed to slightly altered images. In our study, we adopted online augmentation, applying transformations to images as batches are processed during training. This approach accelerates the training process and eliminates the need to store augmented data alongside the original dataset in memory, as required in offline augmentation as per the protocol in study [19]. Specifically, our data augmentation involved shifting image appearance by 0.2 of the image height, adjusting brightness within a range of values between 0.1 and 0.2, and applying zooming. Fig. 5 illustrates the results of the data augmentation process on an image from the LFWA [7] database. The image in the left side of Fig. 5 is the original image after denoising process which loaded from LFWA [7] dataset and the group of images in right side are the data augmentation images after different transformation functions.

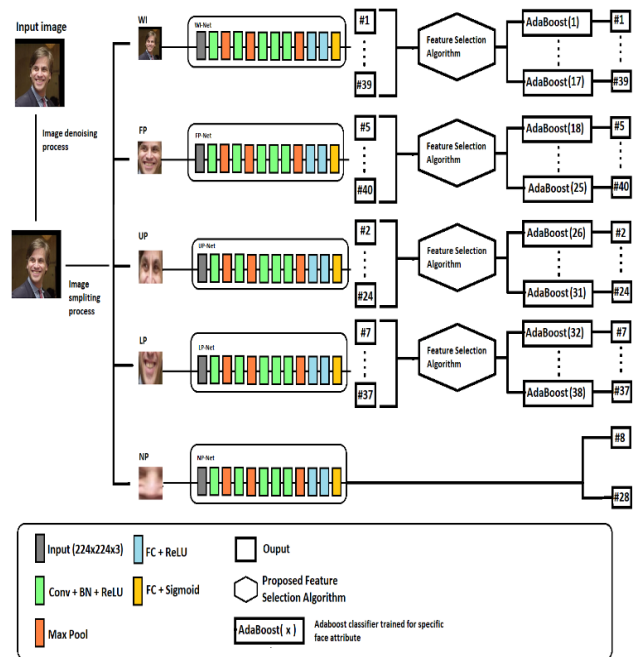


Fig. 4. Overview of the proposed architecture. (Best viewed in color).

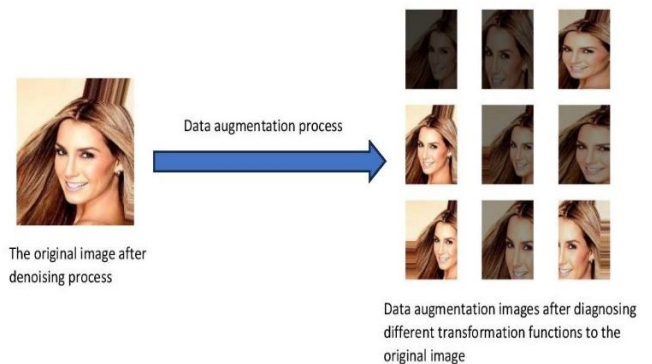


Fig. 5. Data augmentation process of each image in dataset (Best viewed in color).

In addressing this aspect of our methodology, we utilized the Keras ImageDataGenerator class, which offers a convenient and efficient method for image augmentation. This class provides a range of augmentation technics, including standardization, rotation, shifts, flips, brightness adjustments, and more. However, the primary advantage of employing the Keras ImageDataGenerator class is its capability for real-time data augmentation. This means it generates augmented images on-the-fly during the training phase of your model. In short, by utilizing this class, images are loaded in batches, which means saving more memory in training process.

E. Task-Specification Process

To achieve task-specification process, many approaches has been shown in the literature. For example, in study [6] an automatic attributes grouping method has been proposed which take columns of weights matrix returned fully-connected layer ANet as a decision hyperplane to partition the negative and positive samples of attribute. By sample applying k-means to these vectors, the clusters show clear grouping patterns. These are used as system features which passed by SVM to achieve final attribute detection. Furthermore, the work in study [21] used Multi-Kernels Maximum Mean Discrepancies (MK-MMD) proposed in study [22] to show the correlation between features returned by MNet and TNet. Another approach shown with PANDA [10] which take a signal returned by each poselet as pose normalization of each image part and used a linear classifier (Logistic Regression in this case) to achieve a Task-specification process. In addition, FaceTracer [15] used SVM algorithm which deemed by Attribute-Tuned Global SVM to achieve final attribute detection. In short, the task-specification process consists of two steps feature selection combined with an algorithm of classification.

In contrast with the previous approaches, we have been proposed an encoding algorithm for features selection step and AdaBoost algorithm to achieve final prediction of each attribute. The proposed algorithm has been presented in the following section and the based co-occurrence matrices of FP-Net, UP-Net and LP-Net have been presented respectively, by (a), (b) and (c) in Fig. 6. The proposed Algorithm used those matrices to achieve feature selection step.

Algorithm 1: LPBT ← (find list of attributes with probability greater than specific threshold)

```
Initialize
I ← Specific attribute
M ← Matrix of occurrence
T ← Threshold
N ← Number of rows in M
Compute
For i ← 0 to i ← N-1 do
  Update
  Update and analyze
  If M[I][i] >= T then
    L[C] ← i
    C ← C + 1
  End
End
End
```

Algorithm 2: CAGT ← Calculate the margin of error between list attributes return by subnet and ground truth

```
Initialize
I ← Specific attribute
L ← List returned by Algorithm 1 (LPBT function)
N ← lent of L list
T ← Matrix values return by subnet (WI-Net, FP-Net, UP-Net or LP-Net)
P ← Matrix of ground truth values of attributes group according to each subnet.
R ← Number of rows in T matrix
C ← Number of columns in T matrix
Compute
For i ← 0 to i ← N-1 do
  Update
  Update and analyze
  If M[I][i] >= T then
    L[C] ← i
    C ← C + 1
  End
End

For i ← 0 to i ← N-1 do
  Update
  Update and analyze
  TV[i] ← T[:,L[i]]
End

For i ← 0 to i ← R-1 do
  Update
  Update and analyze
  For j ← 0 to j ← C-1 do
    Update
    Update and analyze
    If TV[i][j] == 0 then
      TV[i][j] ← -1
    End
  End
End

For i ← 0 to i ← C-1 do
  Update
  Cpt ← 0
  Update and analyze
  For j ← 0 to j ← R-1 do
    Update
    Update and analyze
    Cpt ← Cpt + TV[j][i]
  End
  CV[i] ← Sigmoid(Cpt)
End
End
```

Algorithm 3: we have in output of this algorithm the list shown the relationship between attributes.

```

Initialize
M ← Co-occurrence matrix in the output of each subnet (WI-Net, FP-Net, UP-Net or LP-Net)
T ← Matrix values return by each subnet
P ← Ground truth matrix for each subnet
R ← Number of rows in T matrix
Compute
For i ← 0 to i ← R-1 do
Update
    Update and analyze
    Accu ← 0
    For j ← 0 to j ← 9 do
Update
        Update and analyze
        S ← j/10
        CO ← LPBT(i, M, S) /* Algorithm 1 */
        If Accu < CAGT(i, CO, T, P) then
            Accu ← CAGT(i, CO, T, P)
            Find ← CO /* Algorithm 2 */
        End
    End
    F[i] ← Find
End
End
    
```

Effectively, in the training set, we have been taken the output of each subnet and we have been applied this algorithm to show the correlation between face attributes. See Table II for more details.

TABLE II. ATTRIBUTES IN FORT CORRELATION WITH EACH ATTRIBUTE RETURNED BY PROPOSED ALGORITHM FOR FEATURE SELECTION STEP IN OUR MODEL

Attribute index	Attributes in correlation	Attribute index	Attributes in correlation
#1	#1, #29, #31, #33, #39	#21	#21, #26, #27
#2	#2, #4, #6, #13, #16, #24	#22	#22, #25, #32
#3	#3, #33, #34, #35, #38	#23	#7, #17, #23
#4	#4, #13, #24	#24	#4, #13, #24
#5	#5, #21, #26, #27	#25	#22, #25, #32
#6	#2, #6, #24	#26	#19, #21, #26, #27
#7	#7, #22, #25	#27	#19, #21, #26, #27, #40
#8	#8	#28	#28
#9	#3, #9, #33, #35, #38, #39	#29	#1, #14, #15, #18, #29, #31, #33, #34, #39
#10	#1, #3, #9, #10, #11, #12, #14, #15, #18, #29, #31, #33, #34, #35, #36, #38, #39	#30	#5, #19, #20, #21, #26, #27, #30, #40

#11	#11, #12, #29, #33, #34, #35	#31	#1, #31, #39
#12	#3, #12, #33, #35, #38	#32	#22, #25, #32
#13	#4, #13, #24	#33	#1, #9, #14, #15, #18, #29, #33, #38, #39
#14	#14, #15, #33	#34	#3, #34, #38
#15	#14, #15, #29, #33, #39	#35	#3, #35, #38
#16	#13, #16, #24	#36	#14, #33, #36, #39
#17	#7, #17, #23	#37	#7, #22, #25, #32, #37
#18	#18, #29, #38	#38	#3, #35, #38
#19	#19, #26, #27, #40	#39	#1, #15, #29, #33, #39
#20	#19, #27	#40	#19, #21, #26, #27, #40

Finally, the results returned by Algorithm for each attribute has been passed in the followed step by Adaboost classifier to achieve the final estimation.

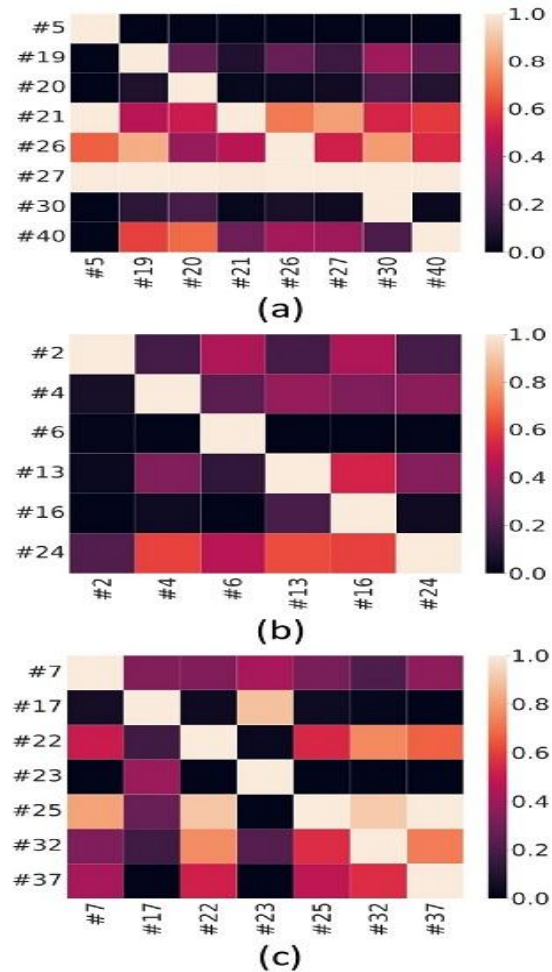


Fig. 6. The matrix of co-occurrence for each subnet in LFWA dataset. (Best viewed in color).

IV. EXPERIMENTAL RESULTS

In the experimental section, we have been shown all the experimental steps provide in this work. In first time, we have

been presented Evaluation Metrics used in this work (subsection A). In the followed subsection B, we have been described the databases used in this work to made training, validation and testing steps. The data pre-processing has been presented in subsection C, which contains image denoising, splitting and database balanced. Further, we have been shown the process to determine the values of some specific parameters of our networks in the subsection D. Finally, the performance of our approach for 40 face attributes, gender recognition and smile estimation have been shown in the subsection E.

A. Evaluation Metrics

The most common metrics for attributes estimation is Accuracy, Precision, Recall and Fi-score. Those metrics can be more representative than others metrics in du literature to evaluate the attributes estimation system, since it can be shown the difference between the estimated value and their ground truth in the statistics manner. Those metrics can be mathematically defined as following:

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - score = \frac{2*Precision*Recall}{Precision+Recall} \quad (4)$$

Where:

- True positive (TP): An instance for which both estimated and ground truth values are positive.
- True negative (TN): An instance for which both estimated and ground truth values are negative.
- False Positive (FP): An instance for which estimated value is positive but ground truth value is negative.
- False Negative (FN): An instance for which estimated value is negative but ground truth value is positive.

B. Datasets

LFWA dataset [7] is a large-scale face attribute database with 13143 images of 5.749 subjects in unconstrained environment. Each image is annotated with 40/73 attributes (see Table I more description). The images in this dataset are in color space and contain large variations in pose, expression, race, background, etc., making it challenging for face attribute estimation. Moreover, the split protocol suggested by dataset is 6263 for training and 6880 for testing. Some images from the dataset have been shown in Fig. 7.



Fig. 7. Samples from LFWA dataset (Best viewed in color).

More descriptions about LFWA dataset have been illustrated in Fig. 8.

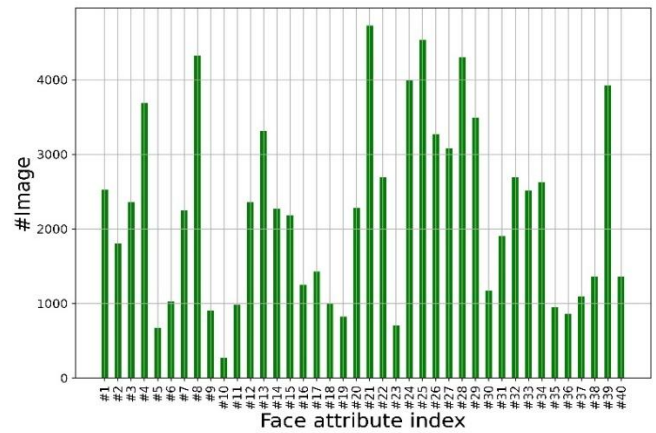


Fig. 8. Details about data distribution in LFWA dataset (Best viewed in color).

Strating from the Fig. 8, the LFWA dataset suffer from large imbalanced in data distribution. For example; number of man subjects in data reach 4727 when female subjects equal to 2153, which make gender estimation task harder for female subjects in compared with man subjects. In the same manner, smiling people in train part of dataset equal to 2687 when no smiling people reach 4193, which make this task hard in the training process. To deal with this problem, we have adopted SMOTE [23] algorithm to balance a training data for each subnet (WI-Net, FP-Net, UP-Net, LP-Net and NP-Net). More details have been described in the followed Subsection C.

The IITM Face Emotion dataset [24] originates from the IITM Face Data and comprises 1,928 images from 107 participants, including 87 males and 20 females. These images have been captured in three distinct vertical orientations (Front, Up, and Down) and has been featured six different facial expressions: Smile, Surprise, Surprise with Mouth Open, Neutral, Sad, and Yawning. The original IITM Face dataset includes additional attributes such as gender, presence of facial hair like mustaches and beards, eyeglasses, clothing, and hair density. For this study, the IITM Face dataset was adapted to focus on facial expressions across different orientations. The IITM Face Emotion dataset features only the facial region segmented for each subject, with all images resized to fixed dimensions of 800 x 1000 pixels, maintaining an aspect ratio of 4:5. This resizing approach ensures consistent scaling across various facial positions for each subject. Some images from the dataset have been shown in Fig. 9.



Fig. 9. Samples from IITM Face Emotion dataset (Best viewed in color).

C. Data Pre-processing

Despite of the approaches in the literatures, we have been intruding denoise process as a data pre-processing step. The method has been used to denoising face image is the method called Non-Local Means proposed in study [25] which based on a simple principle: replacing the color of a pixel with an average of the colors of similar pixels. But the most similar pixels to a

given pixel have no reason to be close at all. It is therefore licit to scan a vast portion of the image in search of all the pixels that really resemble the pixel one wants to denoise. Thus, we have split each image into five parts and we resize them into 224x224 resolution to be compatible with our modified AlexNet input layer. To resize images, we have used Cubic Interpolation algorithm. In short, for denoising and rescaling images, we have been used the implementation of Cubic Interpolation and Non-Local Mean algorithms available in OpenCV library. The face bounding box detection has been achieved by Haar-like detector present in study [26].

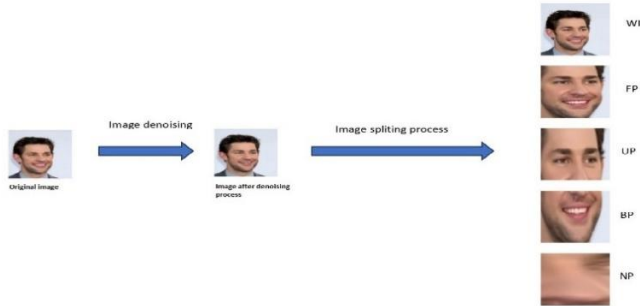


Fig. 10. Example of pre-processing steps for each image in LFWA datasets (Best viewed in color).

Furthermore, we have been processing to cross-validation approaches to predict the skill of our subnets.

In general, cross-validation is a statistical technique employed to assess the performance of machine learning models. When you have both a machine learning model and data at hand, you aim to determine its capability to fit the data. One common approach is to divide the data into training and test sets, training the model on the former and evaluating its performance on the latter. However, a single evaluation may not be sufficient to ascertain whether a favorable outcome is due to genuine model efficacy or mere chance. By conducting multiple evaluations through cross-validation, you can gain greater confidence in the robustness and design of your model.

The method involves a parameter known as 'k,' which denotes the number of subsets the data sample will be divided into. Hence, this technique is commonly referred to as 'k-fold cross-validation.' When a particular value is selected for 'k,' it can replace 'k' in the model reference; for example, setting k=10 would be termed '10-fold cross-validation.'

Unfortunately, k-fold cross-validation may not be suitable for assessing imbalanced classifiers. This is because the data is divided into k-folds based on a uniform probability distribution.

While this approach may be effective for datasets with a balanced class distribution, it can falter when faced with severely skewed distributions. In such cases, one or more folds may contain minimal or no instances of the minority class. As a result, many model evaluations could be misleading, since the model could achieve high accuracy by simply predicting the majority class.

Balanced dataset steps. Machine learning algorithm

performance is often assessed using public datasets like LFWA (see Fig. 10), but this approach can be problematic for imbalanced data. For instance, consider the task of gender classification in face attributes. A typical face dataset might have a distribution of 98% male and 2% female samples. Simply guessing the majority class would result in a predictive accuracy of 98%. However, the application demands high accuracy for detecting the minority class (female) while allowing for some errors in the majority class (male) to achieve this precision. Relying solely on straightforward predictive accuracy is not suitable in such scenarios. This realization underscores the necessity of balancing the dataset to obtain more accurate and meaningful results.

In this study, we adopted the method proposed in study [25] to balance the dataset for each subnet within our proposed pipeline. We opted for this algorithm due to its approach of over-sampling the minority class by generating "synthetic" examples rather than simply duplicating existing ones. This method creates additional training data by applying specific operations to real data, such as selecting k nearest neighbors from the minority class. However, it should be noted that this approach generates synthetic examples in a more generalized manner, operating in "feature space" rather than directly in "data space". Conversely, the majority class is addressed by under-sampling, wherein samples are randomly removed until the minority class comprises a specified percentage of the majority class.

As motioned in the sections above, our model combined five subnets, each one has a specific subset of attributes. Therefore, we have been applied SMOTE algorithm for each part of subnets parts. More details have been shown in figures; Fig. 11, Fig. 12, Fig. 13, Fig. 14, Fig. 15, Fig.16, Fig. 17 and Fig. 18.

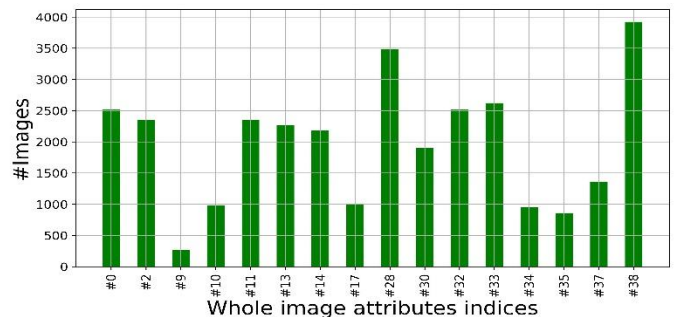


Fig. 11. Data distribution (LFWA dataset) before applied SMOTE algorithm to 16 face attributes according to WI-Net (Best viewed in color) (I).

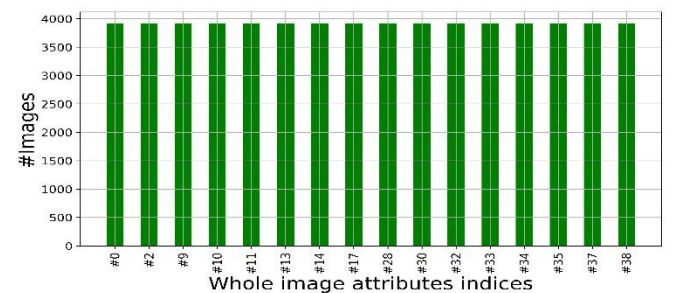


Fig. 12. Data distribution (LFWA dataset) after applied SMOTE algorithm to 16 face attributes according to WI-Net (Best viewed in color) (II).

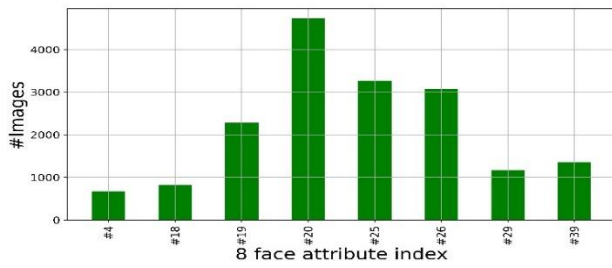


Fig. 13. Data distribution (LFWA dataset) before applied SMOTE algorithm to 8 face attributes according to FP-Net (Best viewed in color) (I).

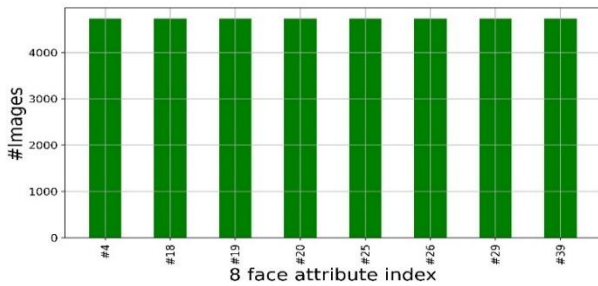


Fig. 14. Data distribution (LFWA dataset) after applied SMOTE algorithm to 8 face attributes according to FP-Net (Best viewed in color) (II).

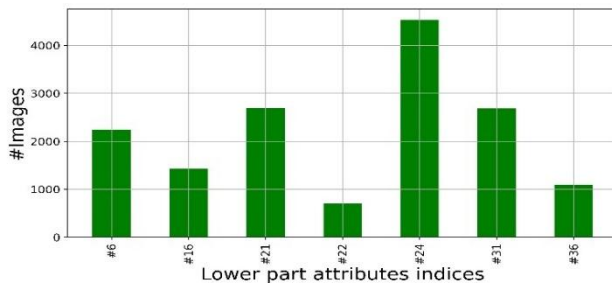


Fig. 15. Data distribution (LFWA dataset) before applied SMOTE algorithm to 7 face attributes according to LP-Net (Best viewed in color) (I).

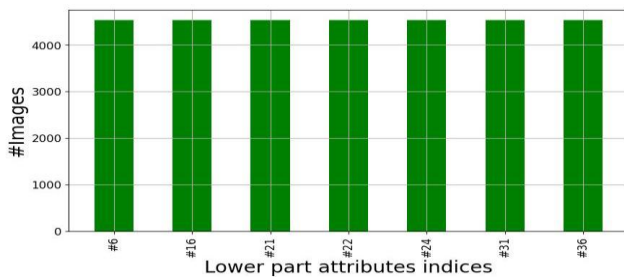


Fig. 16. Data distribution (LFWA dataset) after applied SMOTE algorithm to 7 face attributes according to LP-Net (Best viewed in color) (II).

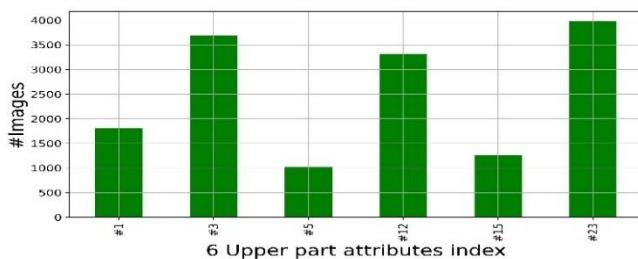


Fig. 17. Data distribution (LFWA dataset) before applied SMOTE algorithm to 6 face attributes according to UP-Net (Best viewed in color) (I).

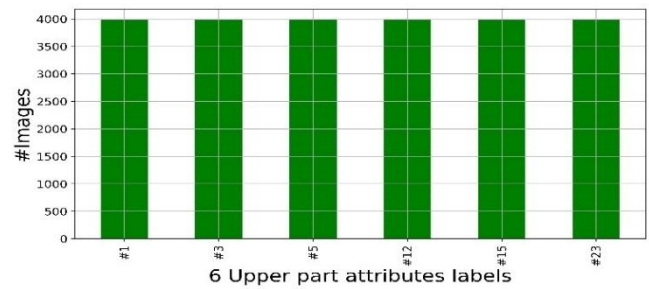


Fig. 18. D Data distribution (LFWA dataset) after applied SMOTE algorithm to 6 face attributes according to UP-Net (Best viewed in color) (II).

For nose part of our model, we have approximation the same number of classes; 4302 for Pointy Nose compared with 4321 for Big Nose. In this case there is no need to applied SMOTE algorithm.

D. Network Parameters

In the previous subsections, we have been described evaluation metrics, data augmentation and data pre-processing steps and this subsection, we will give more details about some parameters of our method. Therefore, we have been used the grid search algorithm [27], implemented in the kears framework to determine the values of some parameters; optimizer, Mini-batch size and Initial learning rate. See Table III for more details about search space of those parameters.

TABLE III. DETAILS OF SEARCH SPACE FOR EACH PARAMETER IN GRID SEARCH ALGORITHM [27]. WE HAVE BEEN INSPIRED FROM WORK IN [28] TO CHOOSE THE INTERVAL OF VALUES SPECIFIED FOR MINI-BATCH SIZE

Parameters	Values
Optimizer ^a	SGD; RMSprop; Adam; AdamW; Adadelta; Adagrad; Adamax; Adafactor; Nadam and Ftrl.
Mini-batch size	16; 28; 32; 64; 128; 256
Initial learning rate	0.1; 0.01; 0.001

^a. All optimizer function names are reported from there implementation in Keras framework.

Thought a set test, we have concluded that the best values of the previous parameters are; SGD (the Stochastic Gradient Descent) as optimizer, mini-batch size equal to 28 and Initial learning rate equal to 0.001. All test has been made on FP-Net subnet for gender estimation attribute (#21) with a number of epochs equal to 100. In the next step, we have been increased the number of epochs from 0 to 600 in the training experiments, in order to set the epoch number which, get the better results in term of accuracy. This experiment shown that the achieves 100% and 0%, respectively, for Accuracy and Loss at the number of epoch equal to 500 (See Fig. 19 for more details). Therefore, we have been based on this conclusion to applied the parameters for all five subnets listed above (WI-Net, FP-net, UP-Net, LP-Net and NP-Net).

E. Experimental Results

This subsection summarizes the results that were obtained from the experiments for both datasets LFWA and IITM face emotion. The most methods in the literature use LFWA to make those evaluation. In addition, we have been choosing to use IITM face emotion, which has Asian people as subject, since the LFWA dataset has Asian people as minority class compared

with others ethnicity (White, Africans). Another reason to use IITM face emotion dataset is all images has been taken in constrained condition for head pose, emotions and light when LFWA is unconstrained dataset. In short, we have been used the LFWA dataset to shown the performance of our approach compared with state-of-the-art methods and we have been used IITM dataset to shown behavior of our method in constrained conditions on Asian people and to show the general ability of proposed model.

Through recent research, we have Gender and Smile are the most interested among all face attributes. To evaluate our proposed method on those two attributes, we have been used the FP-Net subnet to evaluate Gender estimation and LP-Net subnet to evaluate Smile estimation. The experimental set has been made on LFWA [7] and IITM face emotion [24] since those two datasets provide gender and smile information; therefore, LFWA dataset specified the smile by attribute number 32 (#32) and gender by attribute number 21 (#21) when the IITM face emotion dataset given that two information on image name (see [24] for attribute description).

The remaining subsections of this part are organized as follows. The results obtained on 40 face attributes compared with several state-of-the-art methods is described in subsection a). The subsection b) shown the performance of our model in Gender estimation. Finally, we have been shown the behaviors of our model on Smile estimation task in subsection c).

1) *40 face attribute estimations*: We have been training the five subnets on LFWA dataset trough 500 epochs after having applied pre-processing and data augmentation steps. All subnet parameters have been fixed like described in previous Network parameters subsection. However, the results of competing methods are reported from those originals paper whose respect the same protocol provided by dataset owners. Despite, in our method we have been applied SMOTE algorithm on dataset

before training process. The classification results of proposed method and the competing methods on LFWA dataset have been presented in Table IV.

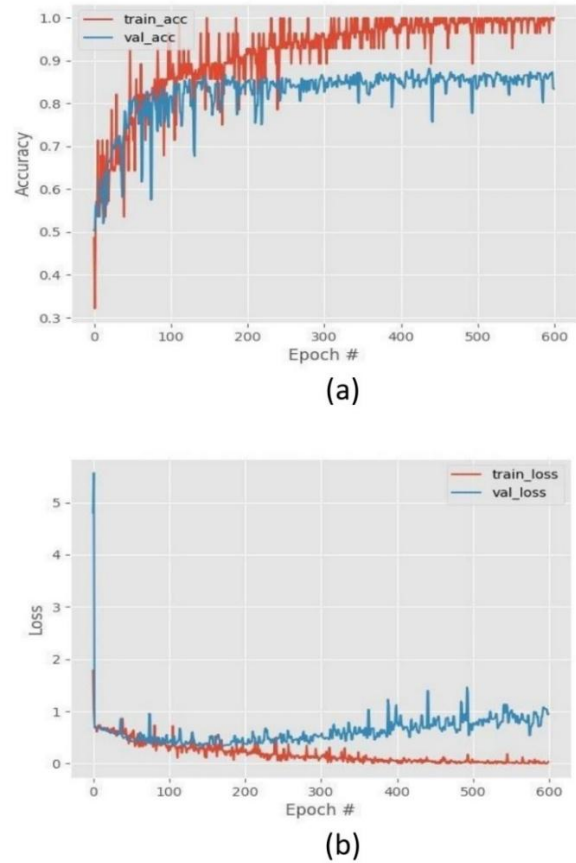


Fig. 19. Training and validation Accuracy/Loss for FP-Net on LFWA dataset (a and b) (Best viewed in color).

TABLE IV. ATTRIBUTE ESTIMATION ACCURACIES (IN %) FOR THE 40 BINARY ATTRIBUTES ON THE LFWA DATABASE BY THE PROPOSED APPROACH AND STATE-OF-THE ART METHODS [5], [6], [10], [14], [16], [15], [29]. THE AVERAGE ACCURACIES OF [5], [6], [10], [14], [16], [15], [29], AND THE PROPOSED APPROACH ARE 86.0%, 83.8%, 81.0%, 86.3%, 86.1%, 73.9%, 84.7% AND 86.8% RESPECTIVELY. SEE TABLE I FOR MORE DESCRIPTION ABOUT ATTRIBUTES LABELS

Attribute index	State-of-the Art Methods							Proposed Approach
	FaceTracker[15]	PANDA[10]	LNets+Anet[6]	CTS-CNN[29]	MCNN-AUX[5]	DMTL[16]	MM-CNN[14]	
<u>1</u>	70	84	84	77	77	80	78	85.00
<u>2</u>	67	79	82	83	82	86	81	83.00
<u>3</u>	67	79	82	83	85	82	81	85.00
<u>4</u>	71	81	83	79	80	84	83	85.50
<u>5</u>	65	80	83	83	83	92	93	97.00
<u>6</u>	77	84	88	91	92	93	92	94.00
<u>7</u>	72	84	88	91	90	77	79	82.00
<u>8</u>	76	87	90	90	93	83	84	94.00
<u>9</u>	88	94	97	97	97	92	92	97.00
<u>10</u>	62	74	77	76	81	97	97	97.00
11	78	81	84	87	89	89	85	83.00
<u>12</u>	68	73	75	78	79	81	82	82.00
<u>13</u>	73	79	81	83	85	80	85	86.00

14	73	74	74	88	85	75	76	79.00
15	67	69	73	75	77	78	82	83.00
16	70	75	78	80	82	92	92	95.00
17	90	89	95	91	91	86	84	90.00
18	69	75	78	83	83	88	89	97.00
19	88	93	95	95	96	95	95	86.00
20	77	86	88	88	88	89	87	81.00
21	84	92	94	94	94	93	94	95.00
22	77	78	82	81	84	86	82	88.00
23	83	87	92	94	93	95	94	97.00
24	73	73	81	81	83	82	82	86.00
25	69	75	79	80	82	81	81	86.00
26	66	72	74	75	77	75	79	80.00
27	70	84	84	73	93	91	91	68.00
28	74	76	80	83	84	84	85	87.00
29	63	84	85	86	86	85	87	90.00
30	70	73	78	82	88	86	87	89.00
31	71	76	77	82	83	80	84	88.00
32	78	89	91	90	92	92	91	92.00
33	67	73	76	77	79	79	79	80.00
34	62	75	76	77	82	80	82	82.00
35	88	92	94	94	95	94	94	90.00
36	75	82	88	90	90	92	91	90.00
37	87	93	95	95	95	93	95	86.00
38	81	86	88	90	90	91	90	87.00
39	71	79	79	81	81	81	83	85.00
40	80	82	86	86	86	87	85	78.00
Average	73.9	81.0	83.8	84.7	86.3	86.1	86.3	86.83

2) *Gender estimation*: In this subsection, we have been shown the results of our proposed approach for gender estimation task. As mentioned in section above, we have been pre-processed each image by denoising and splitting to remove the noise and adjusted them to input blocks.

We have been compared our approach with FaceTracer [15], PANDA[10], LNet+ANets [6] and all models (R-CNN_Gender, Multitask_Face, HyperFace and HF-ResNet) proposed in [8]. The gender estimation performance of different methods is reported in Table V. On the LFWA dataset, our method outperforms all competing methods listed above. Unlike all these methods in our approach, we have been processed to balanced dataset before the training step.

The imbalanced distribution in datasets make behavior of each model change from dataset to another. In addition, we have been evaluated the generalization ability of our FP-Net approach with cross-database testing on the IITM face emotion and LFWA. See Table VI for more details.

3) *Smile estimation*: Smile detectors find applications across various sectors, including the media industry. Here, they play a crucial role in enabling companies to gauge public sentiment towards their products and services. For this reason, in this part of our paper, we have been interested to Smile

estimation task. However, we have been presented the Smile estimation performance on LFWA and IITM face emotion datasets since these datasets come with Smile information. We have been compared our LP-Net with MCFA [30], PANDA [10], FMTNet [21] and LNet+ANets [6]. The Smile estimation performance of different method is reported in Table VII. Furthermore, we have been evaluated the generalization ability of our LP-Net approach with cross-database testing on the IITM face emotion and LFWA. See Table VIII for more details.

TABLE V. PERFORMANCE COMPARISON (IN %) OF GENDER ON LFWA DATASET

Method	LFWA dataset
FaceTracer[15]	84.00
PANDA[10]	92.00
LNet+ANets[6]	94.00
R-CNN_Gender[8]	91.00
Multitask_Face[8]	93.00
HyperFace[8]	94.00
HF-ResNet[8]	94.00
Proposed FP-Net	95.00

TABLE VI. CROSS-DATABASE TESTING ACCURACIES (IN %) OF FP-NET USING LFWA AND IIITM FACE EMOTION DATABASES FOR GENDER CLASSIFICATION

Database		Metrics			
Training	Testing	Accuracy	Precision	Recall	F1-score
IIITM	IIITM	99%	100%	99%	99%
IIITM	LFWA	50%	50%	100%	67%
LFWA	IIITM	79%	99%	80%	88%

TABLE VII. PERFORMANCE COMPARISON (IN %) OF SMILE ON LFWA DATASET

Method	LFWA dataset
MCFA[30]	88.00
PANDA[10]	89.00
FMTNet[21]	89.49
LNets+ANets[6]	91.00
Proposed LP-Net	92.00

TABLE VIII. CROSS-DATABASE TESTING ACCURACIES (IN %) OF LP-NET USING LFWA AND IIITM FACE EMOTION DATABASES FOR SMILE CLASSIFICATION

Database		Metrics			
Training	Testing	Accuracy	Precision	Recall	F1-score
IIITM	IIITM	91%	94%	96%	95%
IIITM	LFWA	54%	51%	92%	65%
LFWA	IIITM	54%	99%	48%	65%

V. DISCUSSION

In this section, summarized the discussion of the results achieved about 40 faces attributes, gender recognition, smile estimation and generalization ability of the proposed system. In the first subsection, we have been presented the effectiveness in the most face attributes (30/40) of the proposed approaches and some lower results (10/40). Additional results analysis about gender recognition has been shown in the second subsection. the third subsection contains behavior of our proposed approach in smile estimation task, and finally we a subsection on generalization ability of our system evaluated on cross-database testing.

A. 40 Face Attributes Estimation

The results on LFWA dataset by proposed approach and the state-of-the-art are reported in Table IV. The proposed approach outperforms [5], [6], [10], [14], [16], [15] and [29] for the most of the 40 attributes. Specifically, the proposed approach outperforms competing methods by 30 face attributes. The remaining 10 attributes belongs to WI-Net, FP-Net and LP-Net. Those attributes can be subdivided into three groups; {#14, #35, #36 and #38} from Whole Image group, {#19, #20, #27 and #40} from Face Part group and {#17, #37} from Lower Part group. For WI-Net, we have four attributes has lower results from 17 attributes according to this subnet. All those attributes have a number of attributes in correlation, less than 5 attributes (see Table II) and belongs to image segments which need

Meged-CNN structure like [14], very deep structure like in [29] or more the 4 attributes in correlation like the work in [16]. Further, the subnet FP-Net show 4 attributes with poor results from 8 attributes according to this subnet, while the competing methods in [16] and [6] show better results than our FP-Net. Despite, they use similar network structure to ours (AlexNet) but they use more attributes in correlation task (attributes inter-correlation). In addition, the attributes #17 from LP-Net subnet show similar results to [15] which use pixel of image combined with AdaBoost algorithm to achieve classification process but this approach show a limitation for profile face image even though, our subnet show lower results compared to [6], [5] and [29]. Since, the work in [5] use network structure similar to ours, and they add auxiliary block (coined AUX) which use fully connection approach between all attributes to handle attribute #17 estimation, when our LP-Net use just 3 attributes (see Table II for more details). In the same manner, the work in [6] use more than 3 attributes to make #17 estimation. However, the good results provided in [27] has been achieved by a deeper structure (similar to VGG) than ours. On the other hand, our proposed subnet shows better results about #17 than [10], [14] and [16] who's adopted AlexNet as backbone like ours, but LP-Net outperforms the estimation in task specification step (feature selection and AdaBoost classifier). Finally, for attribute #37 our LP-Net shows lower results compared to all competing methods listed in Table IV. Thus, the works [8], [12] and [13] use similar backbone model like ours when [3], [4] use a deeper structure than AlexNet when the number of attributes used in classification step for all those methods is bigger than 5 attributes (see Table II) used in our proposed method.

Through all previous analysis, we have been concluded for WI-Net and LP-Net necessity to increase a number of attributes used in classification process to a number bigger than 5 and we have been concluded about FP-Net necessity to change AlexNet model by another structure like VGG or ResNet.

B. Gender Recognition

We present the gender recognition performance on LFWA and IIITM Face Emotion datasets since these datasets come with gender information. Our approach shows better results compared with PANDA [10] (see Table V). Based on results reported from there paper, this approach shows good performance in gender estimation on LFWA (92%) even when images are tightly cropped and variation in pose is reduced, but our FP-Net reach 95% with a wide variation in facial pose (our approach gain more from parts-based assumptions in term of head pose challenge). In the same manner, our FP-Net model outperformed FacTracer [15] methods by 11% in term of improvement. The FaceTracer [15] approach spite the input face to 10 parts while our FP-Net split the face into 4 parts only which get further advantage in terms of pre-processing time and has limitation with non-frontal face. In addition, our FP-Net outperforms LNets+ANets[6] by 1% with 5 attributes groups for our and six groups for LNets+ANets (more advantage in terms of group number). However, our FP-Net use one Modified AlexNet structure when LNets+ANets use a cascade of two AlexNet model which make advantage, on parameter number, train/test time complexity and memory consumption. Although, HyperFace models proposed in [8] has the same backbone network like our FP-Net (Modified AlexNet) even the FP-Net

shows better results more than R-CNN_Gender and Multitask_Face. Respectively, by 4% and 2% in the term of improvement. R-CNN_Gender predicts gender task only when our FP-Net can estimate 8 more attributes in addition of gender task (multi labels estimation against one task estimation). Since, Multitask_Face predicts gender in multitask approach which make training process hard than ours (multi labels against multitask process in the training set). Furthermore, the HF-ResNet approach use ResNet-101 model as a backbone network combined with AlexNet model which make it a very deep structure compared with our FP-Net model and slower than our train/test set. However, our FP-Net approach improved HF-ResNet by 1% in the term of accuracy. On the whole, the proposed FP-Net model for joint estimation of gender attribute demonstrates their effectiveness compared with existing approaches at many specific levels like; parameters number, memory consumption, number of attributes subgroups, accuracy and time complexity.

C. Smile Detection

Our proposed LP-Net model outperformed the state-of-the-art methods (see Table VII) in Smile detection on LFWA dataset. All accuracies value presented in Table VII has been reported from their original papers. To illustrate, our LP-Net outperformed the work in [30] by 4% of accuracy and the cited work used a cascade of three VGG-16 models (SNet, MNet and LNet) which made the detection process more complex and harder to train. Even though, our LP-Net use just one Modified AlexNet with seven attributes in the output combined with Adaboost classifier which made it less complex in train and test sets. Further, FMTNet presented in [21] take 40 attributes and split them to many subgroups, while each attribute has weight depending on the number of groups and the number of attributes in each group. Which means this approach investigate 40 attributes to estimate smile attribute when our LP-Net use just 3 attributes selected by cited algorithm from 7 attributes in relation with lower part of face. Therefore, LP-Net proposed in this work use less attributes to show a attributes correlation and gained 3% in the term of accuracy compared with FMTNet. On the other hand, the work in [21] investigate three model use VGG-16 as backbone (FNet, MNet and TNet) when our LP-Net use just one AlexNet model, to handle the smile task. The proposed LP-Net shows improvement reach 1% and 3%, compared the results provide, respectively, by LNet+ANet and PANDA methods. In short, we find that LP-Net proposed in this work performs better for Smile detection task compared to competing methods. In the other hand, our approach shows a discriminating capability for multis and individual tasks.

D. Generalization Ability

We believe that the real scenario is different from the laboratory scenario which mean the generalization ability provided in [16] for the first time (for the best of our knowing) can give more information about our proposed approach. Hence, we evaluate the generalization ability of the proposed approach with cross-database testing on LFWA and IIITM Face emotion databases.

Specifically, cross-database testing of gender and smile estimation between LFWA and IIITM Face emotion databases is performed by training our approach on LFWA and testing it

on IIITM Face emotion, and vice versa. The estimation results with cross-database are shown in Table VI and Table VIII. The results provided by cross-database testing is lower than intra-database testing. Image conditions (constrained in IIITM Face emotion and unconstrained in LFWA) and the number of images (1,928 images in IIITM face emotion and 13,143 images in LFWA) are responsible for the drop in performance. This experiment suggests that varying image sources can introduce additional hurdles for accurately estimating facial attributes. Nevertheless, we maintain confidence that our proposed approach yields commendable results even within this demanding context.

VI. CONCLUSIONS

The paper proposes a method to decode face attributes using a multi-task, part-based approach and attribute relationships. In contrast of exciting works, it introduces two preprocessing steps: image denoising with the Non-Local Means algorithm and dataset balancing using the SOME algorithm. The feature selection has been done by splitting images into five parts (WI, FP, UP, LP, NP) and each one has been processed by a corresponding subnet (modified AlexNet as backbone). The correlation-heterogeneity relationship between attributes has been achieved by a novel feature selection Algorithm (proposed in this work) combined with AdaBoost algorithms.

To evaluated the proposed approach, we have been used LFWA and IIITM Face Emotion datasets. The first one has images in unconstrained conditions and large scale of illumination, head pose, ... when the second one has images with constrained conditions of head pose, illumination and expression. This strategy helps to studies the performance of proposed approach in different conditions and ethnicity.

Trought a set of experiments has been made in this work, our approach performs well the state-of-the-arts methods specifically, on gender and smile attributes. Nevertheless, the results presented in this work shown that our subnets; WI-Net, UP-Net, LP-Net and NP-Net outperforms the competing methods on specific attributes groups, according to those parts of face, when the subnet FP-Net shown some lower results for attributes {#19, #20, #27 and #40}. One possible solution to this issue could be replaced AlexNet with deeper structure similar to VGG or ResNet. While, the lower results present by this work for attributes number {#14, #35, #36 and #38} and {#17, #37} returned respectively, by WI-Net and LP-Net, could be handle by increasing a number of attributes used to shown the relationship in classification process.

Finally, we have been studied the generalization ability of the proposed approach under cross-database testing scenarios on LFWA an IIITM Face Emotion datasets. Through a results analysis, the cross-database testing highlights the importance of training database in real-world face attributes estimation systems.

For future work, we will try to use a deeper structure for attributes with lower results in FP-Net subnet and we will investigate more time in feature selection to get better results for attributes number {#14, #35, #36 and #38} and {#17, #37}. On the other hand, we will adapt the proposed approach to estimate

age task and face emotion. The age task will be studied in regression manner.

REFERENCES

- [1] O. Maarouf, A. Maarouf, R. El Ayachi, et M. Biniz, « Automatic translation from English to Amazigh using transformer learning », *Indones. J. Electr. Eng. Comput. Sci.*, vol. 34, no 3, p. 1924, 2024, doi: 10.11591/ijeecs.v34.i3.pp1924-1934.
- [2] M. Biniz et R. El Ayachi, « Recognition of Tifinagh Characters Using Optimized Convolutional Neural Network », *Sens. Imaging*, vol. 22, no 1, p. 28, 2021, doi: 10.1007/s11220-021-00347-1.
- [3] U. D. Dixit, M.S. Shirdhonkar, « Face-based Document Image Retrieval System », *Procedia Computer Science*, Volume 132, 2018, Pages 659-668, ISSN 1877-0509, doi :10.1016/j.procs.2018.05.065.
- [4] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, « A review of video surveillance systems », *Journal of Visual Communication and Image Representation*, Volume 77, 2021, doi:10.1016/j.jvcir.2021.103116..
- [5] E. Hand et R. Chellappa, « Attributes for Improved Attributes: A Multi-Task Network Utilizing Implicit and Explicit Relationships for Facial Attribute Classification », *Proc. AAAI Conf. Artif. Intell.*, vol. 31, no 1, 2017, doi: 10.1609/aaai.v31i1.11229.
- [6] Z. Liu, P. Luo, X. Wang, et X. Tang, « Deep Learning Face Attributes in the Wild », in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile: IEEE, 2015, p. 3730-3738. doi: 10.1109/ICCV.2015.425.
- [7] B. H. Gary, R. Manu, B. Tamara, et L.-M. Erik, « Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments », p. 07-49, 2007.
- [8] R. Ranjan, V. M. Patel, et R. Chellappa, « HyperFace: A Deep Multi-task Learning Framework for Face Detection, Landmark Localization, Pose Estimation, and Gender Recognition », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no 1, p. 121-135, janv. 2019, doi: 10.1109/TPAMI.2017.2781233.
- [9] D. Fan, H. Kim, J. Kim, Y. Liu, et Q. Huang, « Multi-Task Learning Using Task Dependencies for Face Attributes Prediction », *Appl. Sci.*, vol. 9, no 12, p. 2535, 2019, doi: 10.3390/app9122535.
- [10] N. Zhang, M. Paluri, M. Ranzato, T. Darrell, et L. Bourdev, « PANDA: Pose Aligned Networks for Deep Attribute Modeling », in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA: IEEE, 2014, p. 1637-1644. doi: 10.1109/CVPR.2014.212.
- [11] J. Cao, Y. Li, et Z. Zhang, « Partially Shared Multi-task Convolutional Neural Network with Local Constraint for Face Attribute Learning », in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT: IEEE, 2018, p. 4290-4299. doi: 10.1109/CVPR.2018.00451.
- [12] S. Yang, P. Luo, C. C. Loy, et X. Tang, « Faceness-Net: Face Detection through Deep Facial Part Responses », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no 8, p. 1845-1859, 2018, doi: 10.1109/TPAMI.2017.2738644.
- [13] U. Mahbub, S. Sarkar, et R. Chellappa, « Segment-Based Methods for Facial Attribute Detection from Partial Faces », *IEEE Trans. Affect. Comput.*, vol. 11, no 4, p. 601-613, 2020, doi: 10.1109/TAFFC.2018.2820048.
- [14] H. Kawai, K. Ito, et T. Aoki, « Face Attribute Estimation Using Multi-Task Convolutional Neural Network », *J. Imaging*, vol. 8, no 4, p. 105, avr. 2022, doi: 10.3390/jimaging8040105.
- [15] N. Kumar, P. Belhumeur, et S. Nayar, « FaceTracer: A Search Engine for Large Collections of Images with Faces », in *Computer Vision – ECCV 2008*, vol. 5305, D. Forsyth, P. Torr, et A. Zisserman, Éd., in *Lecture Notes in Computer Science*, vol. 5305, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, p. 340-353. doi: 10.1007/978-3-540-88693-8_25.
- [16] H. Han, A. K. Jain, F. Wang, S. Shan, et X. Chen, « Heterogeneous Face Attribute Estimation: A Deep Multi-Task Learning Approach », *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no 11, p. 2597-2609, 2018, doi: 10.1109/TPAMI.2017.2738004.
- [17] J. Xiang et G. Zhu, « Joint Face Detection and Facial Expression Recognition with MTCNN », in *2017 4th International Conference on Information Science and Control Engineering (ICISCE)*, Changsha: IEEE, 2017, p. 424-427. doi: 10.1109/ICISCE.2017.95.
- [18] S. Ioffe et C. Szegedy, « Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift », 2015, doi: 10.48550/ARXIV.1502.03167.
- [19] C. Shorten et T. M. Khoshgoftaar, « A survey on Image Data Augmentation for Deep Learning », *J. Big Data*, vol. 6, no 1, p. 60, 2019, doi: 10.1186/s40537-019-0197-0.
- [20] A. Shannaq et L. Elrefaei, « AGE ESTIMATION USING SPECIFIC DOMAIN TRANSFER LEARNING », *Jordanian J. Comput. Inf. Technol.*, no 0, p. 1, 2020, doi: 10.5455/jcit.71-1571410322.
- [21] N. Zhuang, Y. Yan, S. Chen, H. Wang, et C. Shen, « Multi-label learning based deep transfer neural network for facial attribute classification », *Pattern Recognit.*, vol. 80, p. 225-240, 2018, doi: 10.1016/j.patcog.2018.03.018.
- [22] H. Song et H. Chen, « A Fast and Effective Large-Scale Two-Sample Test Based on Kernels », 2021, doi: 10.48550/ARXIV.2110.03118.
- [23] F. Charte, A. J. Rivera, M. J. Del Jesus, et F. Herrera, « MLSMOTE: Approaching imbalanced multilabel learning through synthetic instance generation », *Knowl.-Based Syst.*, vol. 89, p. 385-397, 2015, doi: 10.1016/j.knosys.2015.07.019.
- [24] Rishi Raj Sharma , K V Arya, April 3, 2023, "IIITM Face Emotion (An Indian Face Image Data)", *IEEE Dataport*, doi: 10.21227/rens-ck04.
- [25] A. Buades, B. Coll, et J.-M. Morel, « Non-Local Means Denoising », *Image Process. Line*, vol. 1, p. 208-212, 2011, doi: 10.5201/ipol.2011.bcm_nlm.
- [26] P. Viola et M. Jones, « Rapid object detection using a boosted cascade of simple features », in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. CVPR 2001, Kauai, HI, USA: IEEE Comput. Soc, 2001, p. I-511-I-518. doi: 10.1109/CVPR.2001.990517.
- [27] P. Liashchynskiy et P. Liashchynskiy, « Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS », 2019, doi: 10.48550/ARXIV.1912.06059.
- [28] I. Kandel et M. Castelli, « The effect of batch size on the generalizability of the convolutional neural networks on a histopathology dataset », *ICT Express*, vol. 6, no 4, p. 312-315, 2020, doi: 10.1016/j.icte.2020.04.010.
- [29] Y. Zhong, J. Sullivan, et H. Li, « Face Attribute Prediction Using Off-the-Shelf CNN Features », 2016, doi: 10.48550/ARXIV.1602.03935.
- [30] N. Zhuang, Y. Yan, S. Chen, et H. Wang, « Multi-task Learning of Cascaded CNN for Facial Attribute Classification », 2018, doi: 10.48550/ARXIV.1805.01290.

Evaluation of Eye Movement Features and Visual Fatigue in Virtual Reality Games

Yuwei Ji

Department of Basic Education, Zhengzhou Urban Construction Vocational College, Zhengzhou-451263, China

Abstract—VR games make people happy physically and mentally, but also lead to eye health problems. At present, the existing VR systems lack fatigue detection technology, which makes it difficult to help users use their eyes reasonably. In order to improve the user experience of VR gamers, this paper proposes a visual fatigue detection algorithm based on eye movement features, which uses the relationship between the lateral and longitudinal displacements of the human head and the displacement of the center point of the human eye to locate the position of the human eye. Moreover, in this paper, the human eye position tracking model is input into the three-frame difference algorithm to detect eye movement features. In addition, for tiny motion interference such as eyebrows, the image opening operation of eroding first and then expanding is used to remove it. Through experiments, it is found that the eye movement feature detection method adopted in this paper can greatly improve the detection speed with less accuracy loss, meet the sensitivity requirements of eye movement feature capture, improve the real-time performance of the system, and effectively improve the real-time analysis of player status. Therefore, integrating this algorithm into the virtual game system can help players adjust their own state, which has a positive effect on improving the game experience and reducing eye damage.

Keywords—Virtual reality; games; eye movement features; visual fatigue

I. INTRODUCTION

In three-dimensional game design, the application value of VR technology cannot be ignored. First of all, VR technology can provide an immersive gaming experience, which is one of its most significant advantages. Through technical means such as headsets and surround sound effects, VR technology can completely immerse players in a virtual environment, making them feel as if they are in the three-dimensional world of the game. This immersion greatly enhances the player's gaming experience, allowing them to get a more realistic feel for everything that's happening in the game. From a technical point of view, VR technology adjusts the game screen in real time by tracking the movement of the player's head, making the player's field of vision and observation angle in the game more natural. At the same time, surround sound technology can make players feel sounds from all directions, further enhancing the realism of the game.

Secondly, immersive experiences bring more possibilities to game design. In traditional game design, there is a certain sense of distance between players and the game world, and the application of VR technology can break this boundary and enable players to integrate more deeply into the game world. This provides a broader creative space for game designers, and

can design richer and more complex game scenes and plots, thus enhancing the attractiveness and interest of the game [1].

However, long-term use of VR (more than 30 minutes) may have a certain negative impact on the visual health of the whole body and eyes. The specific manifestations include dizziness, nausea and other symptoms. At the same time, users are accompanied by dry eyes, and even symptoms such as diplopia, tearing, eye pain, eye soreness, and inability to concentrate, as well as related symptoms such as visual asthenopia and video terminal syndrome. The illusion caused by the virtual environment can produce uncomfortable symptoms, such as eye fatigue, dizziness, and other visual fatigue symptoms [2].

There are depth cues in virtual environment scenes, which stimulate eye movements. There is a strong correlation between eye movements and asthenoptic fatigue. Many studies have shown that eye movement behavior can reflect people's thinking movement, and rich information can be obtained from eye tracking movement. Its core purpose is to obtain the gaze point trajectory of human eyes during observation. Combined with knowledge in various fields, we can conduct in-depth analysis of users' visual behavior. Eye trackers are instruments that carry eye tracking technology to track and analyze eye movements, and eye tracking is the core function of eye trackers. From a mental and physical perspective, eye movements are a fundamental reflection of the human state. Meanwhile, eye movements are arguably the most frequent of all human movements, and eye movements are essential to the work of the human visual system. In addition, multiple observations of the eyes are not smooth movements, but multiple eye movement patterns are performed concurrently [3].

Visual fatigue and visual discomfort can be used alternately, but there are still differences between them. Visual fatigue refers to the decline of human visual system performance, which can be measured objectively, and visual discomfort is the subjective response of the visual system. Some researches on eye movement behavior in virtual reality focus on comparing the States before and after use, and some researches also compare the eye movement differences under watching different video content. However, up to now, there has been no research on the differences of visual fatigue and eye movement in different interactive environments of VR. Eye movement behavior is closely related to visual fatigue and the physiological and pathological state of the eye. The changes of eye movement speed, fixation time and blink frequency determine fatigue and mental load.

In response to the design requirements of high speed, small size and non-contact of detection equipment, the face image is

preprocessed by image processing technology such as noise reduction; Then the YCbCr color space domain conversion algorithm is used to segment and locate the face and eye region; The blink frequency is counted by frame difference multiple moving object detection algorithm in the face region after positioning; By comparing the real-time detected blink frequency data with the given threshold, it can provide real-time fatigue data reference for VR game users, which is convenient for timely and effective control of game duration and timely and effective protection.

In order to improve the user experience of VR gamers, this paper proposes a visual fatigue detection algorithm based on eye movement features, which uses the relationship between the lateral and longitudinal displacements of the human head and the displacement of the center point of the human eye to locate the position of the human eye. Moreover, in this paper, the human eye position tracking model is input into the three-frame difference algorithm to detect eye movement features. In addition, for tiny motion interference such as eyebrows, the image opening operation of eroding first and then expanding is used to remove it.

II. RELATED WORK

1) *Research on fatigue*: A lot of research has been done on fatigue at home and abroad. In early foreign studies, fatigue was defined as the loss of energy resources after overwork. The resources here are reflected in two aspects. On the one hand, it represents the loss of internal motivation and action from the psychological aspect, and on the other hand, it represents the decrease of external work performance. The study in [4] hold that the decrease of efficiency caused by excessive physical or mental activities is a manifestation of fatigue. Researchers divided fatigue into psychological fatigue and physical fatigue, central fatigue and peripheral fatigue, cognitive fatigue and exercise fatigue, subjective fatigue and objective fatigue, overall fatigue and local fatigue from different dimensions, among which psychological fatigue and physical fatigue are the most widely studied. At present, the most widely studied is to divide fatigue into mental fatigue and physical fatigue. At the same time, it involves psychological state changes in multiple dimensions such as behavioral response, attention, emotion and motivation [5].

2) *Visual fatigue*: Visual fatigue is one of the types of fatigue. Studies have shown that the characteristics of visual display terminal (VDT) work are directly related to eye discomfort and psychological symptoms. Visual fatigue is defined as subjective symptom syndrome produced when working with eyes, while VDT visual fatigue is more due to eye discomfort and other comprehensive symptoms caused by eyes staring at video terminals for a long time, such as dry eyes, astringent eyes, tingling, eye fatigue, soreness, photophobia and tears, frequent eye movement behaviors, diplopia, blurred vision, heavy eyelids, etc. At the same time, it is also accompanied by headache, dizziness, loss of appetite, memory loss, neck, shoulder, waist, back, joint dysfunction, etc. [6]. Visual fatigue caused by electronic screen operation has

become one of the key research fields of human factors engineering since 1970s. The most common fatigue symptom in video display task is eye fatigue, which includes both physiological fatigue and psychological fatigue. The former is manifested as the general symptoms of eye fatigue and fatigue caused by excessive eye use with the extension of working hours, including symptoms that occur after excessive eye function is stressed. It is manifested by low function of central nervous system, a large decrease in flicker fusion frequency, long-term tension of ciliary muscle, significantly low function of eyeball accommodation system, eye fatigue, eye tingling, and temporary decrease of eyesight. The latter, like mental fatigue, is manifested as cognitive fatigue, such as difficulty in maintaining the initial state and continuing to complete the current task, decreased task performance, decreased attention, etc. [7].

3) *Eye movement behavior*: There are three main types of eye movement behaviors: Saccade, Fixation and SmoothPursuit. The saccades are rapid eye movements that align the fovea with the target. During the experiment, it is important to ensure that the subject does not have saccades while chasing the target smoothly. This eye movement is called catch-up saccades and is more common when catching up at high speeds. Fixation is to keep the foveal visual field on the target for a certain period of time to obtain sufficient visual image details. Smooth pursuit is a kind of fairly slow eye movement that minimizes the movement of retinal targets, and it keeps the eyes fixed on moving objects. The saccade eye movement differs from the smooth pursuit eye movement in that the initial acceleration and deceleration and peak velocity of the former are both higher [8]. A large number of related studies in physiology and psychology have confirmed that some behaviors of human eyes are related to the degree of visual fatigue. Therefore, when collecting eye movement data, it is collected by instruments, and the visual fatigue state of subjects is detected by analyzing the data of eye movement indexes. Through the collation of a large number of references, this paper mainly selects six eye movement parameters, namely, average eye movement duration, average fixation duration, number of eye movement behaviors, number of fixation points, average saccade amplitude and average pupil area, for analysis and research [9].

The average duration of eye movement behavior (unit: ms) refers to the duration of the average single eye movement behavior in the sample. The length of eye movement behavior can also reflect the current level of mental activity and drowsiness, which is closely related to fatigue. By watching different types of videos, study in [10] found that although there was no significant difference in the average duration of eye movement behavior on the whole, it showed an increasing trend with time. The study in [11] studied the continuous viewing of movies, and compared which display can induce visual fatigue more in two kinds of visual display terminals (linear polarization and circular polarization), and found that the average duration of eye movement behavior in the two display terminals increased significantly. The study in [12] explored the effects of different

polarized light displays on human visual comfort, and found that the average duration of eye movement behavior increased with the extension of viewing time of subjects who watched the video content displayed by linearly polarized light and the video content displayed by circularly polarized light liquid crystal.

The average fixation duration (unit: ms) refers to the average of the time allocated by subjects at each fixation point in an experimental process, usually in milliseconds. The length of fixation time can reflect the difficulty of information capture and processing, and can indirectly reflect whether the subjects are tired [13]. Generally speaking, the longer the average duration of the fixation point, the deeper the processing degree of the fixation point, and it also reflects the more concentrated the attention of the current subjects. Therefore, the quality of the subjects' fixation ability also reflects the depth of information processing degree and the concentration and distraction of attention to a certain extent. The study in [14] conducted a comparative study on mental fatigue between the elderly and young people, and found that the average fixation duration of both the elderly and young people increase with the development of fatigue, and the average fixation duration of the elderly is significantly higher than that of the young people. The study in [15] divided the subjects into fatigue group and non-fatigue group, and found that the average fixation duration in the fatigue group is significantly lower than that in the non-fatigue group. Relevant studies have proved that the average fixation duration can indirectly reflect the degree of fatigue of the subjects.

Number of eye movement behaviors (unit: units/min) refers to the number of times the upper and lower eyelids are closed per unit time. Number of eye movement behaviors involves the interaction of efferent nerves between the brain mechanisms responsible for controlling eyelid muscles and other muscle groups, and is closely related to mental activity and visual fatigue. A large number of studies have proved that eye movement behavior is related to visual fatigue, which can be used as an index to evaluate visual fatigue. Generally speaking, the number of eye movement behaviors will increase with visual fatigue. The study in [16] found that the number of eye movement behaviors increases significantly with time by watching different types of videos. The study in [17] studied the continuous viewing of movies, and compared which display can induce visual fatigue more in two kinds of visual display terminals (linear polarization and circular polarization). The results found that the number of eye movement behaviors in the two display terminals increases significantly with time. The study in [18] evaluated the visual fatigue caused by long-term viewing of visual display terminal (VDT) and reading hard-copy, and found that the number of eye movement behaviors under VDT increases significantly from the 2nd hour. The study in [19] studied fatigue driving and found that the number of eye movement behaviors increases with the increase of driving fatigue.

To sum up, visual fatigue includes two meanings: on the one hand, it is eye fatigue caused by some reason, and on the other hand, it refers to psychological fatigue caused by boredom of something and cognitive load.

III. VISUAL FATIGUE DETECTION BASED ON EYE MOVEMENT FEATURES

In terms of limitations, although a lot of research has been done on VR game fatigue, there are still some challenges and limitations. Firstly, the hardware limitation of VR technology is an important factor leading to fatigue. For example, the weight and volume of VR head display equipment are relatively large, and wearing it for a long time can easily lead to discomfort and fatigue. The resolution and scene rendering ability of the device also need to be improved to provide a clearer and more realistic visual experience. Secondly, the rendering amount of VR game images is much higher than that of general games, which requires higher computing power to support, which is a technical bottleneck in the actual development. In addition, the current research on how to effectively alleviate the fatigue of VR games lacks systematic solutions and guidelines.

To sum up, the research on VR game fatigue has made some progress, but it still needs more in-depth research and exploration in hardware technology, image rendering and mitigation strategies. Therefore, this paper attempts to summarize the anti-fatigue system suitable for VR games based on eye movement feature recognition, so as to reduce the impact of game fatigue on players' physical and mental health.

By analyzing the relationship between binocular center point displacement and face displacement, a dynamic video eye position tracking model is established, and then the eye movement behavior rate is detected by frame difference algorithm in the human eye area, and finally the game fatigue judgment is completed.

A. Eye Movement Behavior Detection Algorithm

In this project, the eye movement behavior detection algorithm based on eye features, the eye movement behavior detection algorithm based on background differential moving target detection algorithm and the eye movement behavior detection algorithm based on inter-frame differential moving target detection algorithm are analyzed and studied. The eye movement behavior detection algorithm is comprehensively evaluated from three aspects: effective detection rate, false detection rate and detection speed, and finally the eye movement behavior detection algorithm suitable for this project is selected.

1) *Haar-Adaboost human eye coarse positioning*: Haar-Adaboost algorithm combines Haar-like features with Adaboost cascade classifier. The core idea of the algorithm is to take the Haar-like sub-window as the input of the weak classifier, and use the window template to traverse each region of the image to calculate the features of the window. It then uses the trained Adaboost cascade classifier to screen the feature. If the feature passes each strong classifier screening in the cascade classifiers, the region is determined to be the human eye. The process of Haar-like sub-window traversing the image is shown in Fig. 1.

Fig. 1 shows a Haar-like linear feature template traversing from bottom to top and left to right in an image. Adaboost iterative algorithm is to train weak classifiers to form a strong classifier with better classification effect. Multiple strong classifiers are arranged from low to high according to their

complexity, and the detection results of each level can only be passed to the next level classifier after being screened, and the detection results can only be output after being screened by all

strong classifiers. The flow of human eye coarse positioning of Haar-Adaboost algorithm is shown in Fig. 2.

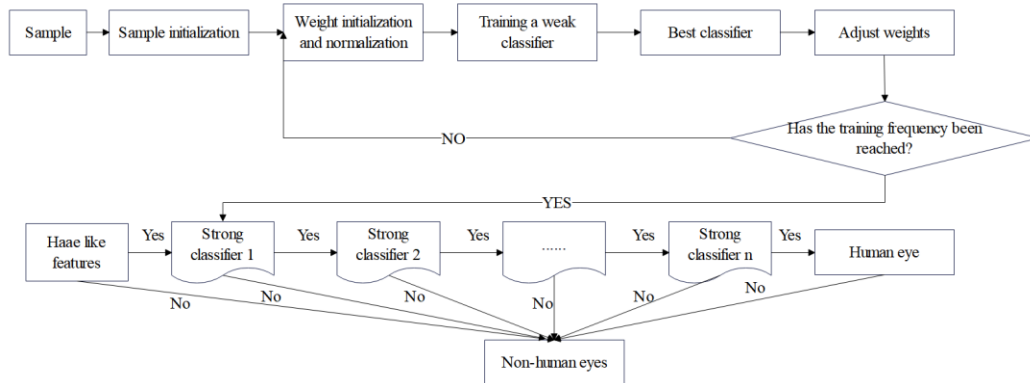
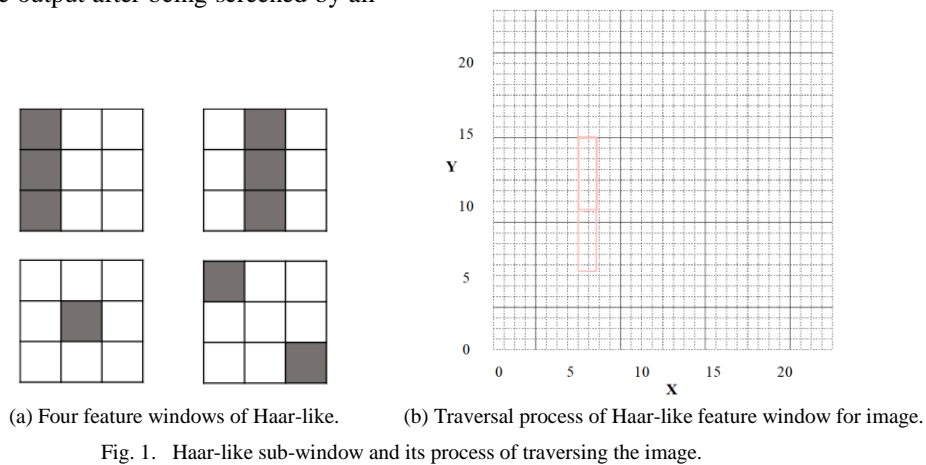


Fig. 2. Human eye coarse positioning process of Haar-Adaboost algorithm.

2) *ERT human eye fine positioning and eye movement behavior detection*: On the basis of coarse localization of human eye area, ERT algorithm is used to accurately segment human eye area and judge eye movement behavior. The ERT algorithm needs to establish a GBDT (Gradient Boosting Decision Tree), which is an object detection method based on the idea of face alignment. Human eye detection using face feature point matching is shown in Fig. 3 [20].

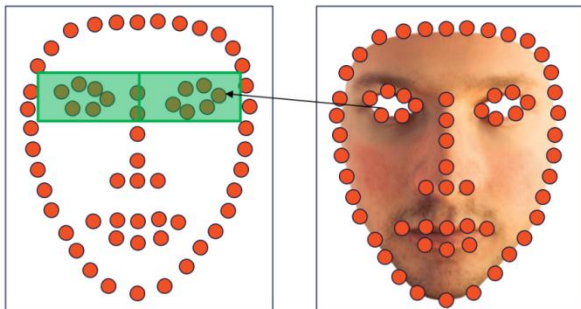


Fig. 3. Accurate positioning of human eye area based on face feature points.

The background difference algorithm can detect moving objects only by comparing the gray value of the current frame

image with the standard background gray value. The algorithm is simple and easy to implement, and it is widely used in the field of video surveillance.

The video frame image at time t is $F(x, y, t)$, the standard background image is $G(x, y, t)$, the binarized image is $B(x, y, t)$, the dynamic judgment threshold is T , and the binarized gray difference image is $D(x, y, t)$, which can be expressed as Eq. (1).

$$D(x, y, t) = \begin{cases} 1, & |F(x, y, t) - G(x, y, t)| > T \\ 0, & \text{Othes} \end{cases} \quad (1)$$

After the image is processed by Eq. (1), the part with gray value of 1 is the moving target, and the portion with a gray value of 0 is the background of the image. The limitation of background difference algorithm is that it needs to set the standard background in advance, and it requires high stability of the background. Because of the dynamic changes of the background, it is difficult to selectively reconstruct the background. In addition, when the image environment is shaken, the illumination changes suddenly, and the like, it is also

considered that a moving target appears after background difference.

Using inter-frame difference algorithm to detect moving objects in face area can achieve the purpose of detecting eye movement behavior. Similar to the principle of background difference algorithm, it assumes that the image function of the current video frame is $F(x, y)$, the image function of the first frame of the current image is $F_{k-\tau}(x, y)$, and the gray difference function $D_k(x, y)$ of the two images can be expressed by Eq. (2) [21]:

$$D_k(x, y) = |F(x, y) - F_{k-\tau}(x, y)| \quad (2)$$

The target region gray function in (2) is binarized, as shown in Eq. (3).

$$B_{D_k}(x, y) = \begin{cases} 1, D_k(x, y) > T \\ 0, Others \end{cases} \quad (3)$$

Three-frame difference method obtains the same part of two difference images by AND operation of two difference images, which avoids the image hole phenomenon of adjacent difference algorithm and enhances the robustness of the algorithm. When the three-frame difference method is used, Eq. (3) becomes Eq. (4).

$$\begin{cases} D_k(x, y) = |F_k(x, y) - F_{k-1}(x, y)| \\ D_{k-1}(x, y) = |F_{k-1}(x, y) - F_{k-2}(x, y)| \end{cases} \quad (4)$$

In Eq. (4), $D_k(x, y)$ is the expression of the difference between the k -th frame image and the $k-1$ -th frame image, and $D_{k-1}(x, y)$ is the expression of the difference between the k -th frame image and the $k-2$ -th frame image. By bringing Eq. (4) into Eq. (3), the binarization, Eq. (5) of the two difference images is obtained.

$$\begin{cases} B_{D_k}(x, y) = \begin{cases} 1, D_k(x, y) > T \\ 0, Others \end{cases} \\ B_{D_{k-1}}(x, y) = \begin{cases} 1, D_{k-1}(x, y) > T \\ 0, Others \end{cases} \end{cases} \quad (5)$$

In Eq. (5), T is the moving target judgment threshold. When $B_{D_k}(x, y)$ and $B_{D_{k-1}}(x, y)$ are equal to 1, it indicates that there are moving targets in two adjacent frames. Because the movement of the moving target is continuous and continuous, the calculation results of two consecutive differences will be 1. If the false detection is caused by noise such as holes, it will be eliminated by AND operation. The target object to be detected obtained by doing and operation in Eq. (5) is shown in Eq. (6).

$$A(x, y) = B_{D_k}(x, y) \& B_{D_{k-1}}(x, y) \quad (6)$$

In Eq. (6), $\&$ represents the AND operation, and $A(x, y)$ is the detection objective function. When the value of $A(x, y)$ is 1, it indicates that there is a moving target in the measured picture, otherwise it is stationary.

The frame difference method has a good effect in detecting eye movement behavior, but it can't judge the non-eye movement area. The three-frame difference method will still be affected by small moving objects such as hair and clothes pleats. Therefore, it is necessary to improve the three-frame difference method to improve the detection accuracy.

B. Three-Frame Difference Eye Movement Behavior Detection Algorithm

In this section, an eye movement behavior detection algorithm based on the combination of binocular position tracking in face region and three-frame difference method is proposed. The improved eye movement behavior detection algorithm is shown in Fig. 4.

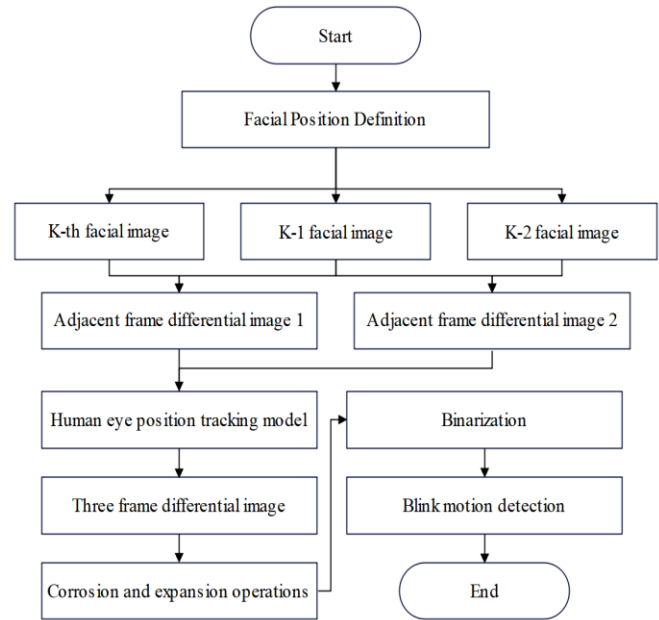


Fig. 4. Flowchart of improved algorithm.

As shown in Fig. 4, the optimization of the three-frame difference method in this project mainly includes the following points.

1) *Definition of face area:* In the face detection algorithm, the coordinate vertices of the face area are successfully found and input into the three-frame difference algorithm, and only the areas within the coordinates are differentiated. The gray value of the non-face areas outside the coordinates is 0, so that the interference of moving targets outside the face area on eye movement behavior detection is eliminated, as shown in Fig. 5.

2) *Human eye position tracking:* Before obtaining the three-frame difference map, the human eye position tracking model is introduced to estimate the position of both eyes, and the difference operation is only carried out in the human eye area, and the rest areas are set to 0, so as to eliminate the

interference of other moving targets in the face area on eye movement behavior detection.

3) *Corrosion expansion*: The image obtained by further processing the corrosion expansion operation is used to remove sharp noise around the eyes and eliminate motion noise such as eyebrows that may affect the detection results.

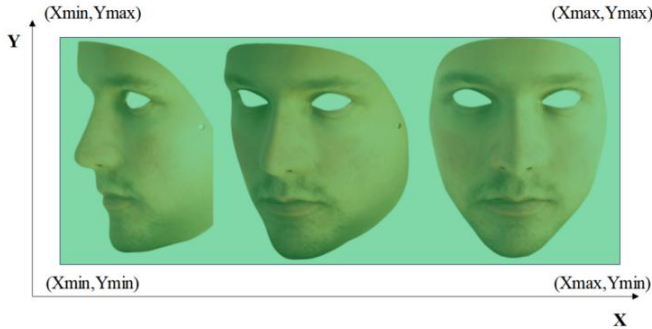


Fig. 5. Face area coordinate vertices.

All the gray values of pixels in areas other than the face are set to 0, and the coordinate level values (X_{min}, Y_{min}) , (X_{min}, Y_{max}) , (X_{max}, Y_{min}) , and (X_{max}, Y_{max}) selected by the face frame are found. When the adjacent difference operation is performed, only the difference operation is needed on the targets falling within the coordinate level value range, as shown in Eq. (7).

$$\bar{F}_k(x, y) = \begin{cases} F_k(x, y) \& M_k(X_i, Y_i), \text{Human face area} \\ 0, \text{Others} \end{cases} \quad (7)$$

In Eq. (7), $\bar{F}_k(x, y)$ represents the first image used as the adjacent difference operation, $M_k(X_i, Y_i)$ represents the detected face region function, $F_k(x, y)$ represents the first frame image, $\&$ represents an AND operation, which eliminates other regions other than the face by the AND operation of the image to be detected with the face position, and X_i and Y_i represents the face region coordinate level value:

$$\begin{cases} X_i = X_{min}, X_{max} \\ Y_i = Y_{min}, Y_{max} \end{cases} \quad (8)$$

After introducing the face region coordinate level values, the following form is obtained.

$$\begin{cases} D_k(x, y) = |\bar{F}_k(x, y) - \bar{F}_{k-1}(x, y)| \\ D_{k-1}(x, y) = |\bar{F}_{k-1}(x, y) - \bar{F}_{k-2}(x, y)| \end{cases} \quad (9)$$

After the operation of Eq. (8) and Eq. (9), the difference operation of two adjacent frames is only performed in a specific area containing the face, which can avoid bringing moving targets other than the face into the three-frame difference algorithm, and eliminate the interference of moving targets other than the face to eye movement behavior detection.

A human eye position estimation algorithm under head movement is proposed. The algorithm ideas are divided into the following points.

1) A head surveillance video of a player playing a game is selected, and the video is decomposed into 100 frames of images.

2) The improved YCbCr algorithm is used to extract the player's face image, and the lateral displacement Δx_k^h and Δy_k^h of the player's head in two consecutive frames is calculated by combining the adjacent frame difference method, and a total of 99 groups of displacement data are extracted, as shown in Eq. (10).

$$\begin{cases} \Delta x_k^h = |x_k^h - x_{k-1}^h| \\ \Delta y_k^h = |y_k^h - y_{k-1}^h| \end{cases} \quad (10)$$

3) The minimum area of the human eye position in each frame image is manually framed, the center position of the box is found, the displacement of the center position of the box is used to represent the displacement of both eyes, and the sum of the transverse and longitudinal displacement differences of the center position of the box is calculated by using the method of step 2, as shown in Eq. (11).

$$\begin{cases} \Delta x_k^e = |x_k^e - x_{k-1}^e| \\ \Delta y_k^e = |y_k^e - y_{k-1}^e| \end{cases} \quad (11)$$

4) Two groups of data of transverse displacement of human head and transverse displacement of binocular center point, longitudinal displacement of human head and longitudinal displacement of binocular center point are extracted respectively, and the functional relationship fitting is carried out to find the relationship expression of the two groups of data. The calculation of the horizontal and longitudinal displacement of the face and the horizontal and longitudinal displacement of the center of the human eye in two consecutive images is shown in Fig. 6.

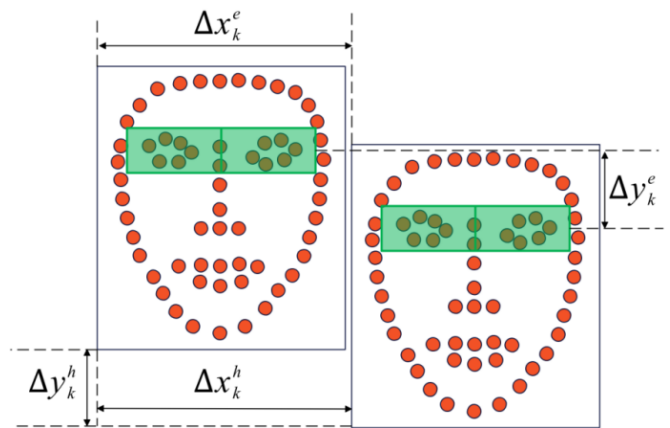
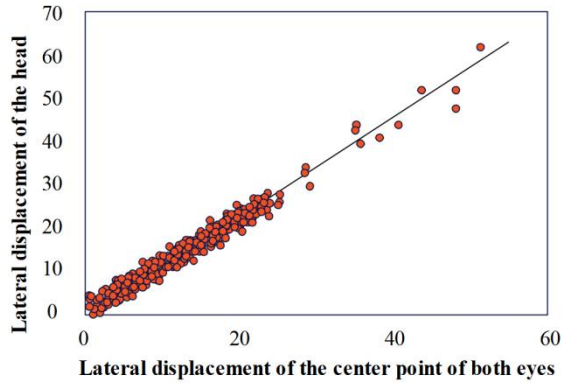
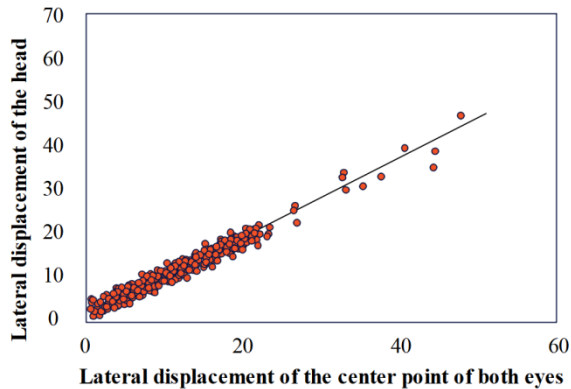


Fig. 6. Head displacement and binocular center point displacement.

The 99 groups data of Δx_k^h , Δx_k^e , Δy_k^h and Δy_k^e are counted and data relationship fitting is performed, and the results are shown in Fig. 7. Fig. 7 (a) is Relationship between the lateral displacement of the center point of both eyes and the lateral displacement of the head, Fig. 7 (b) is Relationship between longitudinal displacement of center point of binocular eyes and longitudinal displacement of head.



(a) Relationship between the lateral displacement of the center point of both eyes and the lateral displacement of the head.



(b) Relationship between longitudinal displacement of center point of binocular eyes and longitudinal displacement of head

Fig. 7. Relationship between human eye center displacement and head displacement.

As can be seen from Fig. 6, the distribution of the two sets of data shows a strong linear relationship, and the relationship between Δx_k^h and Δx_k^e and Δy_k^h and Δy_k^e is shown in Eq. (12) after data fitting.

$$\begin{cases} \Delta x^e = 1.22\Delta x^h - 1.33 \\ \Delta y^e = 0.94\Delta x^h + 0.15 \end{cases} \quad (12)$$

In Eq. (12), Δx^e represents the lateral displacement of the eye to be measured, Δy^e represents the longitudinal displacement of the eye to be measured, Δx^h represents the lateral displacement of the head of two adjacent images calculated by the adjacent frame difference method, Δy^h represents the longitudinal displacement of the head of two adjacent images calculated by the adjacent frame difference

method. Through Eq. (12), the detection range is further compressed to the human eye area, the motion noise in the skin color area is filtered, and the detection accuracy is improved.

IV. SYSTEM CONSTRUCTION AND EXPERIMENT

A. System Construction

Based on the MCIA architecture designed in this paper, the sharing framework of eye tracking data is shown in Fig. 8.

In this paper, the eye tracking and gesture data of users in the scene are captured. After that, this paper builds a collaborative scene to transmit and visualize the data of both users in the scene, thus realizing the sharing of users' eye movements and gestures in VR scenes and providing data support for visual fatigue analysis in VR games.

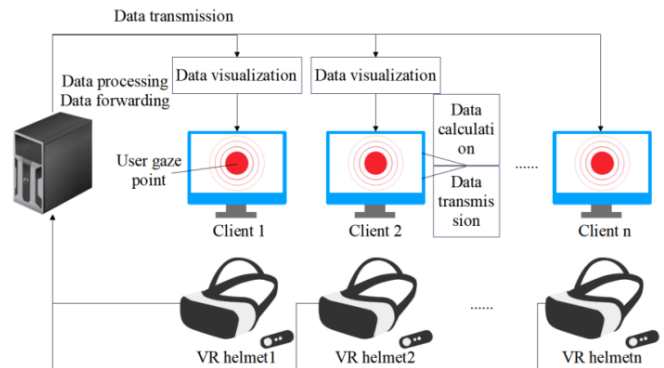


Fig. 8. Eye tracking data sharing framework diagram.

In order to verify whether the algorithm model designed in this paper is suitable for visual fatigue analysis in VR games, and to verify the influence of the best eye-hand visualization mode on the communication degree of both parties, this paper designs a VR game visual fatigue analysis system as shown in Fig. 9. The system is mainly divided into two modules: eye-hand data processing module and virtual scene interaction module. The eye-hand data processing module mainly provides data support for the virtual scene interaction module. The virtual scene interaction module will visually display user data and give users visual feedback. After receiving the visual feedback, the user's behavior is corrected again, so that the ability of collaborative interaction between the two sides is trained through the real-time interaction and real-time feedback of the system.

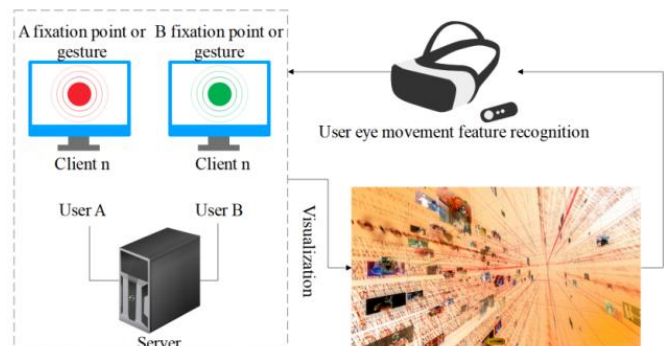


Fig. 9. System structure.

The data set in this article is a self-built data set, which is tested by 40 volunteers. These 40 volunteers are all VR game enthusiasts, so they meet the experimental verification needs of visual fatigue in this article. This paper conducts experiments through popular games, and selects two popular plot games to conduct experiments.

B. Results

In this project, the head images of 40 volunteers in VR game state are taken, and the video is decomposed into 10,000 frames. The number of eye movement behaviors is counted by using the ERT algorithm based on Haar-Adaboost feature, the unimproved three-frame difference method and the improved

three-frame difference method based on binocular position tracking in face area. The results are shown in Table I and Fig. 10 and Fig. 11.

Since this project judges whether there is visual fatigue based on the eye movement behavior rate of VR games, it is particularly important to obtain the frequency data of eye movement behavior under normal circumstances. Through the VR game test of the respondents in this paper, the frequency of eye movement behavior is artificially counted to explore the relationship between the frequency of eye movement behavior and fatigue. The results are shown in Table II.

TABLE I. COMPARISON OF STATISTICAL EFFECTS OF DIFFERENT ALGORITHMS ON EYE MOVEMENT BEHAVIOR

Algorithms	Game 1			Game 2		
	Detection rate (%)	False detection rate (%)		Detection rate (%)	False detection rate (%)	
Human eye feature ERT algorithm	94.45	4.26	Human eye feature ERT algorithm	94.45	4.26	Human eye feature ERT algorithm
Traditional three frame difference method	85.93	20.69	Traditional three frame difference method	85.93	20.69	Traditional three frame difference method
Improve the three frame difference method	91.18	8.61	Improve the three frame difference method	91.18	8.61	Improve the three frame difference method

TABLE II. CORRESPONDENCE BETWEEN EYE MOVEMENT BEHAVIOR RATE AND FATIGUE STATE (+ INDICATES NORMAL STATE, - INDICATES FATIGUE STATE)

Serial Number	Number of Times	Frequency (Times/Min)	State	Serial Number	Number of Times	Frequency (Times/Min)	State
1	53	18	+	21	69	24	-
2	43	15	+	22	30	11	-
3	41	14	+	23	16	6	-
4	48	17	+	24	57	27	-
5	43	15	+	25	22	22	-
6	48	17	+	26	27	27	-
7	34	12	+	27	25	7	-
8	50	17	+	28	62	21	-
9	39	14	+	29	28	10	-
10	49	17	+	30	52	18	-
11	33	12	+	31	21	8	-
12	40	14	+	32	60	31	-
13	42	15	+	33	58	20	-
14	53	18	+	34	27	10	-
15	30	11	+	35	53	18	-
16	37	13	+	36	46	16	-
17	41	14	+	37	35	12	-
18	57	20	+	38	28	10	-
19	29	10	+	39	61	21	-
20	31	11	+	40	19	7	-

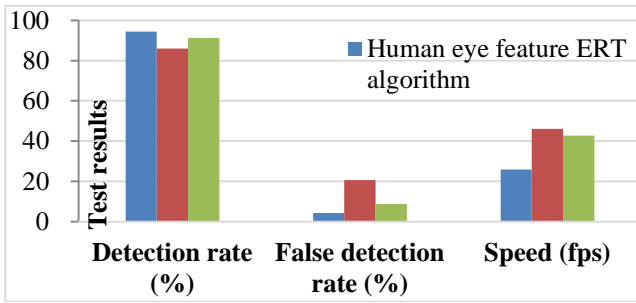


Fig. 10. Comparison of detection effects of different algorithms on eye movement behavior (Game 1).

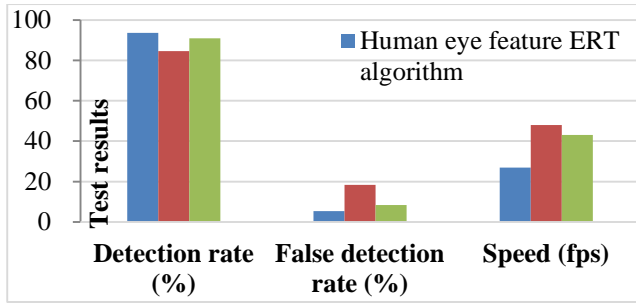
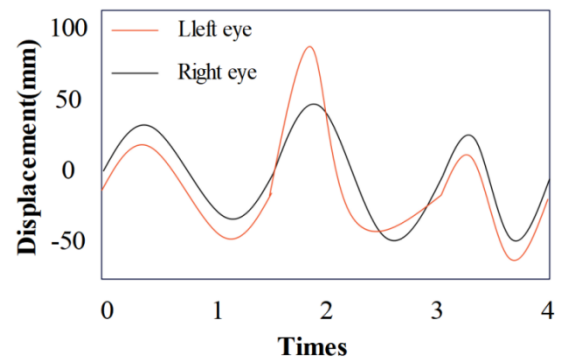


Fig. 11. Comparison of detection effects of different algorithms on eye movement behavior (Game 2).

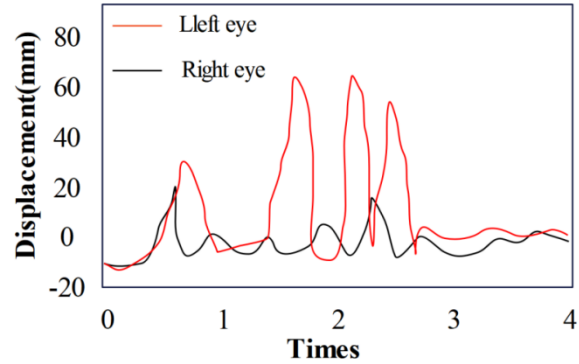
In order to verify the influence of players' body movements on the detection results of eye movement features in VR games, this paper takes racing games as an example to analyze, and designs the tilt experiment of VR racing games passing through speed bumps. When the vehicle speed is 10 km/h, the tires on one side of the car body pass through the speed bump, and the tester's head also shakes violently. The experimental results obtained by processing are shown in Fig. 12. Fig. 12 (a) shows the moving directions of the right and left eyes in the X axis, Fig. 12 (b) shows the moving directions of the right and left eyes in the Y axis, and Fig. 12 (c) shows the moving directions of the right and left eyes in the Z axis.

In the Y direction, the black line and the red line obviously cross, indicating that the movement states of the left eye and the right eye are different at this time. Furthermore, the movements of the right and left eyes are also different in the X and Z directions. The root cause of these differences is that the right and left eyes move in opposite directions on the same axis (namely, X, Y, and Z axes). For example, the difference between the right and left eye on the Y axis indicates that the left eye moves down and the right eye moves up, or the left eye moves up and the right eye moves down. Therefore, by detecting the motion states of the right eye and the left eye, it is possible to determine whether there is a lateral tilting motion of the human body.

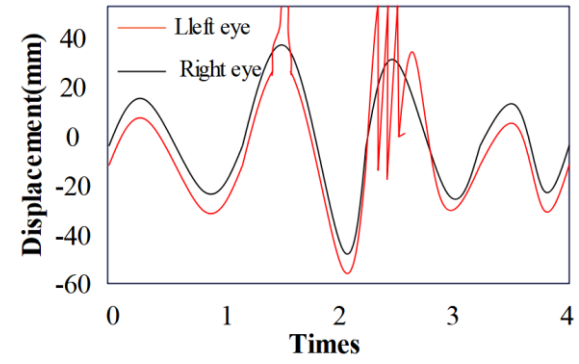
In order to further verify the accuracy of this model in VR game fatigue detection, this model is compared with study [4] (visual tracking), study [10] (EEG), study [16] (brain computer interface), and the accuracy of the above methods in VR game fatigue detection is verified through comparative tests, A total of six groups of tests were conducted, and the comparison results in Table III are obtained.



(a) X-axis



(b) Y-axis



(c) Z-axis

Fig. 12. Experimental results of roll motion in VR game.

TABLE III. COMPARISON OF VR GAME FATIGUE DETECTION ACCURACY

	Visual tracking	EEG	Brain computer interface	This study
1	69.18	77.10	80.82	91.09
2	70.46	75.07	79.23	91.18
3	72.39	76.74	78.95	88.95
4	68.84	76.10	82.61	88.37
5	72.64	82.08	79.78	90.55
6	70.80	75.16	80.67	91.83

C. Analysis and Discussion

For humans, the position of the human eye on the head is relatively fixed, and the area other than the head needs to be removed. Therefore, the area of the human eye can be located only by estimating the position of the human eye on the head. Based on the fact that the human eye is fixed at the position of

the face, when the player's head remains stationary, it is enough to detect moving targets in more than two-thirds of the skin color area of the face. However, in VR ordered scenes, the player's head is often in a constant state. In a state of shaking, finding the approximate position of the human eye when the head is shaking is the difficulty of algorithm improvement. Through observation, it can be found that when the player's head shakes, the movement trajectory of the center point of both eyes and the movement trajectory of the head show a certain regularity, and the relationship between the movement trajectory of the center point of both eyes and the displacement of the head can be located for the human eye only by finding the relationship between the movement trajectory of the center point of both eyes and the movement trajectory of the head.

The eye movement behavior detection algorithm based on human eye features needs to locate the human eye first, and then judge whether there is eye movement behavior according to the coordinate changes of human eye feature points. In the process of cascade classifier training and human eye feature matching, a large number of image samples need to be trained, and to traverse the whole detected image, the algorithm complexity is high, and it is extremely difficult to implement in FPGA development board. After observing the player's face in the game state, it is found that in the normal game state, the player's face is basically stationary, and only the eyes and mouth will appear tiny movements. Therefore, the detection of eye movement behavior can be completed only by detecting whether there is moving target in the face range, and no longer need to waste a lot of computing power for data sample training and feature point acquisition. At present, most algorithms have realized the location of face area. Therefore, the background difference algorithm is considered to detect the moving target of human face and then judge whether there is eye movement behavior.

It can be seen from Table I and Fig. 10 and Fig. 11 that the ERT algorithm based on Haar-Adaboost human eye features has the highest detection rate and the lowest false detection rate for human eye movement behavior, but the processing speed of the algorithm is the slowest among the three algorithms, and the algorithm is complex, which is difficult to implement in FPGA development board. The traditional three-frame difference algorithm has the fastest detection speed, but its detection is the lowest among the three algorithms, and the false detection rate reaches about 20%. The detection rate of eye movement behavior of the improved three-frame difference algorithm is about 91%, the average detection rate is nearly 7% higher than that of the traditional three-frame difference method, and the average false detection rate is 56% lower than that of the traditional three-frame difference method. Because the algorithm introduces the human eye position estimation model, the processing speed is lower than that of the traditional three-frame difference method, but the processing frame rate still reaches 42 fps, which meets the performance requirements of visual fatigue detection of VR games.

It is pointed out that under normal circumstances, the frequency of players' eye movement behavior is between 12-17, and with the increase of game duration and fatigue, the highest rate of eye movement behavior reaches 40 times per minute. As can be seen from Table II, under normal conditions, the

frequency of normal people's eye movement behavior is mainly distributed between 11 and 19. Considering that VR gamers are more concentrated in the game situation, and the rate of eye movement behavior is relatively low, this project will set the threshold of vision through attention to 11-19. Combined with the actual experimental data, it is found that setting the threshold of fatigue state to 11-19 times per minute has the greatest correlation with the fatigue state of VR gamers. When the eye movement behavior fatigue is 5 to 10 times per minute or more than 18 times per minute, it is determined that the VR game player has abnormal eye movement behavior phenomenon and visual fatigue.

It can be seen from Fig. 12 that the experimental results in this paper are consistent with the actual situation, which further verifies that the proposed stereo vision measurement method can be used for tracking measurement of eye movement and verifies the robustness of the algorithm under real conditions.

In Table III, visual tracking has the lowest accuracy, and the highest accuracy is only 72.64%. The recognition results of EEG and brain computer interface are similar, and the highest recognition accuracy can reach 82%. The accuracy of this research model in VR game fatigue detection can reach more than 88%, and the highest can reach 91.83%.

Through comparative analysis, it is verified that this research model has excellent performance in VR game fatigue detection

Taken together, the eye movement feature method proposed in this paper has a good effect in visual fatigue monitoring of VR games and can effectively improve the real-time analysis of player status. Therefore, integrating this algorithm into the virtual game system can help players adjust their own state, which has a positive effect on improving the game experience and reducing eye damage.

There are significant differences in the fatigue state of different individuals while playing games.

This difference is mainly reflected in individuals' reactions to the side effects of VR games. For example, some people may experience severe dizziness, nausea, eye fatigue, and overall fatigue after playing high-intensity VR games, while others may have mild or almost no symptoms. This difference is influenced by various factors, including individual sensory adaptation, age, gaming experience, and duration of continuous use of VR devices.

Specifically, an individual's sensory adaptation plays a crucial role in understanding the severity of VR vertigo. Some people are more likely to experience subjective visual vertical line changes after exposure to VR, especially under high intensity, which may be related to their milder VR dizziness symptoms. On the contrary, those who suffer from the most severe VR vertigo are unlikely to change the way they perceive vertical lines.

In addition, the study also found that women are more likely to experience screen sickness when using VR than men, which may be consistent with statistical data showing that women are also more prone to motion and screen sickness in other environments. This gender difference is particularly important in the widespread application of VR technology, as it may affect

the acceptance and user experience of VR technology among users of different genders.

In addition to gender and sensory adaptation, an individual's gaming experience can also affect their fatigue state. Novice players may feel more exhausted and uncomfortable due to unfamiliarity with the VR environment, while experienced players may be better able to adapt and reduce fatigue. Therefore, for different individuals, the fatigue state while playing games is a complex and variable problem, influenced by multiple factors.

The system has a good statistical effect on capturing eye movement behavior and eye movement behavior rate. However, although there is a great correlation between eye movement behavior rate and fatigue degree, it is difficult to accurately judge the mental state of VR game players only through this single index. In addition, due to factors such as living habits, there are differences in the frequency of eye movement behavior among individuals. Therefore, we should further eliminate individual differences through experiments, and make a comprehensive evaluation of fatigue state in combination with factors such as head posture and mouth characteristics.

V. CONCLUSION

In order to improve users' immersion and experience in VR game environment, this paper proposes a visual skin fatigue recognition algorithm based on eye movement tracking, which uses the relationship between the lateral displacement and longitudinal displacement of the human head and the displacement of the center point of the human eye to locate the position of the human eye, and inputs the human eye position tracking model into the three-frame difference algorithm to detect eye movement behavior. In addition, for tiny motion interference such as eyebrows, this paper adopts the image open operation of eroding first and then expanding to remove it. The eye movement behavior detection method adopted in this paper greatly improves the detection speed, meets the sensitivity requirements of eye movement behavior capture, and improves the real-time performance of the system with less accuracy loss. Moreover, the correlation between eye movement behavior frequency and fatigue is based on relevant reference and actual experiments, and the data are reliable.

The system has a good statistical effect on capturing eye movement behavior and eye movement behavior rate. However, although there is a great correlation between eye movement behavior rate and fatigue degree, it is difficult to accurately judge the mental state of VR game players only through this single index. In addition, due to factors such as living habits, there are differences in the frequency of eye movement behavior among individuals. Therefore, we should further eliminate individual differences through experiments, and make a comprehensive evaluation of fatigue state in combination with factors such as head posture and mouth characteristics.

REFERENCES

- [1] Merzon, L., Pettersson, K., Aronen, E. T., Huhdanpää, H., Seesjärvi, E., Henriksson, L., MacInnes, W. J., Mannerkoski, M., Macaluso, E., & Salmi, J., "Eye movement behavior in a real-world virtual reality task reveals ADHD in children," *Scientific reports*, vol. 12, no. 1, pp. 20308, 2022.
- [2] Nam, Y., Hong, U., Chung, H., & Noh, S. R., "Eye movement patterns reflecting cybersickness: evidence from different experience modes of a virtual reality game," *Cyberpsychology, Behavior, and Social Networking*, vol. 25, no. 2, pp. 135-139, 2022.
- [3] Pastel, S., Marlok, J., Bandow, N., & Witte, K., "Application of eye-tracking systems integrated into immersive virtual reality and possible transfer to the sports sector-A systematic review," *Multimedia Tools and Applications*, vol. 82, no. 3, pp. 4181-4208, 2023.
- [4] Souchet, A. D., Philippe, S., Lourdeaux, D., & Leroy, L., "Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review," *International Journal of Human-Computer Interaction*, vol. 38, no. 9, pp. 801-824, 2022.
- [5] Rappa, N. A., Ledger, S., Teo, T., Wai Wong, K., Power, B., & Hilliard, B., "The use of eye tracking technology to explore learning and performance within virtual reality and mixed reality settings: a scoping review," *Interactive Learning Environments*, vol. 30, no. 7, pp. 1338-1350, 2022.
- [6] Kim, J., Jang, H., Kim, D., & Lee, J., "Exploration of the virtual reality teleportation methods using hand-tracking, eye-tracking, and eeg," *International Journal of Human-Computer Interaction*, vol. 39, no. 20, pp. 4112-4125, 2023.
- [7] Stoeve, M., Wirth, M., Farlock, R., Antunovic, A., Müller, V., & Eskofier, B. M., "Eye tracking-based stress classification of athletes in virtual reality," *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, vol. 5, no. 2, pp. 1-17, 2022.
- [8] Alcañiz, M., Chicchi-Giglioli, I. A., Carrasco-Ribelles, L. A., Marín-Morales, J., Minissi, M. E., Teruel-García, G., Sirera, M., & Abad, L., "Eye gaze as a biomarker in the recognition of autism spectrum disorder using virtual reality and machine learning: A proof of concept for diagnosis," *Autism Research*, vol. 15, no. 1, pp. 131-145, 2022.
- [9] Mitre-Ortiz, A., Muñoz-Arteaga, J., & Cardona-Reyes, H., "Developing a model to evaluate and improve user experience with hand motions in virtual reality environments," *Universal Access in the Information Society*, vol. 22, no. 3, pp. 825-839, 2023.
- [10] Baceviciute, S., Lucas, G., Terkildsen, T., & Makransky, G., "Investigating the redundancy principle in immersive virtual reality environments: An eye-tracking and EEG study," *Journal of Computer Assisted Learning*, vol. 38, no. 1, pp. 120-136, 2022.
- [11] Dong, J., Ota, K., & Dong, M., "Why vr games sickness? an empirical study of capturing and analyzing vr games head movement dataset," *IEEE MultiMedia*, vol. 29, no. 2, pp. 74-82, 2022.
- [12] Banstola, S., Hanna, K., & O'Connor, A., "Changes to visual parameters following virtual reality gameplay," *The British and Irish Orthoptic Journal*, vol. 18, no. 1, pp. 57, 2022.
- [13] Ugwitz, P., Kvarda, O., Juříková, Z., Šašínska, Č., & Tamm, S., "Eye-tracking in interactive virtual environments: implementation and evaluation," *Applied Sciences*, vol. 12, no. 3, pp. 1027, 2022.
- [14] Lin, Y., Gu, Y., Xu, Y., Hou, S., Ding, R., & Ni, S., "Autistic spectrum traits detection and early screening: A machine learning based eye movement study," *Journal of Child and Adolescent Psychiatric Nursing*, vol. 35, no. 1, pp. 83-92, 2022.
- [15] Huygelier, H., Schraepen, B., Lafosse, C., Vaes, N., Schillebeeckx, F., Michiels, K., Note, E., Vanden Abeele, V., van Ee, R., & Gillebert, C. R., "An immersive virtual reality game to train spatial attention orientation after stroke: A feasibility study," *Applied Neuropsychology: Adult*, vol. 29, no. 5, pp. 915-935, 2022.
- [16] Hadjjaros, M., Neokleous, K., Shimi, A., Avraamides, M. N., & Pattichis, C. S., "Virtual reality cognitive gaming based on brain computer interfacing: A narrative review," *IEEE Access*, vol. 11, pp. 18399-18416, 2023.
- [17] Lu, A. S., Pelarski, V., Alon, D., Baran, A., McGarrity, E., Swaminathan, N., & Sousa, C. V., "The effect of narrative element incorporation on physical activity and game experience in active and sedentary virtual reality games," *Virtual Reality*, vol. 27, no. 3, pp. 1607-1622, 2023.
- [18] Zhou, Y., Feng, T., Shuai, S., Li, X., Sun, L., & Duh, H. B. L., "EDVAM: a 3D eye-tracking dataset for visual attention modeling in a virtual museum," *Frontiers of Information Technology & Electronic Engineering*, vol. 23, no. 1, pp. 101-112, 2023.

- [19] Jiménez-Rodríguez, C., Yélamos-Capel, L., Salvestrini, P., Pérez-Fernández, C., Sánchez-Santed, F., & Nieto-Escámez, F., "Rehabilitation of visual functions in adult amblyopic patients with a virtual reality videogame: a case series," *Virtual Reality*, vol. 27, no. 1, pp. 385-396, 2023.
- [20] Luong, T., & Holz, C., "Characterizing physiological responses to fear, frustration, and insight in virtual reality," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 11, pp. 3917-3927, 2022.
- [21] Kristjánsson, T., Draschkow, D., Pálsson, Á., Haraldsson, D., Jónsson, P. Ö., & Kristjánsson, Á., "Moving foraging into three dimensions: Feature-versus conjunction-based foraging in virtual reality," *Quarterly Journal of Experimental Psychology*, vol. 75, no. 2, pp. 313-327, 2022.

High-Accuracy Vehicle Detection in Different Traffic Densities Using Improved Gaussian Mixture Model with Cuckoo Search Optimization

Nor Afiqah Mohd Aris, Siti Suhana Jamaian

Department of Mathematics and Statistics-Faculty of Applied Sciences and Technology,
Universiti Tun Hussein Onn Malaysia, Pagoh Education Hub, Pagoh, 84600, Malaysia

Abstract—Background subtraction plays a critical role in computer vision, particularly in vehicle detection and tracking. Traditional Gaussian Mixture Models (GMM) face limitations in dynamic traffic scenarios, leading to inaccuracies. This study proposes an Improved GMM with adaptive time-varying learning rates, exponential decay, and outlier processing to enhance performance across light, moderate, and heavy traffic densities. The model's parameters are automatically optimized using the Cuckoo Search algorithm, improving adaptability to varying environmental conditions. Validated on the ChangeDetection.net 2014 dataset, the Improved GMM achieves superior precision, recall, and F-measure compared to existing methods. Its consistent performance across diverse traffic scenarios highlights its effectiveness for real-time traffic flow analysis and vehicle detection applications.

Keywords—Gaussian mixture model; vehicle detection; adaptive time-varying learning rate; exponential decay; outlier processing; cuckoo search optimization

I. INTRODUCTION

In recent years, the significance of vehicle detection and tracking systems has increased, driven by the growing demand for efficient and intelligent transportation systems. These systems play a pivotal role in diverse applications such as traffic management, accident prevention, and autonomous vehicle navigation [1-5]. To meet these demands, accurate and dependable Background Subtraction (BS) methods are essential in handling the complexities posed by dynamic and different traffic scenarios. Researchers have proposed many methods to study vehicle detection, such as traditional computer vision, machine learning, deep learning, motion-based, radar-based, and fusion techniques. This research specifically focuses on traditional computer vision methods: background subtraction in terms of mathematical contributions.

The BS method is a technique used for object detection. It involves segmenting the foreground from the background scene by generating a binary mask that identifies moving objects. The core principle of the BS method is to compute the difference between the current frame and a reference frame (background image). Thresholding techniques are then applied to classify the segmented pixels as either foreground or background. This process effectively isolates moving objects from the stationary background.

BS is a pivotal component in detection, tracking, and scene understanding, has been extensively addressed through GMM. GMM, known for their efficacy, are widely utilized to model complex and multi-modal background scenes by capturing statistical distributions of pixel intensities over time [6-8]. However, the application of traditional GMMs encounters notable challenges in the domain of vehicle detection, particularly when faced with different traffic densities. These challenges stem from the inherent limitations of constant learning rates in traditional GMMs, which are unable to adjust dynamically to the varying characteristics of the data. Traffic density and vehicle movement patterns can vary significantly over time and across environments. In such scenarios, a fixed learning rate often proves suboptimal, leading to issues such as slow convergence or convergence to suboptimal solutions, thereby impacting detection accuracy.

Moreover, traditional GMMs treat all observations equally, including outliers, which can distort the underlying data distribution and reduce the precision of the segmentation. They also assume that pixels do not closely match the mean belonging to the same statistical cluster, which may not be accurate in highly dynamic environments. These shortcomings hinder the ability of traditional models to adapt effectively to rapid changes in traffic conditions, such as those encountered in heavy or fluctuating traffic densities.

To address these limitations, this study proposes an Improved Gaussian Mixture Model (Improved GMM) that builds upon the strengths of traditional GMMs while introducing key enhancements. The Improved GMM incorporates an adaptive time-varying learning rate, which allows the model to dynamically adjust its parameters based on the characteristics of the current data. This adaptability improves performance across different traffic densities by better accommodating environmental changes.

Additionally, the model introduces exponential decay, which emphasizes pixels closer to the mean, enhancing the model's ability to distinguish between objects and background elements with higher precision. Outlier processing is also incorporated to control the influence of new covariance update observations, ensuring robustness against noisy data and outliers. These modifications collectively enable the Improved GMM to handle the complexities of vehicle detection under varying traffic densities.

Finally, to further enhance the robustness and adaptability of the Improved GMM, the Cuckoo Search (CS) optimization technique is employed for automatic parameter tuning. Inspired by the breeding behavior of cuckoos, this metaheuristic algorithm intelligently selects optimal values for critical parameters, such as the number of Gaussian components and learning rates, by exploring the parameter space efficiently. Unlike manual tuning, CS optimization dynamically adapts to the complexity of different traffic scenes, ensuring that the Improved GMM consistently delivers high detection accuracy across diverse environmental conditions with minimal human intervention.

Traffic density affects the dynamics of the background scene. In light traffic conditions, background updates may occur less frequently as there are fewer changes to the background. In contrast, the background may change rapidly in heavy traffic conditions due to the movement of multiple vehicles. Understanding and adapting to these variations in background dynamics is essential for accurate background subtraction [9-10].

The focus extends beyond the general challenges of vehicle detection to performance under different traffic densities—light, moderate, and heavy traffic conditions. The classification can be defined by the number of vehicles per square foot. Light traffic scenarios may involve less than three vehicles distributed per 500 square feet; moderate traffic represents a balance of vehicles, which is less than five vehicles distributed per 500 square feet, while heavy traffic introduces challenges, such as more than six vehicles distributed per 500 square feet. These variations necessitate developing an adaptive method to effectively address these issues while maintaining high detection accuracy and computational efficiency.

The structure of this paper is organized as follows: Section II reviews related works, providing an overview of existing approaches and background subtraction enhancements. Section III comprehensively discusses GMM and its applications in background subtraction. Section IV details the proposed Improved GMM, outlining modifications and introducing key elements. Section V describes the experimental setup and datasets utilized for evaluation, while results and discussion is given in Section VI and finally, the paper is concluded in Section VII.

II. RELATED WORKS

In the realm of computer vision, the study of background subtraction has been both significant and extensively explored. This led to the development and presentation of numerous methods and techniques to overcome diverse challenges, particularly in the context of detecting vehicles within dynamic traffic scenarios. The GMM has been widely adopted for background subtraction in computer vision applications, particularly in the vehicle detection field [7]. Initially, Friedman and Russel introduced the GMMs for background subtraction [11], while Stauffer and Grimson subsequently developed effective modified equations [6]. Numerous researchers also proposed additional modifications and improvements on the original model to enhance its performance in various traffic scenarios. Hence, this section

reviews several key developments and recent field advancements, focusing on an overview of the existing GMM methods and their improvements in background subtraction. Stauffer and Grimson denoted the GMM as a background subtraction method, which gained significant popularity due to its ability to model complex and multi-modal background scenes [6]. The study successfully achieved the objectives by capturing the statistical distribution of pixel intensities over time. Consequently, numerous researchers proposed improvements and modifications to the traditional GMM.

Zivkovic developed an adaptive GMM with a configurable number of Gaussian components, producing improved model adaptations to changing conditions [7]. Meanwhile, Zuo et al. designed an enhanced method for noise interruption for the traditional GMM [12]. The study incorporated several techniques to improve performance, including image block averaging, wavelet semi-thresholding, and adaptive background updating. Thus, the method effectively eliminated noise issues and enhanced the detection performance of the moving targets. The study also utilized an adaptive background update during the background updating phase, resulting in more accurate detection results. Another study by Lin and Chen discovered a novel method for recognizing moving objects that integrated GMM with visual saliency maps [13]. The approach effectively overcame the challenges caused by shadow situations while producing stable detection results, which transformed each image frame to the L^* , a^* , and b^* colour spaces. A Gaussian filter was then utilized to smooth the L^* , a^* , and b^* channels, eliminating small texture features and noise. The saliency maps were estimated for each channel and linearly merged to generate a comprehensive saliency map, which was combined with the foregrounds to obtain the moving objects.

Meanwhile, Zivkovic and Heijden present two efficient adaptive density estimation methods for background subtraction in video surveillance systems. The first method is based on a GMM and uses recursive equations to update model parameters and select appropriate components for each pixel. The second method is a nonparametric kernel-based approach that adapts to changes in the scene by updating the training data set. The performance of both methods is evaluated and compared to other algorithms. The results show that the nonparametric method outperforms the GMM approach in terms of accuracy but at the cost of increased processing time. This research provides valuable insights into the challenges of background subtraction and offers practical solutions for real-world applications [14]. A study by Varadarajan et al. proposes a new approach to modeling and subtracting backgrounds effectively in scenes with complex dynamic textures. The proposed method considers the spatial relationship between pixels, modeling regions as mixture distributions rather than individual pixels. In this research, the researchers derive novel online update equations using expectation maximization (EM) for modeling scenes containing dynamic textures. The effectiveness of the proposed algorithm is experimentally verified on various video sequences and compared with other well-known background subtraction algorithms. The results show that the proposed algorithm performs better than most algorithms and produces comparable results to ViBe, one of

the best background subtraction algorithms currently in the literature [15].

Another study by Cioppa et al. introduces a novel algorithm called Real-Time Semantic Background Subtraction (RT-SBS), which combines real-time background subtraction with high-quality semantic information for improved performance. The algorithm addresses the limitations of traditional background subtraction methods by leveraging semantic information provided at a slower pace and for some pixels. RT-SBS reuses previous semantic information by integrating a change detection algorithm during the decision process, ensuring real-time applicability while maintaining performance comparable to SBS. This work advances real-time background subtraction algorithms, particularly in scenarios with dynamic backgrounds, illumination changes, and moving objects [16].

Işık et al. proposed a novel method for foreground or background extraction in videos, specifically designed to address challenges posed by dynamic backgrounds. The Common Vector Approach for Background Subtraction (CVABS) leverages the Common Vector Approach (CVA) obtained through Gram-Schmidt Orthogonalization to achieve accurate background modeling. By treating background modeling as a spatiotemporal classification problem, the algorithm computes the common vector of frames to acquire the background model, enabling effective foreground object detection. The method incorporates a self-learning feedback mechanism to mitigate the impact of dynamic scenes and illumination changes on foreground detection accuracy. Experimental evaluations on diverse, dynamic backgrounds demonstrate the effectiveness of CVABS, positioning it as a competitive solution in the field of moving object segmentation [17].

In recent years, researchers also emphasized optimizing the GMM framework for specific applications, such as vehicle detection in traffic scenarios. A study by Zhang et al. described GMM with Confidence Measurement (GMMCM) as a potential solution [18]. The study addressed the susceptibility of background subtraction models towards contamination by slowly moving or temporarily stopping vehicles. Furthermore, the GMMCM incorporated a Confidence Measurement (CM) technique, assigning trust values to each pixel in the background model. This method quantified the current reliability of background pixels, and the design was developed to balance the dynamic changes in brightness and background (resolving contamination challenges) in complex urban traffic scenes. Consequently, this method was successful through a self-adaptive learning rate, which ensured the background model remained accurate. Another study by Lima et al. included a method for estimating the region-specific thresholds using a feedback step [19]. The approach employed spatial analysis to select an appropriate threshold for each region, which was utilized for pixel classification. A filtering technique was applied to the segmentation before the threshold estimate to address classification errors. This filtering process eliminated disturbances and consolidated the entire area into a cohesive unit. During the feedback phase, segmentation corrections estimated the thresholds for subsequent iterations. Notably, the filtering stage focused on correcting foreground errors, significantly enhancing the vehicle areas. This

recommended strategy facilitated the segmentation of previously segmented regions and resembled a first-order Markov chain estimate of the threshold.

In a study by Agrawal and Natu, a novel approach was developed by combining GMM with blob analysis, including labelling and morphological operations, to enhance the accuracy of foreground detection [20]. The model computed the difference between the reference frame BMG (x, y) and the current frame while applying a threshold to isolate the region of interest. In constructing the foreground model, a threshold value was selected for each pixel, which was determined using the standard deviation. A study by Luo et al. summarised a motion detection method considering spatial variation in image thresholds [21]. The approach required calculating the projected motion size under different image regions, established using a mapping correlation between the geometric motion features and the appropriate enclosing rectangle (BLOB) level in the spatial domain. This discovery enabled an adaptive threshold for each motion, effectively removing unwanted noise during motion detection.

Chen and Ellis employed a multi-dimensional Gaussian Kernel Density Transform (MDGKT) pre-processor to reduce noise in the spectral, temporal, and spatial domains [22]. This pre-processor applied spatial and temporal smoothing to each spectral component using a multivariate kernel, regarded as the product of two radially symmetric kernels. The MDGKT was a crucial component in improving the reliability of the GMM. Thus, the time interval and resolution of the GMM were changed by modifying the size of the kernel through a pair of bandwidth parameters. Kalti and Mahjoub designed a unique approach that incorporated a fuzzy distance into the Expectation-Maximisation (EM) and Adaptive Distance-based Fuzzy-C-Means (ADFCM) algorithms [23]. The pixel characterization in the study was based on two factors: the inherent attributes of the pixel and the characteristics of its surrounding neighbourhood. The classification was then measured using an adaptive distance that preferred one of the attributes concerning the pixel spatial location within the image. Another study by Wei and Zheng studied a method that calculated the L2 norm between the GMMs to measure the similarity corresponding to two pixels [24]. The study recorded the grayscale information of the pixel and the feature abundance in the local image region. Compared to individual pixels based on their differences, higher accurate pixel intensity measurements and information variation in the surrounding region were obtained. This similarity-based approach enhanced the performance of image-denoising models and preserved the detailed information in the image. Likewise, Chen and Ellis discussed an innovative approach that addressed the global illumination change concern in background model adaptation [22]. The study applied a revised adaptive strategy within the iterative learning process of the Zivkovic-Heijden GMM (ZHGM). This method was implemented by integrating a modified adaptive schedule into an existing filtering system, yielding superior performance than previous approaches (particularly in scenarios involving sudden illumination changes).

Martins et al. designed a novel classification mechanism that combined colour space discrimination, hysteresis, and

dynamic learning rate to address sudden illumination changes in the background model [25]. Each channel element (L*, a*, and b*) was analyzed individually, and the decisions obtained from each channel were merged using the AND rule, producing superior results than majority voting. This approach ensured a faster model and slower adaptations in dynamic and static regions. A higher learning rate (α_{UBG}) was applied if the pixel classification transitioned from foreground to background. Therefore, this mechanism promoted rapid adaptation when the background reappeared, effectively preventing the phantom image from emergence.

Regarding mathematical contributions, Su introduced a GMM with a data model optimization approach to address the adapting challenge of light transitions [26]. The initial step in the process included gradient picture calculation of the video stream using the Scar Operator. Subsequently, the RGB values and gradients were integrated, and noisy movement areas were eliminated using various techniques (combining the remaining sites). The two model outputs were compared to determine the final makeup area in mitigating incorrect diagnosis risk. Thus, the results demonstrated that the approach enhanced the detection process accuracy by minimizing the erroneously detected area occurrences caused by sudden illumination changes.

The advancements in GMM-based background subtraction have significantly addressed noise, dynamic textures, and illumination changes. Approaches such as RT-SBS, CVABS, and adaptive density estimation have demonstrated success in handling specific scenarios, including dynamic backgrounds and complex traffic environments. However, these methods often rely on manually tuned or fixed parameters, hindering their adaptability to fluctuating traffic densities and varying environmental conditions. This study bridges these gaps by proposing an Improved GMM that incorporates adaptive time-varying learning rates, exponential decay, and robust outlier processing, supported by CS Optimization for automatic parameter tuning. This ensures consistent and robust performance across light, moderate, and heavy traffic scenarios, contributing to the development of efficient vehicle detection systems.

III. GAUSSIAN MIXTURE MODEL IN VEHICLE DETECTION

GMM is a Mixture of Gaussians (MoG), a prominent strategy for background subtraction methods in computer vision applications. A study by Stauffer et al. discovered this strategy based on a parametric model in handling multiple modes within the pixel values [6]. The study implied that the background and foreground distributions for the GMM were Gaussian, in which the background area was more visible and exhibited smaller variances than the foreground. This assumption enabled the GMM to effectively manage slow-lighting changes and -moving objects, periodic motion, long-term scene changes, and camera noise. Conversely, the GMM was only used for its computational efficiency and excellent performance in numerous applications, as the previous assumption was not always true.

The GMM aims to construct a background model for each pixel to the time-domain distribution of pixel values in a video sequence. This model represents the weighted sum of a finite

number of Gaussian functions, which describes the multi-peak state of pixels while being suitable for complex background models (light gradients and swaying trees). In the GMM, Gaussian components with large weights represent the background, while those with small weights represent the foreground. Generally, a new pixel is part of the background if it correspondingly matches the Gaussian model. Otherwise, the pixel is treated as a foreground pixel if it does not match a Gaussian model (or match a Gaussian model with only a small weight). The efficacy of GMM has prompted various improvements and extensions in the field, which has become a widespread practice for background extraction in computer vision applications [7, 10, 26, 27, 28]. Hence, the application of GMM in vehicle detection can be expressed as in Eq. (1), where the weighted sum of K Gaussian distributions times the Gaussian component.

$$f(x_t) = \sum_{k=1}^K \Pi_{k,t} \cdot \Phi(x_t, \mu_{k,t}, \sigma_{k,t}) \quad (1)$$

where x_t is the pixel value; $\Phi(x_t, \mu_{k,t}, \sigma_{k,t})$ is the Gaussian component density with mean $\mu_{k,t}$ with covariance matrix $\sigma_{k,t}$; $\Pi_{k,t}$ is the weight associated with the k^{th} Gaussian component. Subsequently, $\Phi(x_t, \mu_{k,t}, \sigma_{k,t})$ is formulated as:

$$\Phi(x_t, \mu_{k,t}, \sigma_{k,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\sigma_{k,t}|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_t - \mu_{k,t})^T \Sigma_{k,t}^{-1} (x_t - \mu_{k,t})} \quad (2)$$

where n is the dimension of the pixel intensity. The covariance matrix is also assumed as $\sigma_{k,t} = \sigma_{k,t}^2 I$. Each new pixel value, x_t is compared with each of the existing K Gaussian distributions. A pixel is considered to match a Gaussian distribution if its value falls within a range of 2.5 standard deviations from the mean of that distribution where the matching condition is $|x_t - \mu_{k,t-1}| \leq 2.5\sigma_{k,t-1}$. The classification process involves categorizing a pixel as background if it matches with the Gaussian distribution identified as background, and as foreground if it matches with the Gaussian distribution identified as foreground. In cases where the pixel does not match with any of the K Gaussians, it is classified as foreground. This process results in the creation of a binary mask. A new Gaussian distribution is added if $k < K$, while the Gaussian distribution is replaced with the lowest priority $k = K$ ($\sigma_{k,t}^2 = \Pi_{k,t} / \sigma_{k,t}$) if $k = K$. The weights of every Gaussian distribution must be updated for the next foreground detection,

$$\Pi_{k,t} = (1 - \alpha)\Pi_{k,t-1} + \alpha\psi_{k,t} \quad (3)$$

where ψ is the indicator function and α is the constant learning rate. The mean and variance that do not find a match remain unchanged. However, for the component that does match, its mean and variance are updated according to the following criteria:

$$\mu_{k,t} = (1 - \beta)\mu_{k,t-1} + \beta x_t \quad (4)$$

$$\sigma_{k,t}^2 = (1 - \beta)\sigma_{k,t-1}^2 + \beta(x_t - \mu_{k,t})(x_t - \mu_{k,t})^T \quad (5)$$

where $\beta = \alpha \cdot \Phi(x_t, \mu_{k,t}, \sigma_{k,t})$. If the k^{th} Gaussian distribution matches x_t , then $\psi = 1$. Otherwise, $\psi = 0$ if the k^{th} Gaussian distribution does not match with x_t . The Gaussian distribution weights are then normalized after being modified. The K Gaussian distribution for each pixel is described after the modification process as:

$$B = \operatorname{argmin} \left(\sum_{k=1}^b \Pi_k > th \right) \quad (6)$$

where th is the threshold. Based on the ratio (Π/σ) , these distributions are listed by priority order, beginning with the B Gaussian distribution. Subsequently, a continuous comparison of the x_t and B Gaussian distribution is performed. The pixel is considered a background point if the x_t distribution matches any preceding B Gaussian distribution points. Alternatively, the pixel is regarded as a foreground point if it does not match, and the moving object detection is considered complete.

IV. PROPOSED IMPROVED GMM

The proposed Improved GMM model is fixed as in Eq. (1) and the Gaussian component in Eq. (2), but there are some modifications in Eq. (3) to Eq. (5), which are the updating parameters of the Gaussian component. This section presents the GMM modifications for achieving high accuracy in vehicle detection across various traffic densities, which addresses the traditional GMM limitations by introducing an adaptive time-varying learning rate, exponential decay, and outlier processing. Traditional GMMs have limitations, such as fixed learning rates, sensitivity to outliers, and difficulties in distinguishing between closely spaced objects. Therefore, these enhancements allowed the Improved GMM to effectively capture the dynamic nature of traffic density variations and improve detection accuracy. The suggested Improved GMM was founded on the improvements to the mean, covariance, and weight equations presented by Stauffer and Grimson [6]. The upgrades are as follows:

$$\mu_{k,t} = (1 - \beta(t))\mu_{k,t-1} + \beta(t)x_t \cdot e^{-\lambda d_t} \quad (7)$$

$$\sigma_{k,t}^2 = (1 - \beta(t))\sigma_{k,t-1}^2 + \beta(t)(1 - \gamma)(1 - e^{-2\lambda d_t})(x_t - \mu_{k,t})(x_t - \mu_{k,t})^T \quad (8)$$

$$\Pi_{k,t} = (1 - \alpha)\Pi_{k,t-1} + \alpha e^{-\lambda d_t} \quad (9)$$

where $\beta(t) = c/(c + t)$.

A. Adaptive Time-Varying Learning Rate

Traditional GMMs use a constant learning rate, which struggles to adapt to changing traffic densities. For example, in heavy traffic, where vehicles move closely together, or in light traffic, where vehicles are sparse, a fixed learning rate may result in slow adaptation or inaccurate background modeling. An adaptive time-varying learning rate controls the weights assigned to the newly arriving data samples [30]. This suggestion enables the algorithm to quickly adapt to traffic flow changes and detect vehicles more accurately by considering the distance between the pixel and the current means. This modification is mathematically represented by Theorem 1.

Theorem 1. The $\beta(t)$ is derived using the Robbins-Monro stochastic approximation method, which involves solving the recursive equation as:

$$\beta(t) = c/(c + t) \quad (10)$$

where c is a constant that controls the learning rate.

Proof. The Robbins-Monro stochastic approximation method is an iterative algorithm to solve root-finding issues for non-linear equations in form $f(x) = 0$, which is based on stochastic gradient descent [31, 32, 33]. The Robbins-Monro conditions are satisfied to demonstrate the validity of the update rule in Eq. (10). This validity ensures the convergence of the stochastic approximation method, which two main criteria of the Robbins-Monro conditions are as follows:

Condition 1: The sum of the learning rates $[\sum_t \beta(t)]$ should diverge and $\sum_t \beta(t) = \infty$. By evaluating the summation, a telescoping series is expressed as:

$$\sum_t \beta(t) = \frac{c}{c+1} + \frac{c}{c+2} + \frac{c}{c+3} + \dots + \frac{c}{c+t} \quad (11)$$

When the terms are rearranged, they can be written as:

$$\sum_t \beta(t) = c \left[\frac{1}{c+1} + \frac{1}{c+2} + \frac{1}{c+3} + \dots + \frac{1}{c+t} \right] \quad (12)$$

Since Eq. (12) diverges, it can reach infinity as t approaches infinity. Therefore, $\sum_t \beta(t)$ also diverges, satisfying Condition 1.

Condition 2: The $\sum_t \beta^2(t)$ should converge and $\sum_t \beta^2(t) < \infty$. By expanding and simplifying the expression, an equation is formulated as:

$$\sum_t \beta^2(t) = \frac{c^2}{(c+1)^2} + \frac{c^2}{(c+2)^2} + \frac{c^2}{(c+3)^2} + \dots + \frac{c^2}{(c+t)^2} \quad (13)$$

As in Eq. (13) converges, a finite sum is demonstrated as t approaches infinity. Hence, $\sum_t \beta^2(t)$ also converges, satisfying Condition 2. When both conditions are satisfied, the updated rule in (10) is proven valid within the context of the Robbins-Monro stochastic approximation method [34]. This updated rule ensures the learning algorithm convergence as the iteration or t approaches infinity. This strategy provides more weight to the newer data pixels while maintaining a certain importance level for the past data pixels. The constant (c) in the Robbins-Monro stochastic approximation method controls the learning rate and should be chosen based on the data characteristics and the specific application [35].

The c parameter value is selected based on a priori data knowledge, such as possible value ranges for the model parameters and the data distribution. Notably, the c value affects the performance of the algorithm, which selecting the incorrect value leads to slow convergence or instability. The value of the iteration or t is typically set to increment by one with each iteration. The t value is also interpreted as the number of algorithm iterations or observations analyzed, which the t initial value and the growth rate impact the convergence speed and algorithm stability. If the t initial value is too small, the step size can be excessively large, causing instability and

overestimating the optimal solution. Similarly, if the t initial value is too high, the step size can be extremely small, leading to slow convergence and the possibility of becoming trapped with a suboptimal solution. The t growth rate also affects the convergence speed and algorithm stability. A rapid increase in t promotes faster convergence, which leads to instability and overestimation. Alternatively, a slower growth in t produce stable behaviour, which leads to slow convergence.

B. Exponential Decay

Exponential decay improves the model's ability to handle dynamic objects and lighting changes by giving more weight to recent pixels closer to the mean. For example, when vehicles move closer to the camera, their pixels influence foreground detection more significantly. The parameter λ adjusts the contributions of each current pixel (mean, covariance, and weight parameters) to the Improved GMM. When this adjustment is incorporated, the vehicle detection precision increases to the distance between the current pixel and the current mean. The $\mu_{k,t-1}$ is the previous mean, x_t is the current pixel, and d_t is the Euclidean distance between x_t and the previous mean $\mu_{k,t-1}$. Furthermore, the introduction of the exponential decay factor ($e^{-\lambda d_t}$) from Eq. (7) to Eq. (9) ensures that the contribution of each pixel decreases as its distance from the current mean increases [36]. This characteristic is consistent with the notion that pixels closer to the mean influence the parameter changes more than those further away [37]. The additional term $(1 - e^{-2\lambda d_t})$ adjusts the contribution of each pixel to the covariance parameter based on its distance from the current mean. From Eq. (7) to Eq. (9), $\Pi_{k,t-1}$ represents the previous weight, and $e^{-\lambda d_t}$ modifies the contribution of each data point to the weight parameter with respect to its distance from the current mean. This modification assures that the contribution of each data point to the weight reduces as its distance from the current mean increases [38]. The updated parameter equation is explained in the Theorem 2:

Theorem 2. Consider $e^{-\lambda d_t}$ to be the exponential decay factor introduced from the Improved GMM ((7) to (9)) for vehicle detection in real-time traffic flow analysis. Based on the distance from the current mean, this exponential decay factor modifies the contribution of each current pixel to the mean, covariance, and weight parameters. Therefore, the equations are:

$$\mu_{k,t} = (1 - \beta)\mu_{k,t-1} + \beta x_t \cdot e^{-\lambda d_t} \quad (14)$$

$$\sigma_{k,t}^2 = (1 - \beta)\sigma_{k,t-1}^2 + \beta(1 - e^{-2\lambda d_t})(x_t - \mu_{k,t})(x_t - \mu_{k,t})^T \quad (15)$$

$$\Pi_{k,t} = (1 - \alpha)\Pi_{k,t-1} + \alpha e^{-\lambda d_t} \quad (16)$$

Proof. Consider two pixels (x_{t1} and x_{t2}), where d_{t1} is the distance between x_{t1} and the mean. In addition, d_{t2} is the distance between x_{t2} and the mean. Assume $d_{t1} < d_{t2}$. Hence, x_{t2} and d_{t2} are substituted into Equation 14 produces a new equation:

$$\mu_{k,t} = (1 - \beta)\mu_{k,t-1} + \beta x_{t2} \cdot e^{-\lambda d_{t2}} \quad (17)$$

Since $d_{t1} < d_{t2}$, the $e^{-\lambda d_{t1}}$ is greater than $e^{-\lambda d_{t2}}$. The x_{t1} contribution to the mean parameter is higher than x_{t2} . Similar

reason is extended to Eq. (15) and Eq. (16) to demonstrate that $e^{-\lambda d_{t1}}$ influences the covariance and weight parameters (in a manner that decreases the pixel contribution as its distance from the mean rises) [39].

The contribution of each pixel to the mean parameter decreases as the pixel moves away from the current mean. This outcome prevents outliers or noisy data points from significantly affecting the estimated mean [40]. If a pixel is far from the current mean, it may not accurately represent the underlying distribution, so its contribution to the mean should be downweighed. In vehicle detection, a pixel corresponds to a vehicle far from the current mean is an outlier or an erroneous detection caused by noise in the sensor data. When downweighing the contribution of each pixel to the mean, the algorithm is more robust to such outliers and minimizes the likelihood of false positives or misclassifications. Therefore, adjusting the contribution of each data point to the mean parameter based on its distance from the current mean improves the accuracy and robustness of the GMM algorithm for vehicle detection in real-time traffic flow analysis.

C. Outlier Processing

Traditional GMMs treat all data points equally, including outliers, which can distort the background model. For instance, sudden reflections, shadows, or noisy pixels may lead to false detections. To address this, a robustness parameter (γ) is added to control how much the model updates with new data [41]. This modification permits the influence control of new observations on covariance matrix updates [42]. The γ in the covariance updated equation of the GMM improves the robustness of the model against outliers and noisy data. This new equation is expressed as:

$$\sigma_{k,t}^2 = (1 - \beta)\sigma_{k,t-1}^2 + \beta(1 - \gamma)(x_t - \mu_{k,t})(x_t - \mu_{k,t})^T \quad (18)$$

The updated equation treats all observations in traditional GMM, including the outliers. Hence, the covariance matrix adapts to the outliers and incorporates their influence, potentially leading to distorted estimates of the underlying data distribution. On the contrary, the Improved GMM presented with γ allows the influence control of the new covariance update observations. When adjusting γ value, the weight provided to outlier-like observations is reduced, effectively downplaying their impact on the covariance estimation. When γ is closer to 1, the impact of new observations is reduced, making the model more robust to outliers. Robust estimators mitigate the influence of outliers or departures from model assumptions, providing more reliable parameter estimates [43].

The model responsivity is controlled by incorporating γ into the covariance update equation. This control mechanism balances integrating new information and protecting against outliers. When γ is adjusted, the weight given to outlier-like observations is reduced, allowing the model to focus more on reliable and representative data points [44]. Extensive experiments on real-world datasets were investigated in this study to assess the effectiveness of the proposed modification. The performance of the traditional GMM formulation was then compared with the modified version, incorporating an adaptive

time-varying learning rate, exponential decay, and outlier process.

V. COMPUTATIONAL SETUP

A. Algorithm

The discussion above outlines the computational algorithm for Improved GMM by introducing an adaptive time-varying learning rate, exponential decay, and outlier processing. The steps are similar to traditional GMM, except for steps (4) and steps (5), where there are modifications to the updating parameters means, variance, and weight. The algorithm can be summarized as:

Algorithm 1: Improved GMM

Initialize $K, k, \alpha, \lambda, \gamma, d$ and c

- Step 1 Obtain the current pixel value x_t from video frames.
 - Step 2 Compare the current pixel value with k existing Gaussian distributions, with a matching condition $|x_t - \mu_{k,t-1}| \leq 2.5\sigma_{k,t-1}$. A match is identified if the pixel value falls within 2.5 standard deviations of the distribution.
 - Step 3 If the matching condition is satisfied, classify the pixel values as background. Otherwise, consider the pixel value part of the foreground. If $k < K$, add a new Gaussian distribution; otherwise, replace the Gaussian distribution with the lowest priority $k = K$.
 - Step 4 Update expressions in (7) and (8) if mean and standard deviation do not match.
 - Step 5 After match inspection, update (9) and normalize it after modification.
 - Step 6 List distributions by priority order based on the ratio Π/σ , starting with the B Gaussian distribution.
 - Step 7 Conduct a continuous comparison of x_t and (6). Classify the pixel as a background point if it matches any preceding B Gaussian distribution points; otherwise, consider it a foreground point, indicating the completion of moving object detection.
-

B. Dataset

The Improved GMM was implemented using Python, leveraging libraries such as NumPy for numerical computations, OpenCV for video processing and background subtraction, Scikit-learn for Gaussian Mixture Model operations, and Matplotlib for result visualization. Jupyter Notebook served as the primary development environment, enabling interactive code execution and analysis. The experiments were conducted on a standard laptop equipped with an Intel Core i5 processor (2.50 GHz), 8 GB RAM, and Windows 11 (64-bit). The system achieved a processing speed of approximately 0.75 seconds per frame.

The evaluation of the Improved GMM was conducted using the CDNet 2014 dataset, a benchmark dataset widely used for background subtraction methods. This dataset includes videos representing various traffic densities, categorized as light traffic (less than three vehicles per 500 square feet), moderate traffic (three to five vehicles per 500 square feet), and heavy traffic (more than five vehicles per 500 square feet).

To prepare the dataset for analysis, each video was divided into individual frames for frame-by-frame processing. The

frames were resized to a consistent resolution of 640×480 pixels to ensure efficient processing. Ground truth masks provided in the dataset were used for validation by comparing them against binary masks generated by the Improved GMM. The implementation of the Improved GMM algorithm followed the modifications outlined in Section III. Key parameters were initialized. The CS algorithm was employed to automatically tune these parameters, ensuring optimal performance for various traffic conditions. During frame processing, the GMM dynamically updated its parameters using adaptive learning rates and exponential decay to refine the model, while outlier processing mitigated noise and sudden changes. Finally, the binary masks were refined through morphological operations, such as dilation and erosion, to eliminate noise and enhance detection accuracy. The experiment involved the Improved GMM method and was compared with masks generated using the traditional GMM [6], Effective Adaptive GMM (EGMM) [14], Region-based Mixture of Gaussians (RMoG) [15], Boosted GMM (BMOG) [25], Competitive Learning for Varying Input Distributions (CL-VID) [45], Real-Time Sematic Background Subtraction Version 2 (RT-SBS-V2) [16] and Common Vector Approach Background Subtraction (CVABS) [17].

C. Cuckoo Search Optimization

The CS optimization algorithm was employed to automatically tune the key parameters of the Improved GMM, including $c, K, \alpha, \lambda, \gamma$, and threshold (th). This algorithm, inspired by the breeding behavior of cuckoos, efficiently explores the parameter space using a balance between local and global search mechanisms. To determine the appropriate parameter ranges, prior literature and empirical experimentation were consulted. Table I provides the initial ranges for each parameter:

TABLE I. RANGE PARAMETERS OF OPTIMIZED PARAMETERS

Parameters	Range
c	$0.01 \leq c \leq 0.1$
K	$2.0 \leq K \leq 5.0$
α	$5.0 \leq \alpha \leq 50.0$
λ	$0.01 \leq \lambda \leq 1.0$
γ	$0.1 \leq \gamma \leq 1.0$
th	$3.0 \leq th \leq 7.0$

The CS algorithm began by initializing a population of candidate solutions, each representing a unique set of parameter values within the predefined ranges. These candidates were evaluated using a fitness function designed to maximize the accuracy of the Improved GMM's background subtraction. The fitness function compared the binary masks generated by the model with the ground truth annotations, considering metrics such as precision, recall, and F-measure. During each iteration, the algorithm updated the candidate solutions by simulating the levy flights of cuckoos, which allow for both local fine-tuning and global exploration of the parameter space. Poor-performing solutions were replaced by better-performing ones, ensuring that the algorithm converged towards an optimal set of parameter values.

Default settings for CS algorithm hyperparameters were replaced with experimental investigations to determine the most effective parameter combinations for the Improved GMM. Specifically, the number of Gaussian components (N) and the probability of adaptation (Pa) were key factors in optimizing the background modeling process.

The number of Gaussian components (N) represents the number of distributions used to model each pixel's background, which is crucial in capturing the complexity of background variations. A higher N allows for more sophisticated background modeling, enabling the algorithm to better handle complex scenes with multiple background states such as changing lighting conditions, moving shadows, or repetitive background patterns. The probability of adaptation (Pa) controls the rate at which the model updates its background distributions. By experimenting with different values of N and Pa, we aim to find the optimal configuration that provides the most accurate and robust background subtraction across various traffic scenarios. Table II and Table III illustrate the experimental results for different values of N and Pa.

TABLE II. EXPERIMENTAL RESULTS FOR DIFFERENT VALUES OF N USING PA=0.25 OVER CDNET2014 DATASET

Frame	Input	N=10	N=30	N=50
#0672				
#0808				
#1328				
#1517				

TABLE III. EXPERIMENTAL RESULTS FOR DIFFERENT VALUES OF PA USING N=30 OVER CDNET2014 DATASET

Frame	Input	Pa=0.25	Pa=0.5	Pa=0.75
#0672				
#0808				
#1328				
#1517				

Table II presents the experimental results for different values of N (number of Gaussian components in the GMM) while keeping the probability parameter Pa = 0.25 constant. It showcases how varying the number of Gaussian components

affects the background subtraction performance across multiple frames. Increasing N typically balances model complexity and overfitting, influencing detection accuracy. Table III evaluates the impact of varying probability thresholds for background classification using a fixed N=30. The results demonstrate how different threshold values influence the system's sensitivity to classify pixels as foreground or background, emphasizing the importance of parameter tuning for effective detection in varying conditions.

The combined insights from these tables underscore the sensitivity of the GMM model to these parameters and highlight the necessity for optimization methods, such as CS Optimization, to automatically determine the best parameter values for achieving optimal detection performance. Through the CS optimization method, the algorithm converged on an optimal set of parameters, as shown in Table IV:

TABLE IV. OPTIMIZED PARAMETER OF IMPROVED GMM AUTOMATIC TUNING BY CS OPTIMIZATION METHOD

c	K	α	λ	γ	th
0.1	2.0	9.1	1.0	0.6	3.2

The effectiveness of the automatic parameter tuning was compared against empirical tuning, as demonstrated in Table V. The automatic tuning approach demonstrates improved adaptability, producing masks closer to the ground truth across different frames. The empirical tuning approach may suffer from inconsistency due to the lack of parameter optimization for specific scenarios. The visual comparison indicates that automatic tuning effectively reduces background noise and captures more accurate object contours, particularly in challenging scenarios.

TABLE V. COMPARISON OF FOREGROUND MASKS OF AUTOMATIC TUNING PARAMETER OF IMPROVED GMM AND EMPIRICAL TUNING PARAMETER OF IMPROVED GMM

Frame	Input	Ground Truth	Automatic Tuning	Empirical Tuning
#0672				
#0808				
#1328				
#1517				

Table VI evaluates the average performance metrics of the Improved GMM with automatic tuning (optimized parameters) and empirical tuning (manually set parameters). The values demonstrate that the automatic tuning of parameters using CS optimization substantially outperforms empirical tuning. The improved precision reduced false positives, and higher overall accuracy highlight the importance of optimization in achieving robust and reliable vehicle detection.

TABLE VI. AVERAGE PERFORMANCE PARAMETER EVALUATION OF AUTOMATIC TUNING IMPROVED GMM AND EMPIRICAL TUNING IMPROVED GMM

Tuning	RC	PR	FM	FPR	FNR	WC	AC
Automatic	0.533	0.634	0.579	0.033	0.467	0.077	92.3%
Empirical	0.480	0.292	0.422	0.232	0.160	0.225	77.5%

Fig. 1 provides a visual comparison of the accuracy values between automatic and empirical tuning methods. The accuracy achieved with automatic tuning is 92.3%, as observed in Table VI. This high value underscores the effectiveness of automatic tuning in optimizing the Improved GMM parameters for robust and consistent detection across frames. Empirical tuning achieves an accuracy of 77.52%, significantly lower than automatic tuning. The drop in performance highlights the limitations of manually setting parameters, which are less adaptable to variations in input conditions such as different lighting, occlusions, or traffic densities. The results validate the incorporation of CS optimization as a robust method for parameter optimization in vehicle detection tasks.

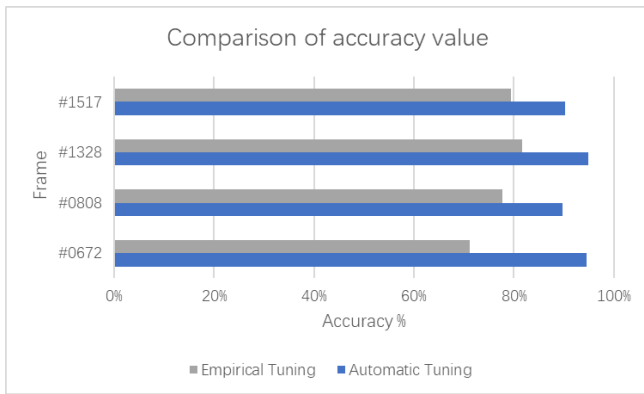


Fig. 1. The core principle of the BS method (a) input image (b) reference image (c) foreground or background.

D. Evaluation Metrics

The poor performance of methods that use GMM for background subtraction is mostly caused by assumptions about the parameters, where different issues produce varying effects on methods performance. Additionally, performance evaluation metrics showcase the advantages of background subtraction methods in vehicle detection accuracy. The pixel features from vehicle detection results were divided into two groups (foreground considered positive and background considered negative) to evaluate the detection accuracy objectively [18].

Pixel-based performance evaluation metrics were widely employed to assess the accuracy of image segmentation algorithms. These metrics compared the pixel-wise segmentation results obtained by the algorithm to the ground truth (GT) annotations. Once the GT was determined, several generally accepted methods were compared to a proposed binary foreground map. In this evaluation, four types of pixels are utilized as follows:

- True positive (TP)
- False positive (FP)

- False negative (FN)
- True negative (TN)

The TP represents pixels correctly classified as positive by the segmentation algorithm and positive in the GT annotation. Meanwhile, FP represents pixels classified as positive by the segmentation algorithm but are negative in the GT annotation. The FN represents negative pixels in the segmentation algorithm but positive in the GT annotation. Subsequently, TN represents pixels correctly classified as negative by the segmentation algorithm and negative in the GT annotation [46].

Based on these four types of pixels, several metrics were computed to evaluate the performance of the segmentation algorithm, including recall (RC), precision (PR), F-measure (FM), FP rate (FPR), FN rate (FNR), and accuracy (AC). The RC is a metric used to evaluate the ability of an algorithm to correctly identify positive pixels, which is calculated as the ratio of TP pixels to the total number of ground TP pixels (also known as a TP rate). The TP rate measures the fraction of foreground pixels accurately identified out of the total number of foreground pixels, which the algorithm has categorized as [47]:

$$RC = TP / (FN + TP) \quad (19)$$

Since PR is the fraction of TP pixels over the number of positive pixels classified by the segmentation algorithm, the proportion of positive predictions that are TPs is measured given by:

$$PR = TP / (TP + FP) \quad (20)$$

The FM, or F1 score, is the harmonic mean of PR and RC. Therefore, FM provides a single metric to evaluate the overall performance of the algorithm as follows:

$$FM = (2 \times RC \times PR) / (RC + PR) \quad (21)$$

The FPR is the fraction of FP pixels over the total GT negative pixels. Hence, the proportion of negative predictions that are FPs is expressed as:

$$FPR = FP / (FP + TN) \quad (22)$$

In FNR, the fraction of FN pixels over the total number of GT positive pixels is obtained. Therefore, the proportion of positive pixels that are missed by the algorithm is described by:

$$FNR = FN / (FN + TP) \quad (23)$$

The AC is the fraction of correctly classified pixels over the total number of pixels and measures the overall correctness of the segmentation of the algorithm is represented by:

$$AC = (TP + TN) / (TP + FP + FN + TN) \quad (24)$$

Finally, wrong classification (WC) is the fraction of wrongly classified pixels over the total number of pixels. The WC also measures the overall error rate of the segmentation of the algorithm is expressed by:

$$WC = (FP + FN) / (TP + FP + FN + TN) \quad (25)$$

These seven indicators explain and evaluate the performance of the proposed model. Generally, these metrics provide a quantitative evaluation of the performance of the segmentation algorithm while aiding in identifying areas where the algorithm requires improvement [48].

VI. RESULT AND DISCUSSION

To validate the effectiveness of the proposed Improved Gaussian Mixture Model (GMM) [29], extensive experiments were conducted using over 1700 frames of moving vehicles on the road. Table VII compares the Improved GMM method with several state-of-the-art background subtraction and vehicle detection methods. The table includes the evaluation metrics used in this study and the results obtained for each method under different traffic densities.

TABLE VII. SUMMARY OF GMM, EGMM, RMOG, BMOG, CL-VID, RT-SBS-V2, CVABS AND IMPROVED GMM AVERAGE METRICS

Method	RC	PR	FM	FPR	FNR	AC	WC
GMM [6]	0.972	0.806	0.891	0.026	0.002	0.970	0.024
EGMM [14]	0.970	0.783	0.876	0.029	0.004	0.970	0.027
RmoG [15]	0.970	0.662	0.795	0.047	0.005	0.957	0.043
BMOG [25]	0.949	0.814	0.906	0.026	0.002	0.974	0.023
CL-VID [45]	0.967	0.849	0.922	0.019	0.013	0.975	0.018
RT-SBS-V2 [16]	0.973	0.790	0.881	0.031	0.004	0.969	0.028
CVABS [17]	0.916	0.849	0.886	0.020	0.084	0.972	0.028
Improved GMM	0.959	0.835	0.890	0.026	0.041	0.973	0.027

A high RC value was desirable to minimize FNs. The FNs occurred when actual positive instances were incorrectly classified as negative, presenting missed detections or opportunities. Thus, a high RC indicated that the model effectively captured the majority of positive instances and was less likely to overlook essential or critical cases [48]. The RC value of the Improved GMM outperformed a few methods with a value of 0.9586. This improvement is particularly impactful in scenarios with heavy traffic densities, where high vehicle overlap increases the likelihood of misclassification in traditional methods. The adaptability of the Improved GMM allows it to differentiate subtle variations in pixel distributions, reducing false negatives in densely packed vehicle scenarios, such as urban intersections or peak highway traffic conditions.

Alternatively, a high PR value suggested that the methods produced a lower FP rate, which could accurately identify positive pixels with a lower tendency to incorrectly classify background pixels as foreground. In other words, a high PR denoted that the methods of a pixel as foreground was more likely to be accurate. The PR value for the Improved GMM was 0.8353, which only slightly differed from the highest values from CL-VID and CVABS. This demonstrates the Improved GMM's effectiveness in reducing false positives, particularly in moderate traffic conditions where objects such as pedestrians, shadows, or reflections might otherwise be misclassified. By maintaining a robust precision level, the Improved GMM ensures reliable vehicle detection, crucial for real-world applications like urban traffic monitoring or congestion management.

The FM is a widely used metric that combines PR and RC to evaluate the overall performance of a method. A higher FM indicated that the methods significantly balanced PR and RC, effectively identifying positive pixels while minimizing FPs. Therefore, a higher FM value suggested better performance. Meanwhile, the FM value for the Improved GMM was among the highest among the other state-of-the-art methods. This positions the Improved GMM as a balanced performer, particularly in traffic scenarios with moderate density, where both precision and recall are crucial for maintaining detection reliability. Compared to CL-VID and BMOG, which excel in heavy and light traffic respectively, the Improved GMM demonstrates consistent and reliable performance across varied traffic densities, making it a versatile choice for real-world applications. The optimal FM value depended on the specific application and the desired trade-off between PR and RC [48].

A low FNR was considered desirable in most situations in which the system accurately identified positive instances and minimized missed detections or incorrect negative predictions. For example, the Improved GMM achieved a lower FNR compared to traditional GMM (0.041 vs. 0.084), highlighting its superior ability to identify true positives effectively. This improvement is particularly beneficial in high-density traffic scenarios, where the risk of missed vehicle detections is higher due to occlusions. Like FPR, the desired FNR depended on the specific context and application [49]. In some cases, a trade-off could occur between the FNR and other factors, including FPR or the overall system cost. The appropriate balance could vary depending on the situation's specific goals, constraints, and acceptable risks. Nonetheless, a low FNR was generally preferable to ensure the highest AC and detection performance [49].

The AC metric measures the percentage of correctly classified pixels in the foreground mask. A high AC suggested that the methods successfully classified a larger proportion of positive and negative pixels relative to the total number of pixels. This value accurately reflected the ability of the methods to distinguish between foreground and background regions. Lastly, WC referred to incidents in which a classification or prediction system improperly categorized them. For instance, in a real-world traffic monitoring system deployed at a busy urban intersection, high WC rates could lead to incorrect vehicle counts or misclassifications, potentially compromising traffic flow optimization or accident detection. The Improved GMM's ability to minimize WC ensures more accurate vehicle detection and tracking, leading to improved reliability in such critical applications. Depending on the applications, high WC rates could produce errors, misinterpretations, and negative consequences [48].

Fig. 2 illustrates the comparative recall performance of eight computational methods across three different traffic densities: light, moderate, and heavy. Recall is a metric that measures the percentage of true positive detections from all the actual positive instances in the video. Recall measures the percentage of true positive detections from all actual positive instances in the video. The graph reveals that some methods, such as CL-VID and RT-SBS-V2, maintain consistently high recall in specific conditions, while others show greater variability. For instance, GMM demonstrates a slight edge in

heavy traffic scenarios, likely due to its strong adaptability to high-density environments. Meanwhile, RT-SBS-V2 excels in light conditions, benefiting from its sensitivity to minimal background dynamics. Conversely, the CVABS method demonstrates a notable decline in recall under moderate traffic conditions compared to its performance in light and heavy conditions. This inconsistency may indicate specific weaknesses, such as sensitivity to moderately complex environments with fluctuating vehicle densities or partial occlusions. By analyzing the recall trends across traffic densities, Fig. 2 highlights the strengths and areas for improvement in the Improved GMM, emphasizing its potential as a balanced performer for diverse real-world applications.

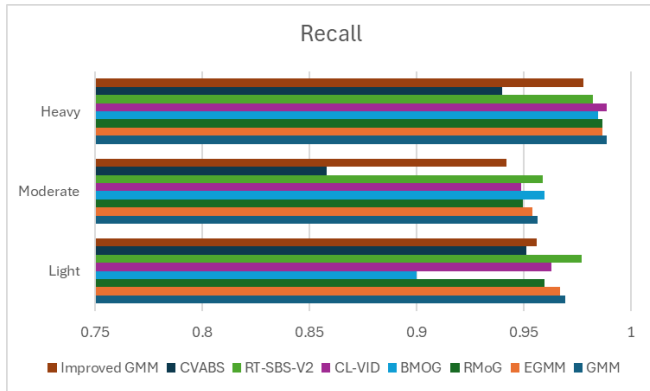


Fig. 2. The comparison of recall results of our proposed method and the other methods.

The PR values are a key performance indicator, particularly in fields with high costs of false positives. Fig. 3 shows that the method labelled CL-VID shows the highest PR value in the heavy category, underscoring its exceptional accuracy under challenging conditions where vehicle overlap, and background complexity are prominent. This suggests that CL-VID is particularly adept at minimizing false positives when the scene dynamics are most intricate. In contrast, RMoG exhibits lower PR values, especially in the moderate traffic category. This performance discrepancy could be attributed to its limitations in handling medium-complexity scenarios, such as moderate occlusions or partially visible vehicles. The proposed Improved GMM achieves consistently high precision across all traffic densities, highlighting its robustness and versatility. While it does not outperform CL-VID in the heavy category, its balanced performance across light, moderate, and heavy traffic conditions signifies its reliability as a general-purpose solution for diverse environments. This generalist approach ensures that the Improved GMM maintains low false-positive rates regardless of traffic complexity, making it a dependable choice for real-world applications.

BMOG and CVABS methods show competitive PR values, particularly in moderate and heavy traffic scenarios. Their ability to sustain high precision in demanding conditions emphasizes their effectiveness in environments with increased vehicle density and dynamic lighting changes. A noteworthy observation is that PR values exhibit less variation across methods compared to RC. This stability indicates that these methods generally maintain a consistent ability to identify true positives. However, differences in precision suggest that their

effectiveness in rejecting false positives varies, which is critical in applications requiring high reliability and minimal false alarms.

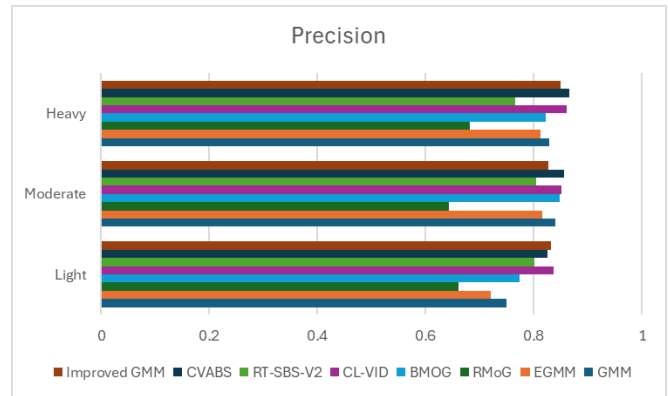


Fig. 3. The comparison of precision results of our proposed method and the other methods.

Fig. 4 illustrates the FM results of eight different background subtraction methods: GMM, EGMM, RMoG, BMOG, CL-VID, RT-SBS-V2, CVABS, and the proposed method. As a harmonic mean of precision and recall, FM provides a comprehensive measure of segmentation accuracy, making it a crucial metric for evaluating the effectiveness of these methods. The graph highlights the superior performance of CL-VID in scenarios with moderate and heavy background motion, showcasing its robustness and adaptability in handling complex conditions such as dynamic lighting and high traffic density. This positions CL-VID as a strong candidate for scenarios requiring high reliability in challenging environments. In simpler conditions, the BMOG method achieves the highest FM in the light traffic category. This indicates its ability to excel in straightforward segmentation tasks where background dynamics are less pronounced, making it suitable for environments with minimal motion complexity.

The Improved GMM demonstrates consistent performance across all traffic categories. While it does not achieve the top FM score in any specific category, its stability across light, moderate, and heavy conditions highlights its versatility and reliability as a general-purpose solution. This balanced performance makes the Improved GMM particularly suitable for applications requiring robust results across diverse operational scenarios. In contrast, RMoG exhibits lower FM values across all categories, signalling potential limitations in adapting to varying background complexities. This underperformance may stem from its inability to handle dynamic textures or abrupt changes effectively. A general trend observed from the results is that most methods perform better in light and moderate categories, with a slight decline in heavy traffic conditions. This trend underscores the increasing challenge posed by higher background complexity and overlapping objects in heavy traffic scenarios, which can impact segmentation accuracy.

The stable performance of the Improved GMM across different conditions underscores its potential for applications requiring consistent and reliable segmentation, such as real-time traffic monitoring and video analytics. While other

methods, such as CL-VID and BMOG, excel in specific conditions, the Improved GMM offers a balanced approach, ensuring dependable performance irrespective of environmental variability.

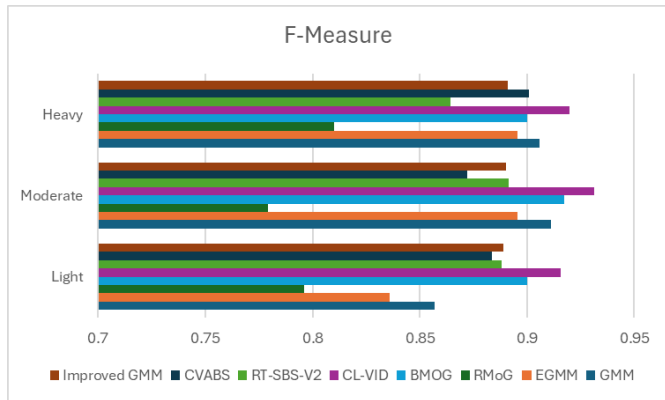


Fig. 4. The comparison of F-measure results of our proposed method and the other methods.

Fig. 5 illustrates the AC results of various background subtraction methods under different traffic conditions. The data reveals that all methods achieve high accuracy, with most exceeding 0.95, reflecting their overall effectiveness in segmentation tasks. However, noticeable differences emerge when comparing performances across specific traffic conditions. BMOG and CL-VID stand out with superior accuracy in light and moderate conditions, respectively, demonstrating their specialised efficiency in less complex environments. Their performances converge under heavy traffic conditions, suggesting that their methodologies are similarly adept at handling challenging scenarios with high background complexity and vehicle overlap. The proposed Improved GMM maintains consistently high accuracy across all traffic densities, highlighting its robustness and adaptability. This consistency suggests that the Improved GMM effectively balances precision and recall, making it well-suited for diverse operational contexts, including scenarios with fluctuating background dynamics. While the RMoG method achieves relatively high accuracy, its lower performance in moderate and heavy conditions indicates potential limitations in handling dynamic or complex backgrounds. This shortfall could be attributed to its reduced ability to effectively manage abrupt changes or intricate textures.

The subtle variations in accuracy among these methods carry significant implications for practical applications where precision is critical. For example, in high-security environments or scenarios requiring detailed video analysis, even small accuracy differences can influence the choice of method. The slight decline in accuracy under heavy traffic conditions observed for most methods emphasizes the need for enhanced robustness against complex backgrounds, direction future research should explore. The consistent performance of the proposed Improved GMM suggests a well-balanced integration of techniques tailored to address varying complexities. This adaptability makes it a compelling option for researchers aiming to develop versatile background subtraction methods that perform reliably across diverse conditions.

Table VIII compares foreground masks obtained using eight different methods. The table shows each method's original frame (input), ground truth (GT), and segmentation masks. The frames displayed are from three different traffic densities:

- 1) Frame 1213: Heavy traffic scenario
- 2) Frame 1480: Moderate traffic scenario
- 3) Frame 795: Light traffic scenario

The first column displays the original input frames, representing the raw video data captured by the camera. The second column shows the GT, which consists of manually annotated masks indicating the precise locations of vehicles in the respective frames. The subsequent columns (third to tenth) present segmentation masks produced by the different methods, allowing a direct comparison against the GT and input frames.

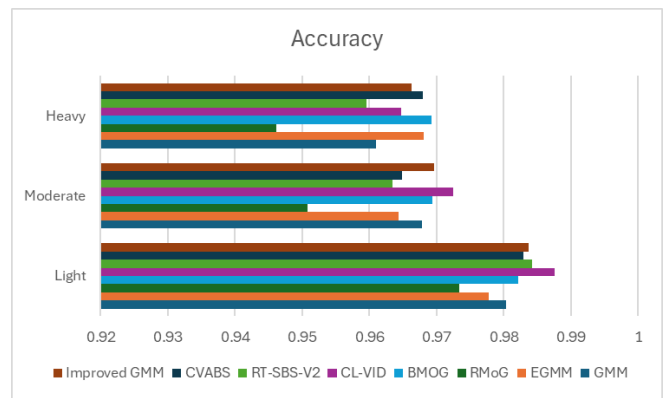
































Fig. 5. The comparison of accuracy results of our proposed method and the other methods.

As illustrated in Table VIII, the Improved GMM demonstrates satisfactory detection performance across all traffic densities. For example, in the heavy traffic scenario, the Improved GMM effectively captures vehicle locations and shapes with notable accuracy, achieving a balance between minimizing false positives and negatives. While the Improved GMM exhibits competitive performance, the results also reveal the distinct strengths and weaknesses of other methods. For instance, CL-VID and BMOG show strong performance in scenarios with moderate traffic, excelling in precision and segmentation clarity. Meanwhile, EGMM and CVABS perform well in light traffic scenarios, accurately identifying individual vehicles with minimal background noise.

False positives and false negatives are present to varying degrees across all methods, including the Improved GMM. This indicates the inherent challenges of background subtraction in dynamic traffic environments, such as occlusions, varying lighting conditions, and overlapping vehicles. The Improved GMM method aims to strike a balance between computational efficiency and segmentation accuracy, making it a viable option for real-time applications where processing time is a critical constraint. However, the increased complexity introduced by the Improved GMM, particularly due to features like adaptive learning rates and exponential decay, warrants further investigation to fully assess its scalability and efficiency under diverse operational scenarios.

TABLE VIII. COMPARISON OF SEGMENTATION MASKS FOR THE CDNET2014 DATASET

Traffic	Light Traffic	Moderate Traffic	Heavy Traffic
Input			
GT			
GMM [6]			
EGMM [14]			
RMoG [15]			
BMOG [25]			
CL-VID [45]			
RT-SBS-V2 [16]			
CVABS [17]			
Improved GMM			

VII. CONCLUSION AND FUTURE WORK

This study proposed an Improved GMM for high-accuracy vehicle detection across varying traffic densities. By incorporating an adaptive time-varying learning rate, exponential decay, and outlier processing, the model effectively addressed limitations in traditional GMM methods, such as fixed learning rates and sensitivity to outliers. The use of CS Optimization for automatic parameter tuning further enhanced the robustness and adaptability of the model. Experimental results demonstrated that the proposed method consistently achieved superior performance metrics, including accuracy, precision, recall, and F-measure, across light, moderate, and heavy traffic scenarios.

Despite these promising results, several limitations warrant consideration. First, the computational complexity of the proposed model may present challenges for deployment on low-resource systems. Second, the model's reliance on hand-crafted features might limit its adaptability to rapidly evolving traffic scenarios. Future work should explore integrating deep learning techniques to complement the Improved GMM for enhanced robustness and scalability. Additionally, incorporating real-time optimization and reducing

computational overhead will be critical for broader adoption in resource-constrained environments.

In conclusion, the proposed method offers a significant step toward reliable and efficient vehicle detection in diverse traffic scenarios, paving the way for further advancements in intelligent transportation systems.

ACKNOWLEDGMENT

The authors are grateful to the anonymous reviewers and editors for their valuable suggestions and comments. This research was supported by the Ministry of Higher Education (MOHE) through Fundamental Research Grant Scheme (FRGS/1/2021/STG06/UTHM/02/1).

REFERENCES

- [1] D. Y. Ge, X. F. Yao, W. J. Xiang, and Y. P. Chen, "Vehicle detection and tracking based on video image processing in intelligent transportation system," *Neural Computing and Applications*, vol. 35, no. 3, pp. 2197–2209, Jan. 2023.
- [2] K. Zhong, Z. Zhang, and Z. Zhao, "Vehicle detection and tracking based on GMM and enhanced camshift algorithm," *Journal of Electrical and Electronic Engineering*, vol. 6, no. 2, pp. 40–45, Apr. 2018.
- [3] N. A. Mohd Aris and S. S. Jamaian, "Background subtraction challenges in motion detection using Gaussian mixture model: a survey," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, no. 3, pp. 1007-1018, 2023.
- [4] R. Deng, D. Yang, X. Liu, and S. Liu., "A background subtraction algorithm based on pixel state," in *Proceedings of the 13th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and its Applications in Industry*, pp. 251–254, Nov. 2014.
- [5] T. Indu, Y. Shivani, A. Reddy, and S. Pradeep, "Real-time classification and counting of vehicles from CCTV videos for traffic surveillance applications," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 14, no. 2, pp. 684–695, May 2023.
- [6] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '99)*, pp. 246–252, Jul. 2006.
- [7] Z. Zivkovic, "Improved adaptive Gaussian mixture model for background subtraction," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, pp. 28–31, Aug. 2004.
- [8] S. Rajkumar, A. Hariharan, S. Girish, and M. Arulmurugan, "An efficient vehicle detection and shadow removal using Gaussian mixture models with blob analysis for machine vision application," *SN Computer Science*, vol. 4, no. 5, pp. 451, Jun. 2023.
- [9] L. Alandkar and S. R. Gengaje, "Dealing background issues in object detection using GMM: a survey," *International Journal of Computer Applications*, vol. 150, no. 5, pp. 0975–8887, Sept. 2016.
- [10] T. Bouwmans, F. El Baf, and B. Vachon, "Background modeling using a mixture of Gaussians for foreground detection—a survey," *Recent Patents on Computer Science*, vol. 1, no. 3, pp. 219–237, Nov. 2008.
- [11] N. Friedman and S. Russell, "Image segmentation in video sequences: a probabilistic approach," *13th Conference on Uncertainty in Artificial Intelligence*, 175-181, 1997.
- [12] J. Zuo, Z. Jia, J. Yang, and N. K. Kasabov, "Moving target detection based on improved Gaussian mixture background subtraction in video images," *IEEE Access*, vol. 7, pp. 152612-152623, 2019.
- [13] L. L. Lin, and N. R. Chen, "Moving objects detection based on Gaussian mixture model and saliency map," *Applied Mechanics and Materials*, vol. 63-64, pp. 350-354, 2011.
- [14] Z. Zivkovic and F. V. D. Heijden, "Efficient adaptive density estimation per image pixel for the task of background subtraction," *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773-780, 2006.

- [15] S. Varadarajan, P. Miller and H. Zhou, "Spatial mixture of Gaussians for dynamic background modeling," In Proceedings of the 2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 63-68, 2013.
- [16] A. Cioppa, M. V. Droogenbroeck, and M. Braham, "Real-time semantic background subtraction," in Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Oct. 2020, pp. 3214–3218.
- [17] Ş. Işık, K. Özkan, and Ö. N. Gerek, "CVABS: moving object segmentation with a common vector approach for videos," IET Computer Vision, vol. 13, no. 8, pp. 719–729, Dec. 2019.
- [18] Y. Zhang, C. Zhao, J. He, and A. Chen, "Vehicles detection in complex urban traffic scenes using a nonparametric approach with confidence measurement," in 2015 International Conference and Workshop on Computing and Communication (IEMCON), pp. 1–7, Oct. 2017.
- [19] K..A. B. Lima, K. R. T. Aires, and F. W. P. D. Reis, "Adaptive method for segmentation of vehicles through local threshold in the Gaussian mixture model," Brazilian Conference on Intelligent Systems (BRACIS), pp. 204-209. 2015.
- [20] S. Agrawal and P. Natu, "An improved gaussian mixture method-based background subtraction model for moving object detection in outdoor scene," In Journal of Electrical and Electronic Engineering, vol. 6, no. 2, pp. 40-45, 2021.
- [21] X. Luo, Y. Wang, B. Cai, and Z. Li, "Moving Object Detection in Traffic Surveillance Video: New MOD-AT Method Based on Adaptive Threshold. ISPRS International Journal of Geo-Information, vol. 10, no. 11, pp. 742, 2021.
- [22] Z. Chen and T. Ellis, "Self-adaptive Gaussian mixture model for urban traffic monitoring system," In 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), pp. 1769–1776, 2011.
- [23] K. Kalti and M. A. Mahjoub, "Image segmentation by Gaussian mixture models and modified FCM algorithm," Int. Arab J. Inf. Technol., vol. 11, no. 1, pp. 11-18, 2014.
- [24] H. Wei and W. Zheng, "Image Denoising Based on Improved Gaussian Mixture Model," Scientific Programming, 2021.
- [25] I. Martins, P. Carvalho, L. Corte-Real, and J. L. Alba-Castro, "BMOG: boosted Gaussian mixture model with controlled complexity for background subtraction," Pattern Analysis and Applications, vol. 21, no. 3, pp. 641–654, Aug. 2016.
- [26] Y. Su, "Target detection algorithm and data model optimization based on an improved Gaussian mixture model," Microprocessors and Microsystems, vol.81, pp. 103797, 2021.
- [27] P. Kaewtrakulpong and R. Bowden, "An improved adaptive background mixture model for real-time tracking with shadow detection," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01), pp. 133–144, Sept. 2002.
- [28] D. S. Lee, "Effective Gaussian mixture learning for video background subtraction," IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 5, pp. 827–832, Mar. 2005.
- [29] H. Wang and P. Miller., "Regularized online mixture of Gaussians for background subtraction," presented at the 2011 8th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS), pp. 249-254, Aug. 2011.
- [30] J. Fried, F. Lizarralde, and A. C. Leite., "Adaptive Image-based Visual Servoing with Time-varying Learning Rates for Uncertain Robot Manipulators," presented at the 2022 American Control Conference (ACC), pp. 3838-3843, Jun. 2022.
- [31] P. Toulis, T. Horel, and E. M. Airoldi, "The proximal Robbins-Monro method," Journal of the Royal Statistical Society Series B: Statistical Methodology, vol. 83, no. 11, pp. 188–212, Feb. 2021.
- [32] H. Robbins and S. Monro, "A stochastic approximation method," The Annals of Mathematical Statistics, vol. 22, no. 3, pp. 400–407, Sept. 1951.
- [33] N. A. Mohd Aris, S. S. Jamaian, and D. R. Sulistyningtum, "Vehicle Detection Based on Improved Gaussian Mixture Model for Different Weather Conditions," Journal of Advanced Research in Applied Sciences and Engineering Technology, pp. 160-170, 2024.
- [34] C. W. Liu, B. Andersson, and A. Skrondal, "A constrained Metropolis-Hastings Robbins-Monro algorithm for Q matrix estimation in DINA models," Psychometrika, vol. 85, no. 2, pp. 322–357, Jun. 2020.
- [35] H. A. Tehrani, A. Bakhshi, and T. T. Y. Yang, "Online jointly estimation of hysteretic structures using the combination of central difference Kalman filter and Robbins-Monro technique," Journal of Vibration and Control, vol. 27, no. 1-2, pp. 234–247, Jan. 2021.
- [36] C. M. Bishop, Pattern Recognition and Machine Learning. New York, USA: Springer Google Schola, 2006. Accessed: August 23, 2016. [Online]. Available: <https://link.springer.com/in/book/9780387310732>.
- [37] S. Ruder, An overview of gradient descent optimization algorithms. Accessed: January 19, 2016. [Online]. Available: arXiv preprint arXiv:1609.04747.
- [38] E. Aboutanios, "Estimation of the frequency and decay factor of a decaying exponential in noise," IEEE Transactions on Signal Processing, vol. 58, no. 2, pp. 501–509, Sept. 2009.
- [39] K. Deng, "Exponential decay of solutions of semilinear parabolic equations with nonlocal initial conditions," Journal of Mathematical Analysis and Applications, vol. 179, no. 2, pp. 630–637, Nov. 1993.
- [40] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," IEEE Transactions on Speech and Audio Processing, vol. 3, no. 1, pp. 72–83, Jan. 1995.
- [41] C. V. Stewart, "Robust parameter estimation in computer vision," SIAM Review, vol. 41, no. 3, pp. 513–537, 2005.
- [42] J. Warwick and M. C. Jones, "Choosing a robustness tuning parameter," Journal of Statistical Computation and Simulation, vol. 75, no. 7, pp. 581–588, Jul. 2005.
- [43] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," IEEE Signal Processing Magazine, vol. 29, no. 4, pp. 61–80, Jun. 2012.
- [44] S. Tadjudin and D. A. Landgrebe, "Robust parameter estimation for mixture model," IEEE Transactions on Geoscience and Remote Sensing, vol. 38, no. 1, pp. 439–445, Jun. 2000.
- [45] E. López-Rubio, M. A. Molina-Cabello, R. M. Luque-Baena, and E. Domínguez, "Foreground detection by competitive learning for varying input distributions," International Journal of Neural Systems, vol. 28, no. 5, pp. 1750056, Jun. 2018.
- [46] T. Schlogl, C. Beleznai, M. Winter, and H. Bischof, "Performance evaluation metrics for motion detection and tracking," in Proceedings of the 17th International Conference on Pattern Recognition, pp. 519–522, 2004.
- [47] K. Oksuz, B. C. Cam, E. Akbas and S. Kalkan, "Localization recall precision (LRP): A new performance metric for object detection," in Proceedings of the European Conference on Computer Vision (ECCV), pp. 504–519, 2018.
- [48] N. Goyette, P. M. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, "ChangeDetection.net: A new change detection benchmark dataset," in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, pp. 1–8, Jun. 2012.
- [49] N. Lazarevic-McManus, J. P. Renno, and G. A. Jones, "Performance evaluation in visual surveillance using the F-measure," in Proceedings of the 4th ACM International Workshop on Video Surveillance and Sensor Networks, pp. 45–52, Oct. 2006.

PSR: An Improvement of Lightweight Cryptography Algorithm for Data Security in Cloud Computing

Dr. P. Sri Ram Chandra^{1*}, Dr. Syamala Rao², Dr. Naresh K³, Dr. Ravisankar Malladi⁴

Computer Science and Engineering, Shri Vishnu Engineering College for Women,
Bhimavaram, West Godavari District, Andhra Pradesh, India¹

Information Technology, SRKR Engineering College, Bhimavaram, West Godavari District, Andhra Pradesh, India²
Department of Computer Science and Engineering,

TKR College of Engineering & Technology, Hyderabad, Telangana, India³

Department of CSE, Koneru Lakshmaiah Education Foundation,
Vaddeswaram, Guntur District, Andhra Pradesh-522302, India⁴

Abstract—Data security in cloud storage is a pressing concern as organizations increasingly rely on cloud computing services. Transitioning to cloud-based solutions underscores the need to safeguard sensitive information against data breaches and unauthorized access. Traditional cryptography algorithms are vulnerable to brute-force attacks and mathematical breakthroughs, necessitating large key sizes for security. Moreover, they lack resilience against emerging quantum computing threats, posing a significant risk to encryption. To tackle these issues, this study presents a novel lightweight cryptography algorithm named as PSR which is aimed at encryption so as to improve data security before storage in cloud systems. The proposed system converts 128 bit plaintext to cipher by employing techniques such as substitution, ASCII and hexadecimal conversions, block-wise transformations including Rail Fence, Grey Code, and XOR operations with random prime numbers. Notably, the proposed algorithm demonstrates superior performance with minimal runtime and memory usage, satisfying the avalanche effect criterion with a noteworthy efficacy in all executions and resistant to brute force attack.

Keywords—Cryptography; cloud security; PSR; encryption; decryption; avalanche effect

I. INTRODUCTION

Cloud computing plays a vital role in modern businesses by offering flexible and scalable solutions for storing and accessing data, facilitating innovation, and enhancing collaboration while reducing infrastructure costs and improving efficiency [1]. Ensuring data security in cloud computing is paramount, safeguarding sensitive information from unauthorized access and cyber threats. It fosters trust among users, promotes compliance with data privacy regulations, and mitigates the risks associated with data breaches. Robust security measures uphold the integrity and confidentiality of data, bolstering the reliability and credibility of cloud-based systems [2]. Traditional cryptography and lightweight cryptography represent two distinct approaches to securing data, each tailored to different needs and constraints. Traditional cryptography typically involves complex algorithms and protocols designed to provide high levels of security but may require significant computational resources and power consumption, making them less suitable for resource-constrained environments [3]. On the other hand, Lightweight cryptography is essential in cloud computing to

optimize performance and resource usage, ensuring efficient data processing and secure communication across distributed networks while minimizing computational overhead [4]. In cloud computing, although current lightweight cryptographic algorithms provide efficiency, the ongoing development of new ones is crucial. Continuous development of new lightweight cryptographic algorithms in cloud computing ensures staying proactive against emerging threats and optimizing performance as environments evolve, supporting stronger, more resilient systems [5].

Encrypting data before storing it in the cloud enhances security by ensuring that only authorized parties with the decryption key can access the data, thus protecting against unauthorized access and data breaches. Additionally, encryption helps organizations meet regulatory compliance requirements and safeguards data integrity during transmission and storage [6]. In this research, the authors made progress in developing a novel cryptographic algorithm named as PSR, with the objective of encrypting data prior to its storage in cloud systems.

The subsequent sections of this paper are structured as follows: Section II outlines fundamental security requirements in cloud computing and explore research on lightweight cryptographic systems. Section III provides a brief overview of the proposed algorithm. Section IV presents the performance and security analysis of the proposed algorithm. Lastly, Section V offers conclusions and outlines future prospects.

II. RELATED WORK

This section includes essential security needs in cloud computing along with research on lightweight cryptographic systems.

A. Security Requirements of Cloud Computing

Key security requirements in cloud computing, as outlined by NIST [7], include confidentiality, availability, integrity, authorization, authentication, accountability, and privacy.

Confidentiality involves restricting access to customer information to authorized individuals. Integrity ensures that information remains unaltered during processing or transmission, and that only authorized individuals can modify or delete it. Authentication verifies the identity of users accessing

*Corresponding Author.

data, typically through account security measures. Availability ensures that customer data and services are consistently accessible. Authorization controls access to data, allowing only authorized individuals to retrieve it [8-9].

B. Cloud Computing Security

Recent studies explore cloud computing security, focusing on cryptography. They analyze encryption algorithms like AES, IDEA, and DES, comparing symmetric and asymmetric methods. Parameters such as Block Size, Key Length, and Execution Time are evaluated for efficiency, especially in the cloud environment [10]. [11] conducted a study on major cloud service providers like Google (Google Drive) and Microsoft (Azure and OneDrive). They examined cryptographic algorithms commonly utilized in cloud computing, including modern cryptography, searchable encryption, homomorphic encryption, and attribute-based encryption (e.g., DES, 3DES, AES, RC6, and BLOWFISH). By combining multiple cryptographic techniques, they introduced a hybrid encryption approach to enhance cloud data security. Additionally, [12] compared IDAs, SHA-512, 3DES, and AES-256, focusing on on-premise data encryption and decryption.

C. Study of Lightweight Cryptography Systems

M. Usman *et al.*, [13] examined the Stable IoT (SIT) lightweight encryption algorithm, which utilizes a 64-bit block cipher and mandates data encryption with a 64-bit address. This approach incorporates elements of both the Feistel structure and a uniform substitution-permutation network. In study [14], a novel symmetric stream cipher, Ultramodern Encryption Standard (UES), is introduced for secure data transmission, utilizing prolific series numbers for key generation and binary/gray code operations for encryption and decryption. Sriram C P *et al.* introduced the Modular Encryption Algorithm (MEA) [15], a novel symmetric block cipher utilizing a trimodular matrix for key generation and employing matrix operations, permutations, and substitutions for encryption and decryption processes. Authors in study [16] present a new symmetric stream cipher called the "Random Prime Key (RPK)" Algorithm, with an evaluation of its resilience against differential cryptanalysis and other pertinent factors, aiming for equilibrium between simplicity and security. In study [17], the "RECTANGLE" cryptosystem is outlined, designed for a 64-bit block size with key lengths of either 80 or 128 bits, and it executes 25 rounds. In study [18], the paper discusses a lightweight encryption algorithm for IoT devices, featuring a 64-bit block cipher and an 80-bit key for data encryption. In study [19], a lightweight cryptosystem features a 64-bit block size and 128-bit key, executed over 32 rounds with XOR operations and rotations. Its goal is hardware deployment in ubiquitous devices like wireless sensors and RFID tags, aiming for AES-level chip size but with faster performance.

III. THE PROPOSED ALGORITHM

To enhance data security within cloud computing, the authors introduced a new lightweight cryptography algorithm named as PSR, aimed at encrypting data prior to storage in cloud systems. PSR encrypts 128-bit binary blocks using a 128-bit key through 10 encryption rounds, relying on mathematical functions for diffusion and confusion in each round. The system

employs techniques like substitution, ASCII and hexadecimal conversions, block-wise transformations including Rail Fence and Grey Code, and XOR operations with random prime numbers. Additionally, it assesses the avalanche effect by comparing differing bits between the original plaintext and resulting cipher text.

A. Encryption Process

To safeguard sensitive data, the proposed algorithm PSR employs an encryption process that involves a sequence of cryptographic techniques. This process converts plain text into cipher-text while ensuring confidentiality and integrity. Here's a summary of the steps involved in one round of encryption process:

Input: Plain text message

Substitution Box Transformation:

```
Define substitution_box mapping characters to substitutes
substituted_text = ""
for each character in input_text:
    substitute = substitution_box[character]
    append substitute to substituted_text
```

ASCII Conversion:

```
ascii_values = []
for each character in substituted_text:
    ascii_value = convert character to ASCII value
    append ascii_value to ascii_values
```

Hexadecimal Conversion:

```
hexadecimal_values = []
for each ascii_value in ascii_values:
    hexadecimal_value = convert ascii_value to hexadecimal
    append hexadecimal_value to hexadecimal_values
```

Hexadecimal to Binary Conversion:

```
binary_values = []
for each hexadecimal_value in hexadecimal_values:
    binary_value = convert hexadecimal_value to 8-bit binary
    append binary_value to binary_values
```

Block-wise Transformation:

```
cipher_text = ""
for each block in binary_values:
    rail_fence_1 = ""
    rail_fence_2 = ""
    for each bit in block:
        if position_of_bit is even:
            append bit to rail_fence_1
        else:
            append bit to rail_fence_2
    grey_code = generate_grey_code(block)
    left_shifted = left_shift(grey_code, 2)
    not_operation_result = apply_not_operation(left_shifted)
    new_substitution_box_result =
    apply_new_substitution_box(not_operation_result)
    prime_number = generate_random_prime(min,max)
    prime_binary = convert prime_number to binary
    xor_result = perform_xor(new_substitution_box_result,
    prime_binary)
    hexadecimal_result = convert xor_result to hexadecimal
    ascii_result = convert hexadecimal_result to ASCII
    character_result = convert ascii_result to character
    append character_result to cipher_text
```

Output: cipher_text

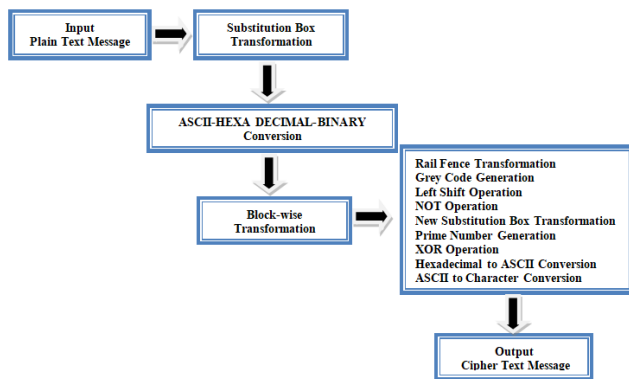


Fig. 1. PSR-Encryption process-flowchart.

The encryption process outlined in Fig. 1 begins by substituting characters in the plain text with predefined substitutes using a Substitution Box Transformation. ASCII Conversion converts substituted characters into ASCII values, followed by Hexadecimal Conversion for easier handling. Hexadecimal to Binary Conversion prepares data for block-wise transformation, where Rail Fence, Grey Code, Left Shift, and NOT operations are applied successively. Subsequent steps include a new Substitution Box Transformation, XOR operation with a key i.e., random prime number, and conversion back to cipher-text.

B. Key Exchange

The random prime keys used during the encryption process need to be exchanged with receiver so as to make use of them in decryption process. For safer exchange of keys between server and client, hybrid model that combines Diffie-Hellman (DH) and New-Hope (NH) is adopted [20].

C. Decryption Process

The decryption process that takes the cipher text as input and reverses the transformation steps to obtain the original plaintext message.

Input: Cipher text message

Character to ASCII Conversion:

```
ascii_values = []
for each character in cipher_text:
    ascii_value = convert character to ASCII value
    append ascii_value to ascii_values
```

Binary to Hexadecimal Conversion:

```
hexadecimal_values = []
for each 8-bit binary_value in ascii_values:
    hexadecimal_value = convert binary_value to hexadecimal
    append hexadecimal_value to hexadecimal_values
```

Block-wise Transformation:

```
original_binary_values = []
for each hexadecimal_value in hexadecimal_values:
    binary_value = convert hexadecimal_value to binary
    append binary_value to original_binary_values
```

Rail Fence and Grey Code Reconstruction:

```
reconstructed_binary_values = []
for each block in original_binary_values:
    not_operation_result = apply_not_operation(block)
    new_substitution_box_result =
```

```
reverse_apply_new_substitution_box(not_operation_result)
xor_result = perform_xor(new_substitution_box_result,
prime_binary)
reconstructed_binary_values.append(xor_result)
```

Binary to ASCII Conversion:

```
decrypted_ascii_values = []
for each binary_value in reconstructed_binary_values:
    decrypted_ascii_value = convert binary_value to ASCII
    append decrypted_ascii_value to decrypted_ascii_values
```

ASCII to Character Conversion:

```
original_text = ""
for each ascii_value in decrypted_ascii_values:
    character_result = convert ascii_value to character
    append character_result to original_text
```

Output: original_text

IV. RESULTS AND DISCUSSION

To validate the proposed algorithm, comparative analysis and a series of security experiments were conducted to gauge the effectiveness of the coding scheme, focusing on evaluation metrics including processing time, confusion and diffusion, avalanche effect.

A. Comparative Analysis of Proposed Encryption Algorithm

A comparison between the proposed method and current encryption techniques mentioned in Table I which reveals notable differences. While existing methods like DES, AES, Blowfish, and LED utilize various structures and operations, the proposed technique introduces a distinct approach. Unlike DES's limited block and key sizes or AES's variable configurations, the proposed technique offers a fixed 128-bit block and key size. Notably, PSR introduces unique mathematical operations like Rail-fence and Grey code conversions, enhancing its security. With a focus on security, PSR demonstrates a highly secure approach, surpassing the proven inadequacies of DES while matching or exceeding the security levels of AES, Blowfish, and LED as shown in Table I.

B. Processing Time

In cryptography, "processing time" is crucial, determining how long it takes to perform cryptographic tasks, impacting the speed of secure communication. Less processing time in cryptography algorithms is important for ensuring efficient and timely secure communication. The data presented in Table II represents the average encryption time obtained from five consecutive experimental runs. Fig. 2 illustrates that the proposed PSR algorithm consistently outperforms or matches the processing times of established algorithms like DES, AES, Blowfish, and LED across different file sizes. This underscores the superior efficiency and competitive advantage of the PSR algorithm in cryptographic operations.

C. Security Analysis

The proposed algorithm enhances the security of data before it gets stored onto the cloud by bolstering confidentiality and integrity while ensuring accessibility when needed. In terms of security, the PSR cryptographic algorithm can withstand well-known threats like weak key attacks, symmetric properties, related-key attacks, and differential and linear cryptanalysis [25, 26].

TABLE I. COMPARATIVE ANALYSIS OF PROPOSED ENCRYPTION ALGORITHM WITH EXISTING TECHNIQUES

Algorithm	DES[21]	AES[22]	Blowfish[23]	LED[24]	Proposed Algorithm-PSR
Structure	Feistel	Substitution-Permutation	Feistel	Feistel + SP	Feistel + SP
Block Size (bits)	64	128	64	64 or 128	128
Key Size (bits)	56	128, 192, 256	32–448	64 or 128	128
No. of Rounds	16	10, 12, 14	16	Variable	10
Key Space	256	2128, 2192, or 2256	232 – 2448	264 , 2128	2128
Mathematical Operations	Permutation, XOR, Shifting, Substitution	XOR, Mixing, Substitution, Shifting, Multiplication, Addition	XOR, Mixing, Substitution, Shifting	XOR, rotations, 2n mod addition, substitution	Substitution, Rail-fence, Binary and Grey code conversions, Shifting, NOT, XOR with Prime Number
S-P Structure	8 S-Box	1 S-Box	4 S-Boxes	4 S-Boxes	2 S-Boxes
Security Rate	Proven inadequate	Secure	Secure	Secure	Highly Secure

TABLE II. COMPARATIVE ANALYSIS OF PROPOSED ALGORITHM’S PROCESSING TIME AND EXISTING CRYPTOGRAPHY ALGORITHMS

Algorithm	DES[21]	AES[22]	Blowfish[23]	LED[24]	Proposed Algorithm-PSR
File Size (KB)	Processing time (seconds)				
28	0.0011	0.0014	0.0012	0.019	0.013
29	0.017	0.016	0.015	0.029	0.0152
210	0.028	0.029	0.031	0.0548	0.0352
213	0.29	0.29	0.24	0.62	0.45
214	0.945	0.805	1.06	1.45	0.789
215	1.91	1.74	2.01	2.53	1.65

Algorithm Processing Time Comparison

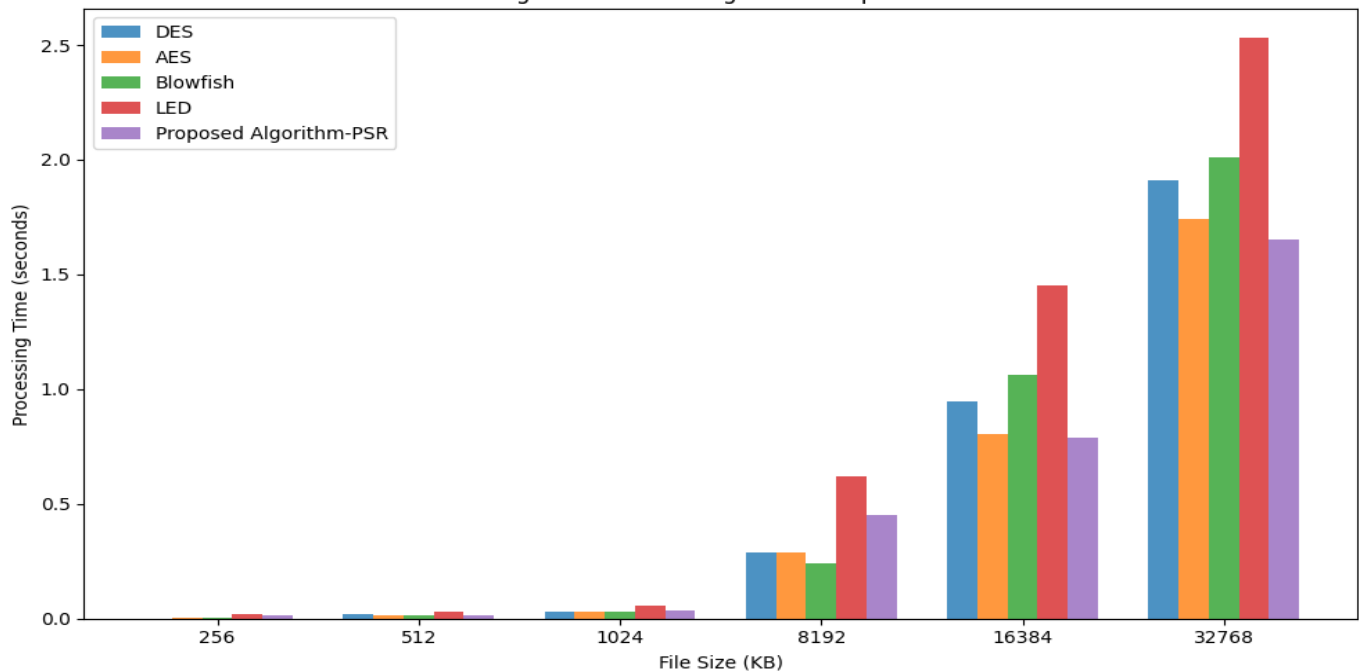


Fig. 2. Processing times of proposed algorithmm PSR cryptography algorithm and existing techniques.

$$\text{Percentage of Avalanche Effect} = \left(\frac{\text{Number of bits flipped in the cipher text}}{\text{Number of bits present in the cipher text}} \right) \times 100$$

D. Impact of Avalanche Effect-SPAC and SKAC

The algorithm must adhere to the Strict Plaintext Avalanche Criterion (SPAC), which implies that even a minor alteration in

the plaintext, while keeping the key constant, should lead to substantial changes in the resulting cipher text. Similarly, it should also meet the Strict Key Avalanche Criterion (SKAC), meaning that with the plaintext fixed, any slight modification in the key should produce significant variations in the generated cipher text [27]. The effectiveness of the PSR cryptography algorithm's security was assessed using SPAC and SKAC. Table III presents the outcomes of this assessment for a fixed plaintext ("Cryptography") across varying keys.

TABLE III. (A) SKAC ANALYSIS OF PSR CRYPTOGRAPHY ALGORITHM

Fixed Plaintext			
Test case	Cipher-text (128 bits)	Number of Bits Changed	Avalanche effect (%)
1	ùÂ¿ØbQDp"Š /	69	53.9
2	°s4ÂROpiÆøe	74	57.59
3	ï"WhÚ Fp*1	71	55.41
4	7hÂB¿ændD¿	78	60.62
5	bEÖÜ»Db@y	80	62.64
6	â2Û>¿-D2"èe	86	67.56
7	Ed¿Ø€ ðÖ,â>P³	109	85.83
8	³^ØÊµØÆhê	64	50.09
9	OðE\öC^ì •	92	72.54
10	7°©&:Â[ï"@@³	99	77.91
Average SKAC			64.40

(B): SPAC ANALYSIS OF PSR CRYPTOGRAPHY ALGORITHM

Fixed key			
Plaintext	Cipher-text (128 bits)	Number of Bits Changed	Avalanche effect (%)
Algorithm	Dâ_@p!¼Ö@s	70	51.66
Percentage	&@Â½/Âê	71	52.58
Avalanche	n*êEØð(77	57.22
Cloud Security	=oX÷Ë#Âèo©&,	92	69.10
Computing	QòkâbT¥	86	64.44
Brute Force Attack	lÂ[J,eZÖ*=&]pÂh´=2	83	62.08
Diffusion	gh ÂtÖjþ	79	58.98
Confidentiality	âê8Â • ø&¶,b7Äv¿	92	68.99
Integrity	!V@è#ð >5	95	71.66
Light weight	@9*#pÖÇ8v	86	64.56
Average SPAC			62.12

The results from the Tables III (A) and III (B) reveal that the percentages for SKAC (Strict Key Avalanche Criterion) and SPAC (Strict Plaintext Avalanche Criterion) are 64.40% and 62.12%, respectively. These thresholds demonstrate the algorithm's effectiveness in achieving a notable avalanche effect, which is critical for ensuring strong diffusion and enhanced security. Among the various test cases evaluated, the PSR

algorithm stands out for delivering the highest avalanche percentages, showcasing its superior performance. This makes it particularly well-suited for encrypting the texts before storing them onto cloud systems, where data confidentiality, integrity, and resilience against brute force attacks are well maintained. With its ability to ensure high levels of diffusion, the PSR algorithm is an excellent choice for protecting sensitive data with respect to cloud-based environments.

E. Confusion and Diffusion

Confusion and diffusion, concepts initially explored by Shannon [28], are fundamental to encryption, aiming to complicate the relationship between encrypted text and keys. Proposed encryption technique employs operations such as substitution, ASCII and hexadecimal conversions, block-wise transformations including Rail Fence and Grey Code, and XOR operations with random prime numbers. Altering a single letter in the original text impacts numerous sections of the encrypted text, while each encryption of identical text generates a varied outcome, enhancing both complexity and security. Consequently, our approach seamlessly integrates the fundamental concepts of confusion and diffusion.

F. Resistant to Brute Force Attack

A brute force attack exhaustively tests all potential combinations to compromise encryption keys. The PSR cryptography algorithm employs a 128-bit binary key, leading to 2¹²⁸ possible key combinations, guaranteeing unique keys for each encryption process.

V. CONCLUSION AND FUTURE SCOPE

Ensuring data security prior to its storage in the cloud has become increasingly crucial. Despite the existence of numerous cryptography algorithms aimed at bolstering data protection, there remains a significant demand for innovative approaches. This paper presents a novel cryptography algorithm, denoted as PSR, which operates on 128-bit plaintext using a 128-bit key to produce 128-bit cipher text. It draws inspiration from the architectural models of Fiestal and SP. The encryption method under consideration utilizes various operations, including substitution, ASCII and hexadecimal conversions, block-level transformations such as Rail Fence and Grey Code, as well as XOR operations involving random prime numbers. Despite boundaries like fixed key length and computational overhead, the PSR algorithm ensures high security with innovative techniques such as prime-based XOR, Rail Fence, and Grey Code. The experimental findings presented in Table I clearly indicate that the PSR algorithm, as proposed, offers a high level of security. Furthermore, Table II illustrates shorter processing times compared to current algorithms, affirming its applicability even in resource-limited environments. Table III data shows PSR cryptography excels in SKAC and SPAC, averaging 64.40% and 62.12%. The PSR algorithm's responsiveness to input variations increases output randomness, a desirable trait in cryptographic algorithms, thwarting attackers' predictions. Future work can focus on benchmarking PSR in real-world scenarios to evaluate its scalability and seamless integration with diverse cloud systems.

REFERENCES

- [1] Pazun, Brankica. (2018). Cloud Computing influence on modern business. Serbian Journal of Engineering Management. 3. 60-66. 10.5937/SJEM1802060P.
- [2] Soofi, Aized & Khan, M & Amin, Fazal-e. (2014). A Review on Data Security in Cloud Computing. International Journal of Computer Applications. 94. 975-8887. 10.5120/16338-5625.
- [3] Tankard, C. Encryption as the cornerstone of big data security. Netw. Secur. 2017, 2017, 5-7.
- [4] K. Huang, X. Liu, S. Fu, D. Guo, M. Xu, A lightweight privacy-preserving CNN feature extraction framework for mobile sensing, IEEE Trans. Dependable Secure Comput. 18 (3) (2019) 1441-1455.
- [5] Thabit, Fursan & Alhomdy, Sharaf & Al-ahdal, Abdulrazzaq & Jagtap, Prof. (2021). A New Lightweight Cryptographic Algorithm for Enhancing Data Security In Cloud Computing. Global Transitions. 2. 10.1016/j.gltp.2021.01.013.
- [6] Belguith, Sana. (2015). Enhancing Data Security in Cloud Computing Using a Lightweight Cryptographic Algorithm", The Eleventh International Conference on Autonomic and Autonomous Systems.
- [7] P. Mell, T. Grance, The NIST definition of cloud computing - SP 800-145, NIST Spec. Publ. (2011), doi: 10.1136/emj.2010.096966 .
- [8] Menezes, A. J., van Oorschot, P. C., & Vanstone, S. A. (1996). "Handbook of Applied Cryptography." CRC press.
- [9] Stallings, W. (2017). "Cryptography and Network Security: Principles and Practice." Pearson.
- [10] D.S. Abd Elminaam , H.M.A. Kader , M.M. Hadhoud , Evaluating the performance of symmetric encryption algorithms, Int. J. Netw. Secur. (2010).
- [11] J.R.N. Sighom, P. Zhang, L. You, Security enhancement for datamigration in the cloud, Futur. Internet (2017), doi: 10.3390/fi9030023.
- [12] D.P. Timothy, A.K. Santra, A hybrid cryptography algorithm for cloud computing security, 2017 International conference on Microelectronic Devices, Circuits and Systems (ICMDCS), Vellore (2017) 1-5, doi: 10.1109/ICMDCS.2017.8211728 .
- [13] M. Usman, I. Ahmed, M. Imran, S. Khan, U. Ali, SIT: a lightweight encryption algorithm for secure internet of things, Int. J. Adv. Comput. Sci. Appl. (2017), doi: 10.14569/ijacsa.2017.080151.
- [14] P. Sri Ram Chandra, G.Venkateswara Rao,G.V.Swamy, 'Ultramodern Encryption Standard Cryptosystem using Prolic Series for Secure Data Transmission', International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137 Volume - 02, Issue - 11, November - 2017, PP - 29-35.
- [15] P.Sri Ram Chandra, Dr. G.Venkateswara Rao and Dr.G.V.Swamy, Modular Encryption Algorithm for Secure Data Transmission Int. J. Sc. Res. In Network Security and Communication ISSN: 2321-3256 Volume-6, Issue-1, February 2018.
- [16] U.Bhanu Prasanna and Dr.P.Sri Ram Chandra, Data Security using Efficient Cryptosystem, TEST Engineering and Management, ISSN: 0193-4120, 15355-15360, January-February2020, The Mattingley Publishing Co., Inc.
- [17] W. Zhang, Z. Bao, D. Lin, V. Rijmen, B. Yang, I. Verbauwhede, RECTANGLE: a bit-slice lightweight block cipher suitable for multiple platforms, Sci. China Inf. Sci. (2015), doi: 10.1007/s11432-015-5459-7.
- [18] A.H.A. Al-ahdal , G.A. Al-rummana , G.N. Shinde , N.K. Deshmukh, A Robust Lightweight Algorithm for Securing Data in Internet of Things Networks, sustainable Communication Networks and Application. Lecture Notes on Data Engineering and Communications Technologies, vol 55. Springer, (2021).
- [19] Z. Gong, S. Nikova, and Y.W. Law, "KLEIN: a new family of lightweight block ciphers," 2012, doi: 10.1007/978-3-642-25286-0_1.
- [20] Hussein, A.I. (2023). Hybrid: (NH-DH) a New Hope and Diffie-Hellman for Transmission Data in Cloud Environment. In: Swaroop, A., Kansal, V., Fortino, G., Hassanien, A.E. (eds) Proceedings of Fourth Doctoral Symposium on Computational Intelligence . DoSCI 2023. Lecture Notes in Networks and Systems, vol 726. Springer, Singapore. https://doi.org/10.1007/978-981-99-3716-5_66
- [21] M.A. Wright, The advanced encryption standard, Netw. Secur. (2001), [https://doi.org/10.1016/S1353-4858\(01\)01018-2](https://doi.org/10.1016/S1353-4858(01)01018-2).
- [22] A.U. Rahman, S.U. Miah, S. Azad, Advanced encryption standard, in: Practical Cryptography: Algorithms and Implementations Using Cpp, 2014.
- [23] M.N. Valmik, P.V.K. Kshirsagar, "Blowfish Algorithm," IOSR J. Comput. Eng. (2014), <https://doi.org/10.9790/0661-162108083>.
- [24] G. Bansod, N. Raval, N. Pisharoty, Implementation of a new lightweight encryption design for embedded security, IEEE Trans. Inf. Forensics Secur. (2015), <https://doi.org/10.1109/TIFS.2014.2365734>.
- [25] M. Usman, I. Ahmed, M. Imran, S. Khan, U. Ali, SIT: a lightweight encryption algorithm for secure internet of things, Int. J. Adv. Comput. Sci. Appl. (2017), doi: 10.14569/ijacsa.2017.080151.
- [26] A.H.A. Al-ahdal , G.A. Al-rummana , G.N. Shinde , N.K. Deshmukh , A Robust Lightweight Algorithm for Securing Data in Internet of Things Networks, sustain- able Communication Networks and Application. Lecture Notes on Data Engineering and Communications Technologies, vol 55. Springer, (2021).
- [27] Norman D. Jorstad.: Cryptographic Algorithm Metrics, January 1997.
- [28] Shannon, C. E. (1949). Communication theory of secrecy systems. Bell System Technical Journal, 28(4), 656-715.

AUTHORS' PROFILE



Dr. P. Sri Ram Chandra serves Shri Vishnu Engineering College for Women, bringing a wealth of academic and professional experience. He holds a B.Tech from Andhra University and both an M.Tech and a Ph.D. from GITAM University. Dr. PSR has made significant contributions to his field, including the publication and review of books. He has also delivered guest lectures under the AICTE-STTP program and evaluated doctoral theses. Dr. PSR has an extensive record in academic service, having reviewed numerous research articles. His teaching career spans over a decade, during which he achieved numerous technical and research certifications. In addition to his academic achievements, Dr. PSR has published many patents and research publications. His research interests include, cryptography and information security, Theory of Computations. Detailed Profile: <https://sites.google.com/view/dr-psr>



Dr. Syamala Rao P. secured his Ph.D degree from Acharya Nagarjuna University. He is a topper in academics and secured gold medal in his M.Tech. He has 20+ years of experience in Academics and he is currently working as an Associate Professor in Information Technology department of S.R.K.R Engineering College, Bhimavaram. His research areas are cryptography, Machine Learning, Artificial Intelligence and Mining.



Dr. NARESH K is an Assistant Professor in the Department of CSE at TKR College of Engineering and Technology, Autonomous, Hyderabad. He received his Ph.D. in Computer Science & Engineering from the NIILM UNIVERSITY, Haryana in 2023. M.Tech in Software Engineering, Jagruthi institute of Engineering and Technology, JNTU HYDERABAD in 2012. His research interests in cryptography, Machine Learning, Deep Learning and Computer Networks.



Dr. M. RaviSankar, working as an Associate Professor in Department of Computer Science and Engineering, K.L. deemed to be University, Guntur dist., Andhra Pradesh, India. He has 24 years of teaching experience. Dr. Ravi has received Excellence in Research Award and Best Senior Faculty Award. He has Published 5 Patents and he has published more than 20 articles in Scopus, SCI, WOS and International Journal., His areas of specializations are cryptography, Data Mining and Artificial Intelligence.

Optimizing Feature Selection in Intrusion Detection Systems Using a Genetic Algorithm with Stochastic Universal Sampling

RadhaRani Akula¹, GS Naveen Kumar²

Research Scholar, Malla Reddy University, Hyderabad, India¹

Associate Professor, Department of CSE (Data Science), Malla Reddy University, Hyderabad, India²

Abstract—The current study presents a hybrid framework integrating the Genetic optimization algorithm with Stochastic Universal Sampling (GA-SUS) for feature selection and Deep Q-Networks (DQN) for fine-tuning an ensemble of classifiers to enhance network intrusion detection. The proposed method enhances genetic algorithms with stochastic universal sampling (GA-SUS) combined with recursive feature elimination (RFE). An ensemble of machine learning methods which includes gradient boosting and XG boost as base learners and subsequently logistic regression as meta learner is developed. A deep Q-networks (DQN) is used to optimize the base algorithms XG boost and gradient boost. The suggested method attains an accuracy of 97.60% on the popular NSL-KDD dataset and proficiently detects several attack types, such as probe attacks and Denial of Service (DoS), while tackling the issue of class imbalance. The multi-objective optimization approach is evident in anomaly detection and enhances model generalization by diminishing susceptibility to fluctuations in training data. Nonetheless, the model's efficacy regarding infrequent attack types, such as User to Root (U2R), remains inadequate due to their sparse representation in the dataset.

Keywords—GA-SUS; anomaly detection; IDS; RFE; DQN

I. INTRODUCTION

An Intrusion Detection System (IDS) is a cybersecurity tool with the primary goal of monitoring and analysing network traffic or system activity for possible malicious behaviour and unauthorized access. IDS can identify attempts that can lead to potential intrusions, that is, whether it be network attacks, unauthorized access to systems, or any other abnormal statistics to detect that we are dealing with malware or other cyber threats [1]. IDS can identify attempts that can lead to potential intrusions, that is, whether it be network attacks, unauthorized access to systems, or any other abnormal statistics to detect that we are dealing with malware or other cyber threats [2]. Anomaly-based detection and signature-based detection are two methodologies employed by Intrusion Detection Systems to identify suspicious activities.

Integrating IDS with machine learning has remarkably improved the potency of IDS to locate cyber threads accurately [3]. However, these are insufficient to address the complex dynamic threats posed by cyber threats [4]. Machine learning mitigates these constraints by allowing Intrusion Detection Systems to learn from data, adapt to emerging threats, and

enhance detection precision over time. Machine learning improves intrusion detection systems by discerning the most pertinent features for spotting intrusions. Emerging issues are seen in the increased incidence of assaults and the advancements in technologies noticed in contemporary IDS systems. Additional recommendations may be required for machine learning approaches while processing extensive data and transitioning throughout networking environments [5]. Consequently, there is a growing apprehension regarding the development of a way to extract superior high-order features when the objective is situated amongst a sea of nonstationary traffic. This necessitates the improvement of the generality and efficiency of the IDS to bolster the network's defences against novel and unidentified attacks [6].

Feature selection (dimensionality reduction) is an essential step in machine learning which entails choosing the most pertinent and informative characteristics from a dataset to enhance model performance. Feature selection minimizes model complexity, boosts generalizability, and frequently improves both accuracy and interpretability of predictions by retaining only the important features [7]. Feature selection Improves predictive accuracy by concentrating on the most pertinent features. Feature selection diminishes the likelihood of overfitting by removing noise and redundant information. It also simplifies the model, resulting in faster training and inference. In addition to that, it will minimize the storage and memory requirements while minimizes the computational complexity. Fig. 1 showcases the importance of feature selection. There exist three kinds of feature selection strategies namely, wrapper models [9], filter methods [8] and embedded methods [10]. Filter-based approaches evaluate feature significance according to the statistical characteristics of the data. They are not related to any specific machine learning algorithm. Parallely, wrapper methods use a specific learning algorithm to evaluate the performance of feature subsets [11]. Embedded approaches conduct feature selection during the model training phase. In this study we approach the feature selection mechanism with the aid of a wrapper method. Genetic Algorithms (GA) [12] are an optimization method derived on the concepts of genetics and natural selection. The genetic algorithms can effectively navigate extensive feature spaces and discern optimal or near-optimal feature subsets, rendering them particularly appropriate for high-dimensional datasets. The GA optimization is used for selecting most relevant features in this work.

To this end, it helps the model reduce overfitting, and thereby the model performs well when tested on unseen data. However, feature selection proved to be helpful in eliminating noisy components, resulting in an improvement in the quality of the provided dataset. In other words, when feature selection is performed properly, one is left with models that are accurate, efficient, and understandable - all qualities that are critical in the quest for insights and trustworthy predictions.

Feature selection is the core of any IDS in which the discovery of discrete features that characterize communications taking place in a network and the capability to discern between anomalous and normal is achieved. Feature selection is essential in the creation of efficient IDS by pinpointing the most pertinent aspects from network traffic data. Because of the complexity and high dimensionality of standard IDS datasets as NSL-KDD, UNSW-NB15 and CICIDS, feature selection enhances analysis by increasing detection accuracy and processing efficiency. By concentrating on the most informative attributes, the IDS can efficiently discern between regular and malicious actions. Minimizing the number of features decreases the computational load, resulting in expedited model training and real-time detection.

The ultimate aim of the current research is to propose and enhance a new approach to enhancing an NDIS by a more refined feature selection and optimization process. The primary contributions of the study are listed below.

- Enhancement of genetic algorithms with stochastic universal sampling (GA-SUS) combined with recursive feature elimination (RFE).
- An ensemble of machine learning methods which includes gradient boosting and XG boost as base learners and logistic regression as meta learner is developed.
- A deep Q-networks (DQN) is used to optimize the base learners XG boost and gradient boost.

The remainder of this paper is structured as below: Section II gives the literature review; Section III proposes the methodology; Section IV gives results and discussion and finally Section V concludes the study.

II. LITERATURE REVIEW

A research by Bakir et al. in study [13] explored innovative ways to enhance IDS using ML, specifically focusing on IoT networks. Using a genetic algorithm for tuning of hyperparameter along with a new hybrid feature selection, the authors propose a substantial increase of IDS effectiveness with the means of security threat identification. The authors combined several approaches looking for the most representative feature subset for detection through a hybrid feature selection methodology. Among others, Mutual Information-based Feature Selection (MIFS) is one among the several ways in which feature selection is performed by MIFS by selecting features from the original set according to their mutual information with the target value while reducing redundancy. Five (Decision Tree, XGBoost, Bagging, Extra Tree, Random Forest) ML algorithms were trained with their existing hyperparameters. The XGBoost classifier elevated the

performance, reaching 99.98% F1 score and 99.98% detection accuracy. The Extra Tree algorithm had a good performance as well, detecting with an accuracy of 99.96%.

A study by Cheng et al. in study [14] developed a pioneering approach known as Detection-Rate-Emphasized Multi-objective Evolutionary Feature Selection (DR-MOFS). The selected features are important for reducing the complex data sets for better efficiency and accuracy of IDS, according to the study. The goal is to decrease the features considered, thereby simplifying the framework and increasing performance. The second main aim of the study highlights optimizing the detection rate, which must be achieved as it minimizes the number of missed attacks. Also it overcomes the limitations of the previous Feature selection approaches based on feature subset size and classification accuracy which often led to low detection rate. Experiments were conducted on well-known network intrusion detection datasets, including UNSW-NB15 and NSL-KDD, in order to validate the suggested method. The results show that DR-MOFS is better than previous methods in most of the measures of less features selected, more accuracy, and more detection rate.

A research work by Ren et al. in study [15] generated a model MAFSIDS that aims to reduce the complexity of the feature selection process by eliminating close to 80% of repeating features in comparison to the base feature set. The MAFSIDS adopts a multi-agent framework in which a large number of feature agents compete with each other. The model provides adaptability to the evolving nature of network attacks (i.e. network IDS becomes more effective against new attacks). MAFSIDS improves the typical feature selection search strategy by formulating the feature selection problem as the target of MAFSIDS implemented in a multi-agent reinforcement learning framework in which the number of features selections in a general case is an exponential 2^N which it can specify those features which make up unit subsets. Here, you will find our model implementation which consists of Deep Q-Learning (a form of deep reinforcement learning). This approach allows the model to learn optimal policies for attacking the environment, through the interactions and feedback given based on the actions taken. GCNs are used to obtain deep features by MAFSIDS. As a result, this approach can significantly improve the feature selection process by allowing the model to better capture complex relationship in the data. While MAFSIDS model did very well with 96.8% accuracy rate on the dataset.

Another work by Ren et al. For example, [16] uses RFE and DT classifiers to remove 80% of all features and finds the most useful subset of features to identify all network attacks, especially unknown attacks. This article is referring on RFE which is used to assign importance the attributes in the ordinal manner of their significance related to target variable (i.e. intrusion detection in the network). Typically, the algorithm removes the least relevant features iteratively from the data. The model is refitted to the features after each iteration. The data is re-coded by way of Mini-Batch processing making the data-set relevant to the DRL model which is helpful in deriving more profound associations between features so it enhances accuracy and efficiency. Using the CSE-CIC-IDS2018 dataset for testing, the model achieved an F1-score of 94.9% and

accuracy of 96.2%. This shows that it is pretty effective at detecting network intrusions.

A study by Thajeel et al. in study [17] proposed DQN-MAFS implements a dynamic feature selection framework that continuously assesses the relevance of features in real-time and updates them accordingly. It is very important for capturing the changes in the data and eliminating irrelevant features for detection. Each feature is treated as an individual agent within the Multi-Agent System framework. Each agent acts to include/exclude a feature with some determination. Its architecture is based on reinforcement learning, which uses a deep Q-learning approach to facilitate online updates. As new labeled data becomes available, agents are rewarded to understand how much to rely on their own features and update their selection accordingly. FARD-DFS is a reward allocation sub-model within the DQN-MAFS framework.

A research work by Kavitha et al. in study [18] introduces a Deep Learning Model and Filter-based Ensemble Feature Selection for Intrusion Detection in Cloud Computing Environment. This research utilizes two publicly available datasets, NSL-KDD and KDDCup-99, for gathering the intrusion data. In the FEFS, three kinds of feature extraction process are involved, which are filter, wrapper and embedded algorithms, and it is obtained from this process that those features are extracted which will help the DLM in the training process. DLM combines RNN with a process known as Tuning Dynamic Optimization (TDO) for its optimization of weighting parameters. The proposed technique acquired a sensitivity of 0.90% and a recall of 0.93%. In relativity, the conventional methods achieved lower recall rates of 0.83% (DNN), 0.88% (RNN), 0.91% (RNN-GA) for recall, and 0.81% (DNN), 0.85% (RNN) for sensitivity.

A study by Mananayaka and Chung in study [19] proposed an innovative methodology for Network Intrusion Detection Systems (NIDS) that integrates Two-Phased Hybrid Ensemble Machine Learning with Automated Feature Selection, employing various ML classifiers to proficiently identify and shortlist the most pertinent attributes for identifying both familiar and unfamiliar attacks, thereby tackling the challenges associated with high-dimensional network data. The framework utilizes an automated feature selection engine that discerns the most pertinent elements from high-dimensional network data. Utilizing four distinct machine learning classifiers, the system may concentrate on the most pertinent information for attack detection, hence improving the accuracy and efficiency of the detection process. The suggested framework exhibited a high detection rate (0.9431) and an exceedingly low false alarm rate (0.0005) in evaluations performed on both wired and wireless networks.

A study by Yin et al. in study [20] aimed to improve the multi-classification efficiency of IDS by the judicious pertinent features selection and the reduction of feature space dimensionality. The IGRF-RFE method integrates wrapper and filter techniques to improve feature importance selection. The initial phase employs Random Forest (RF) and Information Gain (IG) to eliminate less significant features, while the subsequent phase utilizes RFE to further optimize the attribute subset by discarding features which detrimentally affect model

performance. This hybrid methodology seeks to improve the precision of the MLP-based detection of intrusion model utilizing the dataset of UNSW-NB15 through the selection of a more pertinent feature collection. The feature selection procedure decreased the number of features from 42 to 23, hence eliminating redundant and less pertinent characteristics. The MLP model's accuracy increased to 84.24% from 82.25% following the use of the "IGRF-RFE" approach. The weighted F1 score improved to 82.85%, indicating enhanced overall model performance for precision and recall.

The primary goal of the research by Saheed et al. in study [21] is to precisely detect fraudulent activity in computer networks by employing an advanced bat optimization technique in conjunction with the distinctive characteristics of the number system (residue). The work seeks to successfully diminish the complexity of the feature space by integrating the residue number system with the bat algorithm, while preserving or enhancing detection accuracy. The Bat algorithm is efficient for feature selection, although it may exhibit prolonged training and testing durations. The integration of RNS mitigates this constraint by enhancing processing speed. The study additionally utilizes PCA for feature extraction, which further enhances the chosen features. PCA facilitates the transformation of selected features into a lower-dimensional space while maximizing variance retention. The PCA + NB + Bat-RNS algorithm attained an accuracy of 97.82%. The Bat-RNS+PCA+KNN model exhibited an enhanced detection accuracy of 99.15%. The integration of the Bat method with RNS and PCA markedly improves the efficiency of the KNN classifier in intrusion detection.

A study by Francis and Sheeja in study [22] created an Intrusion Detection Model utilizing Bagging and Deep Reinforcement Learning (DRL). The model derives features from pre-processed data via the Enhanced Principal Component Optimization approach in conjunction with the Self-Optimizing Seagull Algorithm. This strategy aids in identifying pertinent features that can improve the model's efficacy. The chosen features are utilized to train the Bagging-DRL Intrusion Detection model, which integrates Convolutional Neural Networks, Multi-Layer Perceptron, Optimized Recurrent Neural Networks. The model is refined utilizing the Self-Improved Seagull Optimization Algorithm to augment detection precision. The model acquired an accuracy of 98.3% on the current dataset and 96% on the CSE-CIC-IDS2018 dataset. The framework demonstrated exceptional specificity rates of 99% for the NSL-KDD dataset and 97.6% for the CSE-CIC-IDS2018 dataset, highlighting its proficiency in accurately identifying non-intrusive cases. The sensitivity rates were robust, registering at 95% for the dataset of NSL-KDD and 98.3% for the CSE-CIC-IDS2018 dataset, indicating the model's efficacy in accurately detecting genuine intrusions.

A research paper by Rabash et al. in study [23] aims to selectively and adaptively identify pertinent characteristics in response to data alterations, tackling the issues presented by feature drift and concept drift in Intrusion Detection Systems. The suggested method employs a multi-objective optimization strategy to equilibrate several criteria, including feature relevance and feature reduction, so assuring that the chosen features enhance the classification model's performance

effectively. The research aims to increase the efficacy and precision of the IDS by the implementation of an “Enhanced Dynamic Filter-Based Feature Selection” (EDFBFS) architecture. The method utilizes a dual-mode strategy to produce optimal dynamic feature selection outcomes. The best feature set length is dictated by either the median or mean of the identified solutions in the Pareto, facilitating improved adaptation to varying circumstances. The method functions via iterative cycles encompassing initialization, crossover, and mutation processes. Throughout these cycles, objective functions are assessed according to feature relevance and feature reduction, directing the selection process. The E-DFBFS architecture proficiently tackles the issues of concept drift, facilitating enhanced adaptability in dynamic settings. Table I summarizes the contribution of previous researchers.

TABLE I. BACKGROUND WORK ANALYSIS

Study	Dataset(s)	Feature Selection Technique	Models
[13]	CICIDS2017	Mutual Information-based Feature Selection using genetic algorithm	Bagging, Random Forest XGBoost, Extra Tree and Decision Tree
[14]	NSL-KDD, UNSW-NB15	Multi-objective evolutionary algorithm	CART Decision tree, Logistic Regression, Random Forest
[15]	CSE-CIC-IDS2018, NSL-KDD	multi-agent feature selection	GCN
[16]	CSE-CIC-IDS2018	DT+RFE for feature selection	deep reinforcement learning
[17]	Four benchmark XSS datasets, which are, D3-30, D1-66, D4-30 and D2-167. T	Multi-agent feature selection and Deep Q-network	Multiple classifiers
[18]	KDDCup-99, NSL-KDD	Filter, wrapper, and embedded algorithms are classified as filter-based ensemble feature selection.	DLM is the short of RNN along with TDO
[19]	Aegean Wi-Fi Intrusion Detection Dataset	Automatic feature selection include (AFS-SVM, AFS-RF, AFS-ANN, and AFS-DT)	Two-phased Hybrid Ensemble learning
[20]	UNSW-NB15	Information gain and random forest with recursive feature elimination (RFE)	MLP
[21]	NSLKDD network data.	Bat algorithm with Residue Number System	NB, KNN
[22]	NSL-KDD and CSE-CIC-IDS2018 databases	Seagull algorithm for the enhancement of Enriched Principal Component Optimization	DRL uses MLP, CNN, while O-RNN interacts optimally with the surroundings or environment.

III. METHODOLOGY

The primary goal of this work is to develop a hybrid, ML model for network intrusion detection, in terms of feature selection, dimensionality reduction, and ensemble machine learning. The ameliorative model includes genetic algorithm

(GA), recursive feature elimination (RFE), kernel linear discriminant analysis (KL), principal component analysis (PCA), deep Q-network (DQN optimization steps) and stacked ensemble learning about it. The subsequent sections define and explain each phase of the identified methodology sequentially starting from the data pre-processing phase right up to the phase dealing with the evaluation of the final model. In this part of the research, we present the architecture of the proposed model in Fig. 1.

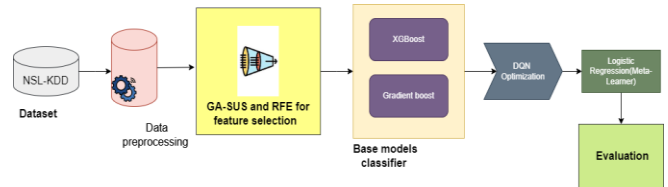


Fig. 1. Schematic architecture diagram of proposed system.

A. Preprocessing

The dataset used in this work has undergone a series of preprocessing methods to make it fit for the subsequent analysis. Firstly, the raw data is converted into a feature matrix with a corresponding vector label. The feature matrix contains a set of relative parameters that describe the network traffic, such as protocol type, packet size, and connection duration. The label vector comprises binary indicators that classify traffic into normal or incursion categories, facilitating supervised learning for ID.

To enhance the reliability and generalizability of the framework, the dataset was partitioned into testing and training subsets, a standard procedure for assessing model performance. The data division generally adheres to an 80:20 ratio, with 80% of the dataset designated for framework training and the residual 20% assigned for evaluating its predicted accuracy.

Prior to model training and feature selection, the data underwent supplementary preprocessing processes, encompassing demeaning and normalization of the features. Demeaning entails centering feature values around zero by subtracting the mean of each feature, whereas standardization adjusts the characteristics to achieve a standard deviation of one. These actions are essential for machine learning models, particularly when features display varying ranges or units of measurement. Standardization guarantees that all features contribute uniformly to the model, preventing those with more volatility or bigger magnitudes from overshadowing the learning process. This phase is crucial for models like as ensemble approaches and Support Vector Machines, which are sensitive to the relative scales of input features.

Standardizing the dataset before feature selection ensured a balanced representation of all features, enabling the feature selection method to discover the most pertinent qualities without bias. This thorough methodology strengthens the model's resilience, enabling it to more effectively identify trends in both legitimate and malicious network data.

B. Feature Selection using Genetic Algorithm (GA)

The process of feature selection involves reducing the number of attributes and identifying a subset of the original features. This technique is commonly utilised in data

preparation to uncover significant aspects that are often not known in advance and to eliminate superfluous or redundant features that have little bearing on classification tasks. In machine learning workflows, feature selection plays a pivotal role, particularly in enhancing the performance evaluation of classification models. The fundamental aim is to pinpoint the most crucial and informative features within the dataset, thereby improving accuracy.

Holland's genetic algorithm (GA) represents a computational optimisation methodology rooted in evolutionary biology principles. This technique operates in binary search spaces, managing a population of potential solutions. Each solution is encoded as a chromosome, comprising a finite sequence of binary digits. A fitness function assesses the viability of these solutions, with survival probability directly correlating to chromosomal fitness. The GA process commences with a randomly generated initial population, which then undergoes three primary mechanisms: selection, crossover, and mutation. The selection process identifies superior individuals for immediate progression to the next generation. Crossover involves the random exchange of chromosomal segments between two parent solutions to create offspring. Mutation introduces random alterations within individual chromosomes, contributing to genetic diversity.

This study employs Genetic Algorithms to remove inconsequential features. To achieve this objective, we designated chromosomes as a mask for attributes. For fitness evaluation, each individual in the population was assessed based on its ability to train a Random Forest classifier. If an individual selects at least one feature, the classifier is trained using these features, and its accuracy in the validation set determines the fitness score of the individual. If no features were selected, the fitness score was set to zero.

Selection was performed using stochastic universal sampling. First, the total fitness of the population was computed. The step size is then determined based on the total fitness and population size. Parents are chosen using a random start and pointers for a given size; the size is divided within the step size with the probability of high fitness being selected higher. Cross-over occurs whereby two selected parents are combined to form the offspring. A link was selected randomly and the child received some specific trait from both parents, or the first part was of one parent and the rest of the part was of other parent. Mutation is used in generating new offsprings by randomly setting bits to 0 or 1 adding new genetic feature to the population. A new population of the same size replaces the old one and this process a predefined number of generations or when some stopping criteria is fulfilled. Lastly the best from the final generation was chosen because it had the best fitness score out of all the individuals. This individual pertains to the best subset of features that are being searched sequentially by a genetic algorithm. The algorithm of GA along with mathematical formulae is given in Algorithm 1.

Algorithm 1: Genetic Algorithm for Feature Selection

Initialization:

Initialize the population $P = \{p_i \mid i = 1, 2, \dots, P\}$, where $p_i \in \{0, 1\}^N$ is a binary array representing a subset of features.

Fitness Evaluation:

For each individual $p_i \in P$, compute the fitness:

Let $F(p_i)$ be the set of selected features:

$$F(p_i) = \{j \mid p_i[j] = 1, j = 1, 2, \dots, N\} \quad (1)$$

If $F(p_i) \neq \emptyset$:

Then use the features of the dataset to train a random forest classifier

The accuracy $acc(p_i)$ of the classifier is calculated.

Otherwise, $acc(p_i) = 0$

Selection (Stochastic Universal Sampling):

Calculate the total fitness:

$$total_fitness = \sum_{i=1}^P acc(p_i) \quad (2)$$

Determine the step size:

$$step_size = \frac{total_fitness}{\frac{P}{2}} \quad (3)$$

Select parents:

Start point:

$$start_point = uniform(0, step_size) \quad (4)$$

Pointers: pointers = {start_point + k * step_size | k = 0, 1, ..., [P/2] - 1}

The indices based on cumulative fitness are selected.

Crossover:

For each pair of parents, p_i , and p_j :

Random crossover point $c = random(0, N - 1)$

Generate child:

$$c_k = (p_i[:c] \oplus p_j[:c]) \quad (5)$$

c_k inherits the first c bits from p_i and the remaining bits from p_j

Mutation:

For each child c_k :

For each bit $c_k[j]$:

$c_k[j] = 1 - c_k[j]$ with probability μ

New Generation:

The old population was replaced with the new generation of children.

This process continues for G generations or till we meet a certain criterion is met

Output:

Identify the best individual p^* from the final generation:

$$p^* = acc(p_i) \quad (6)$$

C. Recursive Feature Elimination (RFE)

RFE is a wrapper technique for feature removal. It removes repetitive and ineffective features that minimally affect the training error, while preserving strong and independent features to enhance the framework's generalization activity. It utilizes an sequential approach for feature importance, that is a variant of "backward feature elimination". This technique first develops the model utilizing the entire set of features and then prioritizes the features according to their importance. It subsequently removes the least significant feature, reconstructs the model, and recalculates the feature importance.

Following the feature subset derived by Genetic Algorithm (GA) optimization, Recursive Feature Elimination (RFE) was used to further enhance the selection process and ascertain a more ideal collection of features. RFE functions by iteratively removing the least important features based on the amount of

contribution they make towards the improvement of the model until we arrive at the number of features we need. Feature selection is addressed by using Random Forest algorithm as a model to predict the importance of the features. Subsequent process included turning off one feature after another from the bottom, beginning from the least contributing feature and retraining of the model. This process is continued until arrive at K best features only. These features were used in the subsequent features reduction. The following sections feature reduction and estimation steps.

D. Dimensionality Reduction

To address the curse of dimensionality and further reduce the feature space, two dimensionality reduction techniques are employed: Two methods identified are Principal Component Analysis (PCA) and Kernel Linear Discriminant Analysis (KLDA).

KLDA was used to transform the data onto a shorter feature dimension and also minimizing the interclass distance (normal – intrusion). Based on a kernel function, KLDA can model the nonlinear relationship of features, and then establish a better feature space.

$$Z_{KLDA} = W_{KLDA}^T X_{top} \quad (7)$$

where W_{KLDA} is the projection matrix obtained by maximizing the Fisher criterion.

After that, the features will be transformed by using the PCA in order to select only p principal components for comparison with the KLDA model. PCA removes projection directions determined to present high variability of the data and as such, most of the noise and redundant features.

$$Z_{PCA} = W_{PCA}^T Z_{KLDA} \quad (8)$$

Where W_{PCA} consists of eigenvectors corresponding to largest eigenvalues of the covariance matrix of Z_{KLDA} .

The final reduced dataset is denoted as Z_{final} .

E. Model Training and Stacked Ensemble Learning

1) *Base learners*: In order to construct a robust Intrusion Detection System (IDS), multiple base models were trained in the present study using a dataset that had been transformed into a lesser-dimensional vector space through the application of “Principal Component Analysis” (PCA). PCA, a prominent dimensionality reduction method, was utilized to identify the most critical characteristics while preserving the majority of the dataset's variation. XGBoost and Gradient Boosting Classifier were used as the main base models of the ensemble.

XGBoost is selected for handling large datasets and intricate pattern detection because of the gradient boosting framework upon which it is built. Additionally, GBC extends XGBoost, which iteratively provides better approximations to the model with fewer errors. These models complement each other to a great extent in the sense that they provide the benefit of handling numerous aspects of data complexity and drive up the predictive capability.

2) *Meta classifier*: A powerful binary classifier logistic regression takes the role of a meta-classifier. It is primarily deployed to merge the outcomes of the base, from which a final classification is generated. Logistic regression was again chosen because it is good at weighting the results of other models, and it calculates the best weights for each base model depending on the accuracy of the latter. The goal of this strategic integration is to increase the ability of the model to distinguish normal behaviour from non-normal or abusive behaviour.

3) Deep Q-Network (DQN) optimization

a) *Q-Learning setup*: Realising that the ensemble model could be enhanced, for hyperparameter tuning, we use a deep Q-network (DQN). Reinforcement learning is used in the form of a DQN, which helps in selecting the optimally-suited numerical for the hyperparameters for the best results. In this regime, the DQN influences the model in terms of the hyperparameters, and the response is a set of rewards derived from the model's evaluation results.

b) *Training*: When acquiring DQN, Q-values are updated when the amount of hyperparameters defined rises. The objective is to improve the reward function, which in the present case is the enhancement of the performance of the ensemble model. The same approach that is, following the above outlined feature selection scheme, benefits the DQN in a way that it is able to bring about ‘fine tuning’ of the hyperparameters to a level where classification differences of network activities are enhanced.

F. Proposed Model Algorithm

The combination of shortlisted features, the set of the training parameters, and performance metrics in a final model is preserved for future use. The documentation of the results comprises an evaluation of the proposed hybrid architecture for network intrusion identification. In this detailed record, the actual and the predicted markings are mentioned, which define how accuracy the model is beneficial for classifying the network threats; hence, comprehend how independent utilization of methodologies can be beneficial. The algorithm of the proposed model is given in Algorithm 2.

Algorithm 2: Proposed Machine Learning Framework for Network Intrusion Detection

Initialization

- $X, y \leftarrow$ Load data
- Hyperparameters \leftarrow Set parameters for GA, RFE, KLDA, PCA, DQN, and Stacking models

Feature Selection using Genetic Algorithm (GA)

- Initialize Population:
 - Population \leftarrow Random Initialization of N chromosomes
 - Evaluate Fitness:
 - For each chromosome $c_i \in$ Population:
 - Features \leftarrow Selected by c_i
 - Model \leftarrow Train RandomForest on Features
 - Fitness(c_i) \leftarrow Evaluate model accuracy
-

- Selection:
 - Selected Chromosomes ← Stochastic Universal Sampling (SUS) based on Fitness
- Crossover:
 - Offspring ← Apply Crossover on Selected Chromosomes
- Mutation:
 - Mutated Offspring ← Apply Mutation with rate pm
- Update Population:
 - Population ← Mutated Offspring
- Repeat:
 - Repeat steps for G generations or until convergence.
- Final Selection:
 - c_{best} ← Chromosome with highest Fitness

Recursive Feature Elimination (RFE)

- Feature Ranking:
 - Ranked Features ← RFE with RandomForest on Features selected by c_{best}
- Feature Selection:
 - Top Features ← Select k best features

Dimensionality Reduction

- Apply KLDA:
 - Z_{KLDA} ← KLDA on Top Features
- Apply PCA:
 - Z_{PCA} ← PCA on Z_{KLDA} reducing to p components

Model Training using Stacked Ensemble Learning

- Base Models:
 - Base Models ← Train models (XGBoost, GBC) on Z_{PCA}
- Meta-Classifer:
 - Meta-Model ← Train Logistic Regression on predictions of Base Models

Deep Q-Network (DQN) Optimization

- Q-Learning Setup:
 - States, Actions, Rewards, Q (s,a) ← Define for DQN
- Training:
 - $Q(s,a)$ ← Train DQN to optimize hyperparameters or thresholds Q (s,a)

Model Evaluation

- Prediction:
 - \hat{y} ← Predict using Meta-Model on test data
- Evaluation Metrics:
 - Recall, F1-Score, Accuracy, Precision, Confusion Matrix ← Evaluate on \hat{y}

Output Results

- Save (Features, Model Parameters, Metrics)
- Visualize Performance

IV. RESULTS

This section presents the results of the current study. Basis on the results attained, it is deduced that the intelligent hybrid model of GA-SUS feature selection and stacking ensemble

learning model with deep Q-learning neural network, which is proposed in the current research, is critical for using in network intrusion detection. NSL-KDD was used to benchmark the model with tests conducted to determine success rates, Precision, F1-score, recall, accuracy in differentiating between normal traffic, and anomalous traffic.

A. Dataset

The current dataset, NSL-KDD Dataset [24] is an improved and augmented version of the old KDD Cup 99 dataset, and is more suitable for IDS assessment. This approach eliminates certain inaccuracies in the initial data, for example, the presence of multiple records, which can introduce certain biases in the evaluation of an IDS. NSL-KDD consists of several types of records and probes: normal, DoS, R2L, U2R, and probes in the network traffic records. It is widely used to compare IDS effectiveness because it provides a reasonable distribution that is close to the real traffic distribution [16].

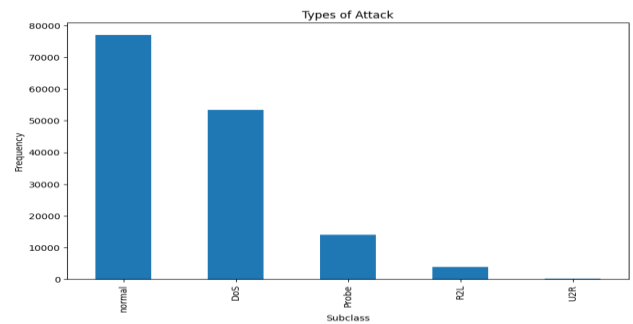


Fig. 2. Class distribution.

Fig. 2 illustrates the proportion of class labels within the dataset with the class label that appears most frequently. Such distribution forms can be skewed where some classes like ‘DoS’ and ‘normal’ are more frequent than classes like ‘U2R.’ Such distribution is important for model training and testing.

B. Evaluation Criteria

The assessment of the suggested model was performed using the following features: accuracy, confusion matrix, recall rate, precision rate, and F1 score. Accuracy gives a general measure of the developed model and checks correctness of the developed model. Precision, and recall measure to some extent how many of the positive instances are correctly classified and how few misclassifications in the form of false positives or false negatives are there. The F1 score is a metric that is in-between recall and precision. The confusion matrix allows estimating all the true, false, negations and positives that can be retrieved from the assessed model. It is the basis for calculating the said metrics. The formulae for the above metrics are given below.

$$Accuracy = \frac{TP1 + TN1}{TP1 + FP1 + TN1 + FN1}$$

$$Recall = \frac{TP1}{TP1 + FN1}$$

$$Precision = \frac{TP1}{TP1 + FP1}$$

$$F1 - Score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$$

Where FN is false negative, FP is false positive, TP is true positive, TN is true negative. These outcomes are shown in various kinds of diagrams and graphs for the objective of understanding and evaluating the performances of the models.

C. Classification Performance

In Table II, the classification report of a model with GA-SUS feature selection is illustrated. The model achieves an appreciable degree of accuracy: the overall accuracy is 0.9761. Outstanding performance for “DoS” (Denial of Service) category, shown that the model made a highly accurate detection of such kind of attacks. The “Probe” category is another category that gives a good result, but ‘DoS’ performance is slightly higher with good identification rate. Needless to say, weaker performance can be observed in the “R2L” category that has lower effectiveness for this kind of recognition. The “U2R” category can be said as very poor with all the parameters being nearly low. Since the presence of this category is negligible in the dataset, the detection capability shows a very poor result. As for the last “normal” group, the model correctly correlates their network activity with high performance indicators. In conclusion, the macro levels of performance at each class are low to moderate but at the same time the weighted levels indicate high competency of the model at identifying certain classes that are more dominant. Figure 3 provides confusion matrix of the model that used GA-SUS feature selection algorithm.

TABLE II. CLASSIFICATION REPORT OF MODEL USING GA-SUS FEATURE SELECTION

	precision	f1-score	recall	support
Probe	0.96	0.95	0.95	2749
DoS	0.99	0.99	0.99	10688
U2R	0.00	0.00	0.00	25
R2L	0.85	0.79	0.74	792
normal	0.98	0.98	0.98	15450
weighted average	0.97	0.98	0.98	29704
macro average	0.76	0.74	0.73	29704

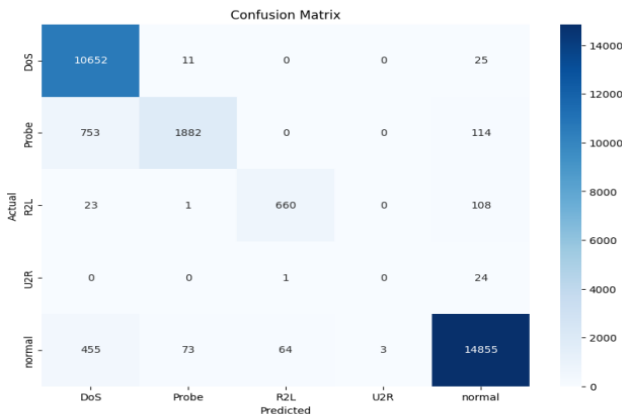


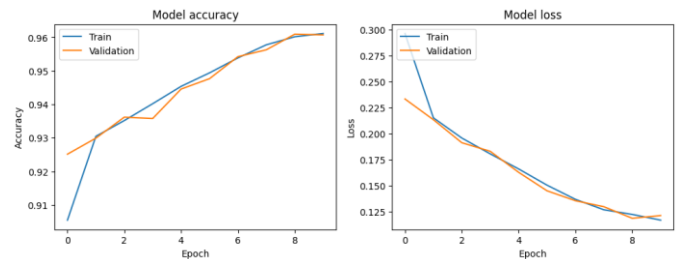
Fig. 3. Confusion Matrix of model using GA-SUS feature selection.

The suggested GA-SUS feature selection technique was contrasted with differential evolution-based algorithms that have the maturity extension feature selection proposed in [22]. When comparing the proposed GA-SUS with RFE ensemble learning approach to DE-ME, differences in performance and technique are evident.

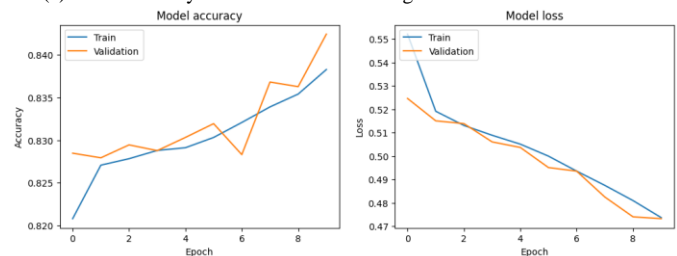
Classification Report of model using DE-ME Feature Selection is shown in Table III. Classification report shows overall high performance of the model in using feature selection from DE-ME is 94.43%. Once more, the accuracy of the model is extremely high when it comes to the detection of “DoS” and “normal” classes due to high coefficients of F1-score, recall, and precision, which equals to 0.90 and above. The macro average F1-score is calculated to be 0.72 and clearly shows the variation in the performance of the model across the classes Hence the weighted average F1-score of 0.94 reveals the complete performance of the framework; however, it somewhat biases towards the majority classes “DoS and “normal”. But this means that the model is more accurate when it comes to frequent attacks but not as effective when it comes to rare attacks.

TABLE III. CLASSIFICATION REPORT OF MODEL USING DE-ME FEATURE SELECTION

	Recall	Precision	F1-score	Support
U2R	0.00	0.00	0.00	25
DoS	1.00	0.90	0.94	10688
R2L	0.83	0.91	0.87	792
Probe	0.68	0.96	0.80	2749
normal	0.96	0.98	0.97	15450
macro avg	0.70	0.75	0.72	29704
weighted avg	0.94	0.95	0.94	29704



(a). The Accuracy and loss of models using GA-SUS feature selection.



(b). The Accuracy and loss of models using DE-ME feature selection.

Fig. 4. Accuracy and loss plot.

Fig. 4 (a) and Fig. 4 (b) illustrates Accuracy and loss plot for GA-SUS and DE-ME feature selection respectively. The accuracy and loss plots compare model performance using two

feature selection methods: DE-ME and GA-SUS. For both methods, the accuracy plot shows how well the models correctly classify data over training epochs, while the loss plot tracks the error reduction. Typically, a rising accuracy and a decreasing loss indicate good model training. Comparing the two, GA-SUS likely shows better stability with smoother curves and higher final accuracy, while DE-ME may have more fluctuations, suggesting GA-SUS's feature selection yields a more consistent and accurate model. The plots help visualize the effectiveness of each feature selection approach.

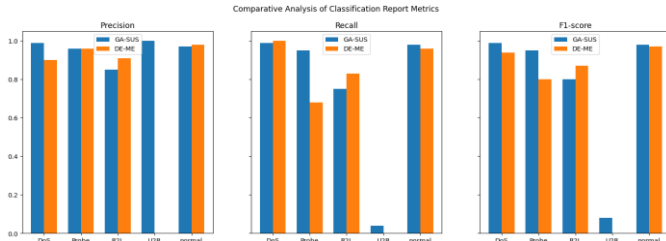


Fig. 5. Comparative performance analysis.

Fig. 5 presents the comparative performances classification algorithms. It visually compares the corresponding performance indices of two different models or features selection algorithms. This is likely to report, on the same screen, metrics such as Recall, Precision, F1-score, and even accuracy for each class, enabling a calibration. This comparison illustrates how various solutions affect the framework's ability in screening different kinds of attacks and normal traffic. In the current case and by the overlap of figure we are able to easily compare which of the GA-SUS feature selection method performs better in general and which one has a problem with certain classes. It offers information about the best and inferior aspects that can be used to strengthen the model.

D. Discussion

This study proposes a novel technique of GA-SUS with RFE for selecting the features for an IDS employing three benchmark datasets. In comparison with the existing approach, the current approach yielded results listed in Table IV.

Various studies on IDS datasets have applied different feature selection and machine learning algorithms. Our proposed model yielded decent results compared with those of other feature selection approaches in the literature.

TABLE IV. COMPARISON OF GA-SUS WITH RFE IN EXISTING STUDIES

Study	Feature Selection algorithm	Model	Accuracy achieved (%)
[25]	BukaGini(gini Importance)	Random forest classifier	99
[26]	Feature importance (RF)	RF	-
[27]	Condensed nearest neighbors (CNN)	CNN	95.54
		Radial basis function (RBF)	94.28
[20]	IGRF-RFE	MLP	-
[28]	GA in Map-Reduce	LR, SVM, RT, NB, ANN	90.45%
Proposed model	GA-SUS with RFE	Ensemble learning -DQN	97.61%

BukaGini, with a Random Forest classifier, obtained a high accuracy of about 99%. Other methods, such as Radial Basis Function (RBF) and convolutional neural network (CNN), yielded accuracies of 95.54% and 94.28%, respectively. The GA in the MapReduce approach combined with LR, RT, ANN, SVM and NB achieved 90.45% accuracy. Our model, utilising GA-SUS with Recursive Feature Elimination (RFE) and ensemble learning optimised by DQN, achieved a notable accuracy of 97.61%, displaying its robustness in intrusion detection.

Although the proposed model offers good results, certain limitations still exist. There appears to be no perfect dataset for studying invertible graphs; however, the current work employed the dataset called NSL-KDD, which has been used in most previous studies but may not portray real-life network traffic and emerging threats. Furthermore, the optimisation process used in DQN is quite efficient, but at the same time, it is costly and time consuming; hence, its applicability to large datasets or real-time data may be problematic. This study also presupposes that the selected features remain the best under various network conditions, which may not be true. Future work could consider extending the work to other types of datasets with larger and diverse groups of users, and also compare the model performance in real-time activities in dynamic network topologies.

V. CONCLUSION

The findings from this study highlight the feasibility of the proposed hybrid model of GA-SUS with RFE for feature selection and DQN for fine-tuning an ensemble learning model of classifiers for network intrusion detection. It reaches an accuracy of 97.60% on the NSL-KDD dataset and is capable of detecting different kinds of attacks, such as revival of DoS and probe attacks, as it solves the problem of class imbalance. The proposed multi-objective optimization harnessing stochastic universal sampling with a Genetic Algorithm for selection and Deep Q-Networks thus contributes to the design of new approaches for improving the generalization of the model by reducing its sensitivity to changes in the training data. As a result, the development of the study has limitations evident as follows; this kind of attack is very rare, but because it is present in the dataset very few times, the performance for such types like U2R remains below par. Future work may investigate better detection rates for these minority classes by investigating better data augmentation techniques or by using enriched deep neural networks. Furthermore, the model could be tested on other datasets as well as real-time environments, and such aspects could also be further explored. Extending this approach to address dynamic cyber threats or using it for more general and larger sets would further improve the approach to help with network security use cases.

REFERENCES

- [1] A. Thakkar and R. Lohiya, "A survey on intrusion detection system: feature selection, model, performance measures, application perspective, challenges, and future research directions," *Artif. Intell. Rev.*, vol. 55, no. 1, pp. 453–563, 2022.
- [2] S. Hajj, R. El Sibai, J. Bou Abdo, J. Demerjian, A. Makhoul, and C. Guyeux, "Anomaly - based intrusion detection systems: The requirements, methods, measurements, and datasets," *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 4, p. e4240, 2021.

- [3] F. Sharif, "The Role of Ensemble Learning in Strengthening Intrusion Detection Systems: A Machine Learning Perspective," 2024.
- [4] S. Ali, S. U. Rehman, A. Imran, G. Adeem, Z. Iqbal, and K.-I. Kim, "Comparative evaluation of ai-based techniques for zero-day attacks detection," *Electronics*, vol. 11, no. 23, p. 3934, 2022.
- [5] M. Di Mauro, G. Galatro, G. Fortino, and A. Liotta, "Supervised feature selection techniques in network intrusion detection: A critical review," *Eng. Appl. Artif. Intell.*, vol. 101, p. 104216, 2021.
- [6] S. Das et al., "Network intrusion detection and comparative analysis using ensemble machine learning and feature selection," *IEEE Trans. Netw. Serv. Manag.*, vol. 19, no. 4, pp. 4821–4833, 2021.
- [7] H. Liu et al., "Evolving feature selection," *IEEE Intell. Syst.*, vol. 20, no. 6, pp. 64–76, 2005.
- [8] N. Sánchez-Maróño, A. Alonso-Betanzos, and M. Tombilla-Sanromán, "Filter methods for feature selection—a comparative study," in *International Conference on Intelligent Data Engineering and Automated Learning, 2007*, pp. 178–187.
- [9] N. El Aboudi and L. Benhlina, "Review on wrapper feature selection approaches," in *2016 international conference on engineering & MIS (ICEMIS), 2016*, pp. 1–5.
- [10] H. Liu, M. Zhou, and Q. Liu, "An embedded feature selection method for imbalanced data classification," *IEEE/CAA J. Autom. Sin.*, vol. 6, no. 3, pp. 703–715, 2019.
- [11] Y. B. Wah, N. Ibrahim, H. A. Hamid, S. Abdul-Rahman, and S. Fong, "Feature selection methods: Case of filter and wrapper approaches for maximising classification accuracy," *Pertanika J. Sci. Technol.*, vol. 26, no. 1, 2018.
- [12] M. Zivkovic et al., "Hybrid genetic algorithm and machine learning method for covid-19 cases prediction," in *Proceedings of international conference on sustainable expert systems: ICSES 2020, 2021*, pp. 169–184.
- [13] H. Bakır and Ö. Ceviz, "Empirical enhancement of intrusion detection systems: a comprehensive approach with genetic algorithm-based hyperparameter tuning and hybrid feature selection," *Arab. J. Sci. Eng.*, pp. 1–19, 2024.
- [14] Z.-H. Cheng, H. Shang, and C. Qian, "Detection-Rate-Emphasized Multi-objective Evolutionary Feature Selection for Network Intrusion Detection," *arXiv Prepr. arXiv2406.09180*, 2024.
- [15] K. Ren, Y. Zeng, Y. Zhong, B. Sheng, and Y. Zhang, "MAFSIDS: a reinforcement learning-based intrusion detection model for multi-agent feature selection networks," *J. Big Data*, vol. 10, no. 1, p. 137, 2023.
- [16] K. Ren, Y. Zeng, Z. Cao, and Y. Zhang, "ID-RDRL: a deep reinforcement learning-based feature selection intrusion detection model," *Sci. Rep.*, vol. 12, no. 1, p. 15370, 2022.
- [17] I. K. Thajeel, K. Samsudin, S. J. Hashim, and F. Hashim, "Dynamic feature selection model for adaptive cross site scripting attack detection using developed multi-agent deep Q learning model," *J. King Saud Univ. Inf. Sci.*, vol. 35, no. 6, p. 101490, 2023.
- [18] C. Kavitha, T. R. Gadekallu, N. K. B. P. Kavin, and W.-C. Lai, "Filter-based ensemble feature selection and deep learning model for intrusion detection in cloud computing," *Electronics*, vol. 12, no. 3, p. 556, 2023.
- [19] A. K. Mananayaka and S. S. Chung, "Network intrusion detection with two-phased hybrid ensemble learning and automatic feature selection," *IEEE Access*, vol. 11, pp. 45154–45167, 2023.
- [20] Y. Yin et al., "IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset," *J. Big Data*, vol. 10, no. 1, p. 15, 2023.
- [21] Y. K. Saheed, T. O. Kehinde, M. Ayobami Raji, and U. A. Baba, "Feature selection in intrusion detection systems: a new hybrid fusion of Bat algorithm and Residue Number System," *J. Inf. Telecommun.*, vol. 8, no. 2, pp. 189–207, 2024.
- [22] E. Geo Francis and S. Sheeja, "Enhanced intrusion detection in wireless sensor networks using deep reinforcement learning with improved feature extraction and selection."
- [23] A. J. Rabash, M. Z. A. Nazri, A. Shapii, and M. K. Hasan, "Non-Dominated Sorting Genetic Algorithm based Dynamic Feature Selection for Intrusion Detection System," *IEEE Access*, 2023.
- [24] S. Mohanty and M. Agarwal, "Recursive Feature Selection and Intrusion Classification in NSL-KDD Dataset Using Multiple Machine Learning Methods," in *2nd International Conference on Computing, Communication, and Learning, CoCoLe 2023, 2024*, pp. 3–14.
- [25] M. A. Bouke, A. Abdullah, K. Cengiz, and S. Akleylek, "Application of BukaGini algorithm for enhanced feature interaction analysis in intrusion detection systems," *PeerJ Comput. Sci.*, vol. 10, p. e2043, 2024. DOI:10.7717/peerj-cs.2043
- [26] N. M. Khan, N. Madhav C, A. Negi, and I. S. Thaseen, "Analysis on improving the performance of machine learning models using feature selection technique," in *Intelligent Systems Design and Applications: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018) held in Vellore, India, December 6-8, 2018, Volume 2, 2020*, pp. 69–77.
- [27] F. Z. Belgrana, N. Benamrane, M. A. Hamaida, A. M. Chaabani, and A. Taleb-Ahmed, "Network intrusion detection system using neural network and condensed nearest neighbors with selection of NSL-KDD influencing features," in *2020 IEEE International Conference on Internet of Things and Intelligence System (IoT&IS), 2021*, pp. 23–29.
- [28] D. Mehanović, D. Kečo, J. Kevrić, S. Jukić, A. Miljković, and Z. Mašetić, "Feature selection using cloud-based parallel genetic algorithm for intrusion detection data classification," *Neural Comput. Appl.*, vol. 33, pp. 11861–11873, 2021

Optimizing Route Planning for Autonomous Electric Vehicles Using the D-Star Lite Algorithm

Bhakti Yudho Suprpto, Suci Dwijayanti, Desi Windisari, Gatot Aria Pratama

Department of Electrical Engineering, Universitas Sriwijaya, Inderalaya, South of Sumatera, Indonesia

Abstract—Every vehicle, including autonomous vehicles, requires a route to navigate its journey. Route planning is a critical aspect of autonomous vehicle operations, as these vehicles rely on guided paths or sequential steps to move effectively. Ensuring that the route is optimal is a key consideration. This study tests the D-Star Lite algorithm to determine the most efficient route. In simulation tests, the D-Star Lite algorithm was compared with the A-Star algorithm. The results showed that D-Star Lite outperformed A-Star, achieving an average distance reduction of 124 meters. Real-time testing involved finding a route from node 36 to node 0, resulting in a total distance of 803 meters. Additional tests focused on route replanning in real-time scenarios. For instance, the initial route passing through nodes 36 → 37 → 38 → 39 → 40 → 41 → 42 → 43 → 44 → 45 → 0 was adjusted to an alternative route: 36 → 37 → 38 → 46 → 26 → 11 → 2 → 4 → 1 → 0. Based on the results, the D-Star Lite algorithm proves effective in identifying the best route for autonomous electric vehicles while also enabling real-time route replanning.

Keywords—Autonomous vehicle; D-Star Lite; path planning; realtime; replanning route; optimal route

I. INTRODUCTION

Recent advancements in computing and communication technologies have significantly contributed to the development of autonomous vehicles. The emergence and evolution of these vehicles are the results of research in fields such as wireless communication technology, navigation, sensor technology, ad hoc networking, data acquisition and distribution, and data analysis [1] [2]. In addition to their ability to navigate autonomously to destinations, other critical factors must be considered in autonomous vehicles, such as the time required to reach the destination [3]. To plan movements efficiently, it is equally important to consider the routes the vehicle will follow [4].

Route planning is a critical aspect of robotics. Autonomous robots and vehicles require guidance paths or next steps to navigate effectively [5]. Defining a destination coordinate as a target ensures that the robot or autonomous vehicle can reach that destination via a predetermined route. This route must be the fastest to optimize efficiency and effectiveness while avoiding unnecessary steps [6]. This aspect is crucial to ensuring the efficiency and accuracy of the vehicle's movement [7]. Therefore, using a path planning method that works in dynamic environments is essential to handle obstacles that may change or be unpredictable [8]. Several methods can be used to determine the route, such as the Dijkstra algorithm, A-star, and D-Star Lite [9].

In research on the Dijkstra algorithm, it is explained that this algorithm is used for route planning in smart cars by implementing both the Dijkstra algorithm and the dynamic window approach [10]. This method was successfully applied to a self-developed smart car to avoid obstacles and reach a predetermined position. The study involved both simulation experiments and real-world testing, demonstrating the effectiveness and reliability of the Dijkstra algorithm and the implemented system.

In research on the A-star (A*) algorithm, improvements were made to the A-star-based path planning algorithm implemented in autonomous vehicles [11]. These improvements cover various aspects, such as the use of evaluation standards to measure performance, incorporating human guidance or global path planning to develop heuristic functions, leveraging key points around obstacles for more effective avoidance, and applying the variable-step-based A-star algorithm to reduce computation time [11].

However, both of these algorithms have their drawbacks. The Dijkstra algorithm is categorized as a greedy algorithm that can optimally find the shortest path solution [12]–[14], but it requires a longer search time. On the other hand, the A-star algorithm is a best-first search algorithm that can find the shortest path more quickly but does not always produce an optimal solution [15]. Nevertheless, A-star has an advantage over Dijkstra in its calculations. The A-star algorithm utilizes a heuristic distance [16] added to the straight-line path, resulting in a more efficient route. A-star is well-suited for situations where it is important to find a path quickly and efficiently in various environments [17]. Both the Dijkstra and A-star algorithms are only effective in solving the path search process in static environments (where conditions do not change) [18]. However, for an autonomous vehicle to adapt to unknown and potentially changing road conditions, a path planning algorithm that can be implemented in dynamic (changing) environments is needed. Therefore, the D-Star Lite algorithm will be developed.

The D-star Lite algorithm can address the efficiency issues of other algorithms when used in dynamic environments [18]. In previous research [10] and [11], path planning algorithms have been implemented using autonomous vehicles, some of which involved simulations. However, none of the studies using Dijkstra or A-star algorithms have been able to perform real-time replanning when obstacles change. In the study [19] that discusses the D-star Lite algorithm, the goal was to design a modified D-star Lite algorithm for global path planning in UAV-based (unmanned aerial vehicle) and mobile robots in large-scale disaster areas. This algorithm aims to address the

challenges of dynamic environments that are only partially known by providing shorter paths and faster execution times, ultimately improving the performance and efficiency of rescue robots in such situations. Although in study [19] explores the use of the D-Star Lite algorithm for path planning in dynamic environments with UAVs and mobile robots, no study has implemented the D-star Lite algorithm for path planning in autonomous vehicles. Therefore, this study uses the D-Star Lite algorithm to determine the fastest route for autonomous electric vehicles.

The contributions of this study are as follows: implementing the D-Star Lite algorithm to determine the fastest route for the autonomous vehicle, with real-time testing performed using a path on the Universitas Sriwijaya campus, which represents road conditions in Indonesia. Additionally, the study compares the D-Star Lite algorithm with other well-known path planning algorithms.

The paper is organized as follows: Section II explains path planning, the method is presented in Section III, Section IV discusses the results and findings, and finally, the paper is concluded in Section V.

II. PATH PLANNING

Path planning is a technique used to determine the best route for an autonomous electric vehicle to move from its current position to the desired destination while avoiding obstacles along the way [9]. Based on the environment in which it is applied, path planning can be performed in either static or dynamic environments.

In a static environment, obstacles have fixed positions and do not change location. In contrast, in a dynamic environment, obstacles may be partially known or entirely unknown, and their positions can change over time.

There are two types of path planning: global and local path planning.

1) *Global path planning*: Global path planning involves determining the route from the starting point to the destination within a larger environment. This requires extensive mapping and information about the robot's initial position and target destination. The focus is on finding the optimal route to reach the goal without considering the detailed surroundings near the robot. The global path planning process typically takes more time, as it involves analyzing the entire environment to identify the best route over a larger distance.

2) *Local path planning*: Local path planning focuses on determining the route around the robot's current position. Its primary objective is to avoid nearby obstacles and ensure the robot reaches its destination safely and efficiently. Local path planning is faster to execute because it focuses on a smaller area surrounding the robot.

Both approaches are essential for enabling autonomous vehicles to navigate complex environments effectively, combining broad-route optimization with immediate obstacle avoidance to ensure safety and efficiency.

B. Lifelong Planning A-star Algorithm (LPA*)

The Lifelong Planning A-Star (LPA*) algorithm is an enhancement of the A-Star algorithm. LPA* is an incremental version of A-Star, enabling it to adapt to changing environments by utilizing two key values: $g(s)$, which represents the cost accumulated so far to move from the current node to the start node (the formula for calculating $g(s)$ is provided in Eq. (1) [20]) and $rhs(s)$, which represents the best-known cost to reach a node from the start node (its formula is provided in Eq. (2) [20]).

By leveraging these two values, the LPA* algorithm efficiently recalculates paths as the environment changes, making it well-suited for dynamic scenarios.

$$g(s) = g(s') + d(s', s) \quad (1)$$

$$rhs(s) = \min_{s' \in neighbor(s)} (g(s') + d(s', s)) \quad (2)$$

where $g(s)$ is the cost to move from the start node to the current node, s is current node, s' is the predecessor node, and $d(s, s')$ is the cost of moving from the predecessor node to the current node.

If $g(s) = rhs(s)$, the node can be considered consistent. However, if the calculated node is inconsistent, it indicates a possible error in the calculation process.

In the LPA* algorithm, a priority queue is used to store nodes that are known and need to be evaluated or updated. Each node in the priority queue is assigned a key value, which determines the priority of the node. Nodes with the smallest key value are evaluated and updated first.

The function used to determine the key value of each node is provided in Eq. (3). This mechanism ensures that the algorithm efficiently processes nodes in the correct order, maintaining accuracy and minimizing computational overhead.

$$k(s) = \min(g(s), rhs(s)) + h(s) \quad (3)$$

where s is current node, $g(s)$ is g -value of the current node, $rhs(s)$ is rhs -value of current node and $h(s)$ is heuristic value of the current node.

C. D-star Lite (D* Lite) algorithm

The D-Star Lite algorithm, first developed by Sven Koenig and Maxim Likhachev in 2002, is a path planning algorithm capable of optimally finding a route between a start point and a goal point in environments that are known, partially known, or dynamic.

This algorithm operates on a data structure consisting of interconnected nodes. A node leading to the current position is called a predecessor node, while a node that will be traversed next is referred to as a successor node.

D-Star Lite is based on the Lifelong Planning A-Star (LPA*) algorithm, an incremental version of A-Star that adapts to changes in the map graph. However, unlike traditional approaches, D-Star Lite performs route planning starting from the goal node (finish) and works toward the start node. In this context, the $g(s)$ value represents the estimated cost from the current node to the goal node.

This reverse planning approach allows D-Star Lite to efficiently handle replanning when changes occur in the environment. The algorithm achieves this by maintaining an estimated cost for each traversed node, representing the distance to the goal node. This capability makes D-Star Lite particularly well-suited for dynamic and unpredictable environments.

D* Lite uses distance as a fundamental component because it is a path planning algorithm designed to find the shortest or least costly path between a start point and a goal. Here's why distance plays such a central role [7] [21] [22] [23]:

1) *Core purposes path planning:* The primary objective of D* Lite is to navigate an autonomous vehicle efficiently from a start point to a goal while avoiding obstacles. Distance or cost is the metric used to evaluate the optimality of the path. This ensures that the agent follows the shortest or least costly route, saving time, energy, or other resources.

2) *Adaptation to dynamic information:* In dynamic and partially known environments, the map can change due to new obstacles or updated information. D* Lite re-evaluates the distance (or cost) between nodes when changes occur, allowing the algorithm to efficiently update the path without recalculating everything from scratch. This incremental approach relies on comparing distances to ensure the agent can still reach the goal optimally.

3) *Grid representation and node expansion:* D* Lite often uses a grid or graph-based representation of the environment where nodes represent possible positions, and edges represent paths between these positions. The algorithm assigns a cost to each edge, typically based on physical distance or other factors like terrain difficulty. Calculating the shortest path through these nodes inherently involves summing distances or costs.

4) *Real-world relevance:* Distance is a straightforward and intuitive metric that directly translates to practical scenarios. Whether it's minimizing travel time, energy consumption, or fuel usage, distance serves as a universal measure of efficiency. For example, in rescue operations, D* Lite's reliance on distance ensures that the robot can reach victims or resources quickly.

D. Euclidean Distance

The Euclidean distance is a technique used to measure the distance between two points by considering the straight-line distance between them, not the angles. In Euclidean distance measurement, the calculation is conducted within a single plane and involves applying the Pythagorean theorem.

This method is commonly used to compute the distance between nodes and to determine heuristic values in the D-Star Lite algorithm. It achieves this by utilizing longitude and latitude values obtained from GPS sensors.

The formula for Euclidean distance is provided in the equation below, offering a straightforward way to calculate the straight-line distance between two points in a given space.

$$h = \sqrt{(x_{destination} - x_{start})^2 + (y_{destination} - y_{start})^2} \quad (4)$$

With x is the heuristic distance value, $x_{destination}$ is the longitude value of the target position, x_{start} is the longitude value of the starting position, $y_{destination}$ is the latitude value of the target position, and y_{start} is the latitude value of the starting position.

Eq. (4) above can be used to calculate the distance between two coordinate points, which will be applied in the D-star Lite algorithm. To obtain the distance in kilometers, Eq. (4) must be multiplied by the Earth's degree value, approximately 111.319888.

III. METHOD

A. Design System

In this study, the system design is presented in the form of a flowchart, as shown in Fig. 1 below. The flowchart illustrates the stages involved in determining the optimal route for an autonomous electric vehicle, as well as the steps taken to replan the route if obstacles are encountered.

In Fig. 1, the process begins with reading GPS data via ROS, followed by inputting the target node. The D-Star Lite algorithm determines the optimal route by identifying the direction of the next node based on the previous heading. The autonomous vehicle then starts moving toward the next node.

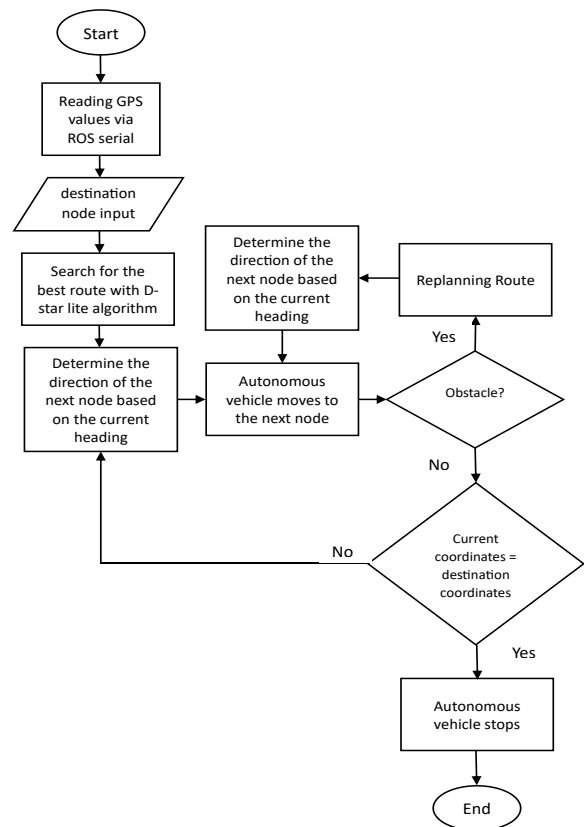


Fig. 1. Flowchart of path planning system design.

If obstacles are encountered along the route, the D-Star Lite algorithm will replan and determine a new direction for the next node. The autonomous vehicle will continue its movement. If no obstacles appear along the path, the system will check the vehicle's current position. If the current coordinates match the target coordinates, the autonomous vehicle will stop, indicating that it has reached the desired destination. The route search process using the D-Star Lite algorithm must be able to replan if obstacles are detected while the autonomous vehicle is moving toward the target point. The flow diagram for the designed software can be seen in Fig. 2.

In Fig. 2, it can be seen that the algorithm initially reads the coordinate values from the GPS system, which are transmitted via ROS serial communication. After obtaining the initial coordinates, the current coordinates are determined. The next step is to define the destination or target node. The D-Star Lite algorithm calculates the global path from the current node to the target node. The target node result is then sent to the controller via ROS serial communication.

The camera sensor provides image data that is sent to a computer for identification processing, which then sends input to the controller. If the camera detects an obstacle, the D-Star Lite algorithm adjusts the route and performs replanning, which is transmitted via ROS serial communication. However, if no obstacle is detected, the movement continues until the target node is reached.

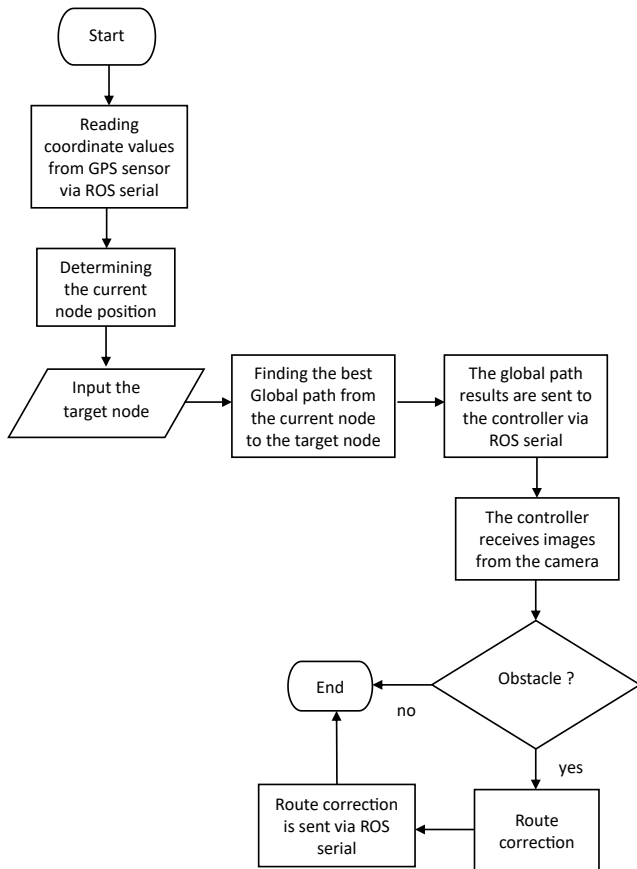


Fig. 2. The flowchart of design software.

B. Route Data

At this stage, longitude and latitude coordinate data are collected directly at each point designated as a node. A total of 47 longitude and latitude data points were obtained during this process, which will be used for testing purposes in both simulations and real-time scenarios. The nodes are labeled with numbers from 0 to 46, as shown in Table I.

TABLE I. NODE POINTS ON THE CAMPUS OF UNIVERSITAS SRIWIJAYA INDRALAYA

Node	Longitude	Latitude	Description
0	-3.21738259	104.64643749	Engineering Faculty
1	-3.21667979	104.64656550	North of the Faculty of Engineering T-junction
2	-3.21545079	104.64774530	The Faculty of Medicine Intersection
3	-3.21548959	104.64955100	The Southern T-Junction of the Rectorate
4	-3.21667550	104.64773890	The Eastern Intersection of the Library
5	-3.21666990	104.64954630	The Western Intersection of the Library
6	-3.21667029	104.65052800	Faculty of Social and Political Sciences Intersection
7	-3.21737769	104.65052250	South of Faculty of Social and Political Sciences Intersection
8	-3.21668260	104.65088070	T-Junction of Faculty of Social and Political Sciences
9	-3.21735050	104.65089470	South of the FISIP T-Junction
10	-3.21399860	104.65086760	The Northern Intersection of the Faculty of Law
11	-3.21385989	104.64773350	Auditorium intersection
12	-3.21820369	104.65055840	South intersection of Faculty of Economics
13	-3.21911079	104.65051810	Intersection of Faculty of Computer Science
14	-3.21950100	104.64873060	West intersection of Faculty of Agriculture
15	-3.21804735	104.64875234	Intersection behind the library
16	-3.21564319	104.64739540	Faculty of Medicine
17	-3.21670539	104.64872703	Library
18	-3.21391629	104.64873580	Landmark UNSRI
19	-3.21644240	104.65090500	Faculty of Social and Political Sciences
20	-3.21536880	104.65088300	Faculty of Law
21	-3.21795930	104.65054590	Faculty of Economics
22	-3.21949390	104.64932110	Faculty of Teacher Training and Education
23	-3.21855209	104.64639790	Faculty of Mathematics and Natural Sciences
24	-3.21950639	104.64806430	Faculty of Agriculture
25	-3.21911473	104.65089650	Faculty of Computer Science
26	-3.21397368	104.64540105	Faculty of Public Health
27	-3.21735675	104.64956288	South of node 5
28	-3.21805961	104.64956063	South of node 27
29	-3.21903399	104.64654099	South of Faculty of Mathematics and Natural Sciences
30	-3.21945477	104.64699562	South of node 29

31	-3.21843234	104.64683263	South of the canteen intersection
32	-3.21803595	104.64686478	Intersection of canteen
33	-3.21940914	104.65020614	West of Faculty of Teacher Training and Education
34	-3.21579602	104.65035667	North of node 6
35	-3.21394241	104.64953128	Rectorate
36	-3.21736794	104.64538617	Department of Electrical Engineering
37	-3.21730889	104.64475116	East intersection of Electrical Engineering Department
38	-3.21735422	104.64370186	T-junction of Faculty of Engineering
39	-3.21894491	104.64379309	T-junction of south node 38
40	-3.21884979	104.64427315	West of node 39
41	-3.21838310	104.64493115	West of node 40
42	-3.21834656	104.64501846	Behind of Department of Mechanical Engineering
43	-3.21793954	104.64534572	Behind of Department of Electrical Engineering
44	-3.21791539	104.64564074	East of node 45
45	-3.21821540	104.64647486	T-junction of Faculty of Mathematics and Natural Sciences
46	-3.21396715	104.64410885	T-junction of Faculty of Public Health

The mapping of these 46 nodes is shown in Fig. 3.

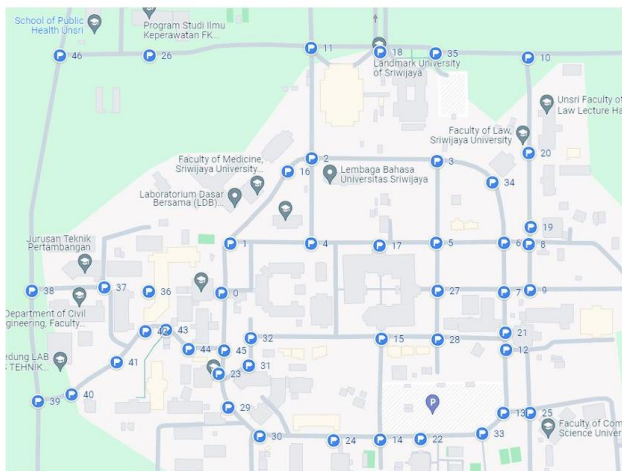


Fig. 3. The mapping routes on the Universitas Sriwijaya Indralaya campus.

In this study, the selected location is the road around the Indralaya campus of Sriwijaya University because the roads in this area have challenging characteristics, such as the absence of road markings, road barriers, and the surface condition of the roads which is not very smooth. The roads around the Indralaya campus reflect those in rural areas of South Sumatra Province in general. In terms of traffic density, it is not as congested as rural roads in Sumatra, but it is already quite busy due to the many students who use the roads by riding motorcycles, driving cars, or taking buses.

IV. RESULTS AND DISCUSSIONS

A. Path Planning Testing Through Simulation

In this testing, the path planning system is evaluated using the D-Star Lite algorithm to determine whether the developed system functions properly. A comparison will also be made between the route search results using the D-Star Lite algorithm and the A-Star algorithm from previous research. This experiment involves finding the best route across 10 different routes. In the first trial, the route search was tested from the Faculty of Engineering to the Faculty of Law. The results of this test are presented in Table II, and the traversed route is shown in Fig. 4.

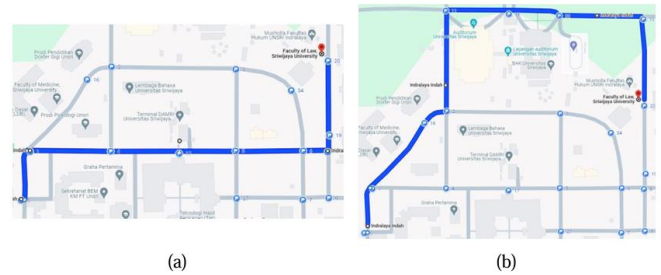


Fig. 4. Route from the faculty of engineering to the faculty of law: (a) D-Star Lite method, (b) A-Star method.

TABLE II. TESTING THE ROUTE FROM THE FACULTY OF ENGINEERING TO THE FACULTY OF LAW

Method	Nodes skipped	Total euclidean distance (m)	Distance based on google maps (m)
D-Star lite	0→1→4→17→5→6→8→19→20	704,9	702
A-Star	0→1→16→2→11→18→35→10→20	948,9	948
Distance difference (m)		244	246

In the second trial, the route search was tested from the Faculty of Engineering to the Faculty of Economics. The results of this test are presented in Table III, and the traversed route is shown in Fig. 5.

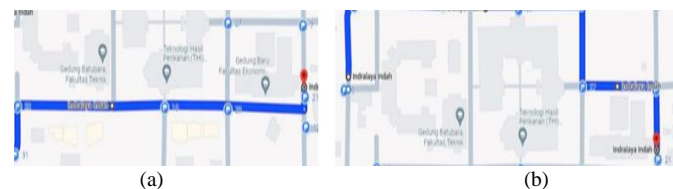


Fig. 5. Route from the Faculty of Engineering to the Faculty of Economics: (a) D-Star Lite Method, (b) A-Star Method

TABLE III. TESTING THE ROUTE FROM THE FACULTY OF ENGINEERING TO THE FACULTY OF ECONOMICS

Method	Nodes skipped	Total euclidean distance (m)	Distance based on google maps (m)
D-Star lite	0→45→23→31→32→15→28→21	633,6	633
A-Star	0→1→4→17→5→27→7→21	658,1	658,3
Distance difference (m)		24,5	25,3

In the third trial, the route search was tested from the Faculty of Engineering to the Rectorate. The results of this test are presented in Table IV, and the traversed route is shown in Fig. 6.

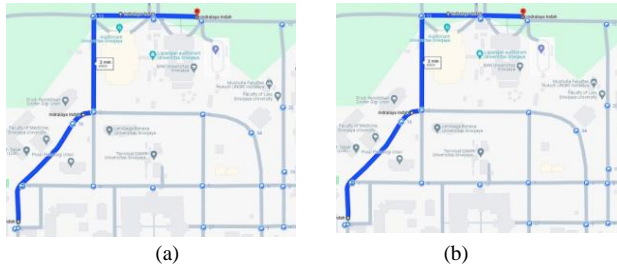


Fig. 6. Route from the faculty of engineering to the rectorate: (a) D-Star lite method, (b) A-Star method.

TABLE IV. TESTING THE ROUTE FROM THE FACULTY OF ENGINEERING TO THE FACULTY OF ECONOMICS

Method	Nodes skipped	Total euclidean distance (m)	Distance based on google maps (m)
D-Star lite	0→1→16→2→11→18→35	648	650,6
A-Star	0→1→16→2→11→18→35	648	650,6
Distance difference (m)		0	0

In the fourth trial, the route search was tested from the Faculty of Economics to the Faculty of Medicine. The results of this test are presented in Table V, and the traversed route can be seen in Fig. 7.

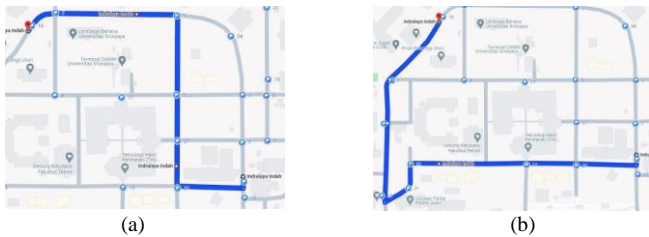


Fig. 7. Route from the faculty of economics to the faculty of medicine: (a) D-Star lite method, (b) A-Star method.

TABLE V. TESTING THE ROUTE FROM THE FACULTY OF ECONOMICS TO THE FACULTY OF MEDICINE

Method	Nodes skipped	Total euclidean distance (m)	Distance based on google maps (m)
D-Star lite	21→28→27→5→3→2→16	640,6	640,2
A-Star	21→28→15→32→31→23→45→0→1→16	860,6	858,6
Distance difference (m)		220	218,4

In the fifth trial, the route search was tested from the Faculty of Agriculture to the landmark. The results of this test are presented in Table VI, and the traversed route is shown in Fig. 8.

From the five trials conducted, the D-Star Lite algorithm shows a larger error in comparison to Google Maps readings than the A-Star algorithm. However, when comparing the routes taken and the best route searches, the D-Star Lite algorithm outperforms the A-Star algorithm. This is evident in the first, second, and fourth trials, with the largest difference being 244 meters in the second trial. This occurs because the A-Star algorithm prioritizes only the nodes leading directly to the destination as the best route, whereas the D-Star Lite algorithm evaluates each node in the dataset to determine the shortest path to the destination. Consequently, the D-Star Lite algorithm sometimes finds a more optimal route than the A-Star algorithm. Therefore, the D-Star Lite algorithm is a viable method for finding the best route.

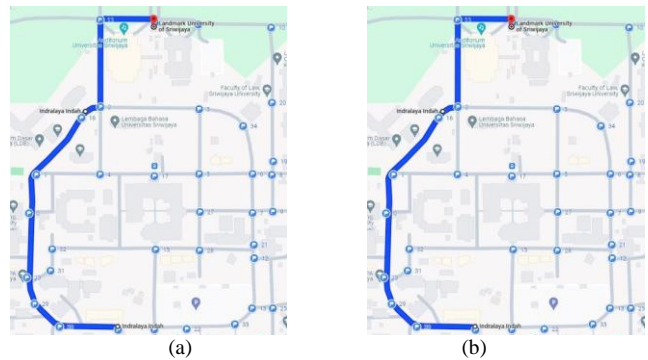


Fig. 8. Route from the faculty of agriculture to the landmark: (a) D-Star lite method, (b) A-Star method.

TABLE VI. TESTING THE ROUTE FROM THE FACULTY OF ECONOMICS TO THE FACULTY OF MEDICINE

Method	Nodes skipped	Total Euclidean distance (m)	Distance based on google maps (m)
D-Star lite	24→30→29→23→45→0→1→16→2→11→18	933,3	934,2
A-Star	24→30→29→23→45→0→1→16→2→11→18	933,3	934,2
Distance difference (m)		0	0

B. Route Replanning Testing via Simulation

In this simulation test, the replanning system using the D-Star Lite algorithm was tested to determine whether it could successfully perform route replanning when an obstacle appeared on the route. This experiment included five tests to evaluate whether the D-Star Lite algorithm's replanning system could be used in real-time conditions.

In the first trial, a route search was conducted from the Faculty of Engineering to the Faculty of Law. The best route identified passed through nodes 0 → 1 → 4 → 17 → 5 → 6 → 8 → 19 → 20. After establishing the route, node 17 was designated as an obstacle or closed, prompting the D-Star Lite algorithm's replanning system to search for the best alternative route avoiding the closed node. The resulting route passed through nodes 0 → 1 → 4 → 2 → 3 → 34 → 6 → 8 → 19 → 20, as shown in Fig. 9.

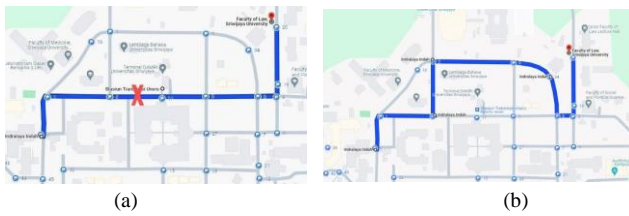


Fig. 9. Replanning route from the faculty of engineering to the faculty of law
(a) Before replanning (b) After replanning.

In the second trial, a route search was conducted from the Faculty of Economics to the Faculty of Medicine. The best route identified passed through nodes 21 → 28 → 27 → 5 → 3 → 2 → 16. After establishing the route, node 3 was designated as an obstacle or closed, prompting the D-Star Lite algorithm's replanning system to search for the best alternative route, avoiding the closed node. The resulting route passed through nodes 21 → 28 → 27 → 5 → 17 → 4 → 2 → 16, as shown in Fig. 10.

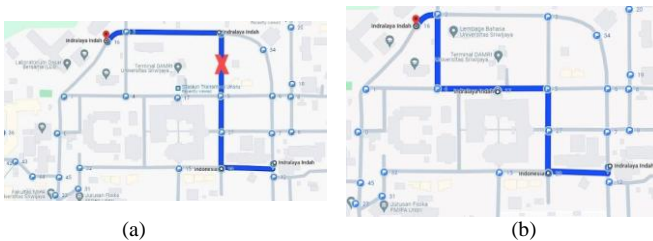


Fig. 10. Replanning route from the Faculty of Economics to the Faculty of Medicine (a) Before replanning (b) After replanning.

In the third trial, a route search was conducted from the Faculty of Mathematics and Natural Sciences to the Faculty of Economics. The best route identified passed through nodes 23 → 31 → 32 → 15 → 28 → 21. After establishing the route, node 28 was designated as an obstacle or closed, prompting the D-Star Lite algorithm's replanning system to search for the best alternative route, avoiding the closed node. The resulting route passed through nodes 23 → 31 → 32 → 15 → 14 → 22 → 33 → 13 → 12 → 21, as shown in Fig. 11.

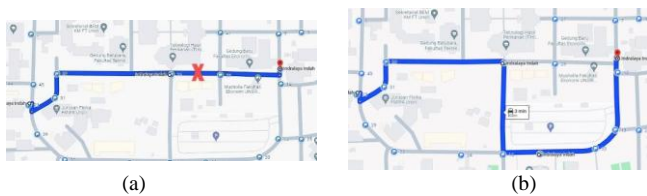


Fig. 11. Replanning route from the faculty of mathematics and natural sciences to the faculty of economics (a) Before replanning (b) After replanning.

In the fourth trial, a route search was conducted from the Faculty of Law to the Faculty of Agriculture. The best route identified passed through nodes 20 → 19 → 8 → 9 → 7 → 27 → 28 → 15 → 14 → 24. After the route was established, node 27 was designated as an obstacle or closed, prompting the D-Star Lite algorithm's replanning system to search for the best alternative route, avoiding the closed node 27. The resulting route passed through nodes 20 → 19 → 8 → 9 → 7 → 21 → 28 → 15 → 14 → 24, as shown in Fig. 12.

In the fifth trial, a route search was conducted from the Faculty of Public Health to the Faculty of Law. The best route identified passed through nodes 26 → 11 → 18 → 35 → 10 → 20. After establishing the route, node 18 was designated as an obstacle or closed, prompting the D-Star Lite algorithm's replanning system to search for the best alternative route, avoiding the closed node. The resulting route passed through nodes 26 → 11 → 2 → 3 → 34 → 6 → 8 → 19 → 20, as shown in Fig. 13.

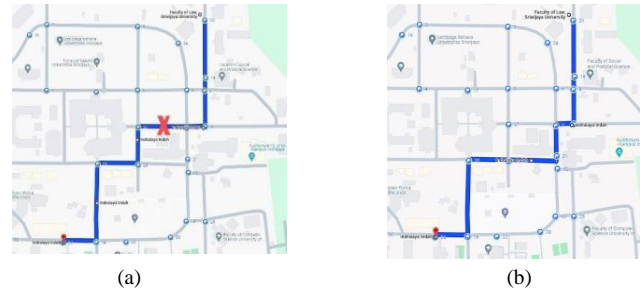


Fig. 12. Replanning route from the faculty of law to the faculty of agriculture
(a) Before replanning (b) After replanning.

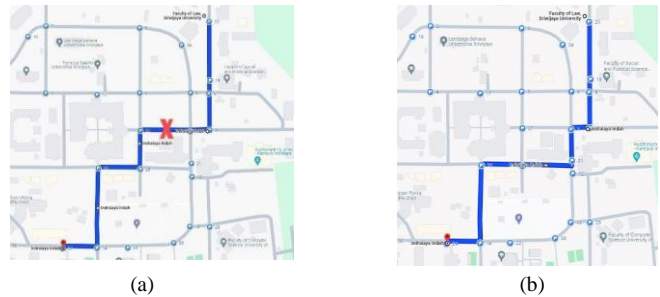


Fig. 13. Replanning route from the faculty of public health to the faculty of law (a) Before replanning (b) After replanning.

From the five route replanning trials conducted, it is evident that the route replanning system using the D-Star Lite algorithm successfully performs the route replanning process. Therefore, when an obstacle or blockage occurs, it generates a new optimal route to follow. Consequently, the D-Star Lite algorithm is suitable for real-time route replanning system testing.

C. Real-time Path Planning Testing

Next, this section discusses the path planning system testing under real-time conditions. In this test, an autonomous electric vehicle is used, with its position monitored in real-time via GPS. The objective is to evaluate the path planning system, designed with the D-Star Lite algorithm, to guide the autonomous electric vehicle towards its destination by following the optimal route.

In this test, the autonomous electric vehicle will move from its starting position, the Digital Control Laboratory in the Electrical Engineering Department (node 36), to its destination, the Faculty of Engineering Dean's office building (node 0). The best route will then be determined from the starting position to the destination. The optimal route found passes through nodes 36 → 37 → 38 → 39 → 40 → 41 → 42 → 43 → 44 → 45 → 0. For the autonomous electric vehicle to reach the

destination, it must pass through 10 node stages. The path taken is shown in Table VII.

In the real-time path planning tests conducted with an electric vehicle, as shown in Table VII, the autonomous electric vehicle successfully reached the target position by following the optimal route determined by the D-Star Lite algorithm. This demonstrates that the D-Star Lite algorithm is an effective method for finding the best route for autonomous electric vehicles.

TABLE VII. REAL-TIME PATH PLANNING TESTING

Node stages	Total distance (m)	Google maps distance (m)	Route based on google maps	Route taken electric vehicle
36 → 37	70,8	70,9		
37 → 38	116,6	116,5		
38 → 39	177,1	181,1		
39 → 40	54,3	54,5		
40 → 41	89,6	89,1		
41 → 42	67	67,8		
42 → 43	32,8	32,9		
43 → 44	46,6	48		
44 → 45	55,6	56,8		
45 → 0	92,6	92,4		

D. Real-Time Route Replanning Test

In this real-time route replanning experiment, a direct test will be conducted using an autonomous electric vehicle to determine whether the route replanning system of the D-Star Lite algorithm can effectively adjust the route when encountering obstacles in real-time conditions.

In this test, the autonomous electric vehicle is programmed to move from its starting point at the Digital Control Laboratory to the Faculty of Engineering Dean's office. The planned route passes through the following nodes: 36 → 37 → 38 → 39 → 40 → 41 → 42 → 43 → 44 → 45 → 0.

However, when the autonomous electric vehicle reaches node 38 and detects an obstacle blocking the path to node 39, the system identifies this path as impassable. The road closure toward node 39 is illustrated in Fig. 14.



Fig. 14. Road closure condition toward node 39.

Fig. 14 shows the visual closure of the road to node 39. This road closure occurs when the autonomous electric vehicle detects an obstacle blocking the path. The D-Star Lite algorithm handles this condition by dynamically recalculating an alternative route in real-time to ensure the vehicle can continue its journey toward the destination.

Once the road closure is detected, the replanning system in the D-Star Lite algorithm is activated to recalculate and adjust the route, ensuring that the autonomous electric vehicle can still reach its predetermined destination. After the replanning process, the new route is as follows: 36 → 37 → 38 → 46 → 26 → 11 → 2 → 4 → 1 → 0. A comparison between the original route (before replanning) and the new route (after replanning) is shown in Fig. 15.

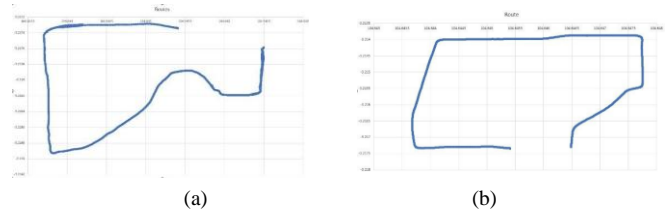


Fig. 15. Route replanning from the control and robotic laboratory to the faculty of engineering (a) Before replanning, (b) After replanning.

The real-time route replanning test demonstrated that the designed system can dynamically adjust the route in real-time whenever obstacles are encountered during the autonomous electric vehicle's journey toward its destination.

D-Star Lite uses distance as a core metric because it aligns with the algorithm's goal of finding optimal paths while efficiently adapting to dynamic environments. Distance serves as a universal measure of cost that simplifies computations, ensures practical relevance, and facilitates heuristic optimization.

If we compare the D-Star Lite algorithm with Dijkstra, the core characteristics are as follows: Dijkstra's algorithm is one of the earliest graph-based approaches for finding the shortest

path between nodes. It is deterministic and guarantees an optimal solution by systematically exploring all possible paths in a static and fully known environment. The algorithm's primary strength lies in its simplicity and optimality for static graphs. On the other hand, D-Star Lite is a dynamic and incremental path planning algorithm designed for environments that are partially known or subject to change. It builds on the principles of Dijkstra's algorithm but introduces significant enhancements to handle real-time updates efficiently. By focusing only on affected nodes when the environment changes, D-Star Lite reduces the computational overhead typically associated with path recalculations in dynamic scenarios. For the performance: Dijkstra's algorithm guarantees optimal paths in static settings but suffers from high computational complexity in large graphs due to its exhaustive exploration. This limitation becomes apparent when applied to vast areas or dense graphs, as the algorithm must evaluate all possible nodes and edges systematically [7]. D-Star Lite, however, is optimized for efficiency in dynamic and partially known environments. It updates only the necessary parts of the graph when changes occur, significantly reducing computational demands. Techniques like auto-clustering further enhance its performance by segmenting large maps, as demonstrated in Heo et al. (2022), where the Auto-Splitting D-Star Lite method reduced unnecessary node expansions [23].

Dijkstra and D-Star Lite algorithms cater to distinct path planning requirements. Dijkstra's algorithm is ideal for static, structured environments where optimality and simplicity are paramount. D* Lite, on the other hand, is tailored for dynamic and partially known environments, offering computational efficiency and adaptability. The choice between these algorithms depends on the specific use case, environmental constraints, and computational resources. Future advancements, such as hybrid approaches or machine learning integration, may further enhance their capabilities, bridging the gap between static and dynamic path planning needs.

V. CONCLUSIONS

After conducting five trials to compare the D-Star Lite algorithm with the A-Star algorithm, it was concluded that D-Star Lite demonstrates more optimal route-finding capabilities than A-Star. The average difference in route distance between the two algorithms was 97.7 meters, with D-Star Lite consistently providing shorter routes. Additionally, D-Star Lite's ability to calculate the distance to the target at each node enables it to perform route replanning effectively when encountering obstacles.

In the conducted tests, the D-Star Lite algorithm proved capable of finding the shortest route in real-time, covering a distance of 803 meters from the starting point at the Digital Control Laboratory to the Faculty of Engineering. Furthermore, the D-Star Lite algorithm successfully performed route replanning. Initially, the route was: 36 → 37 → 38 → 39 → 40 → 41 → 42 → 43 → 44 → 45 → 0. After replanning due to an obstacle, the route was adjusted to: 36 → 37 → 38 → 46 → 26 → 11 → 2 → 4 → 1 → 0. This study has shown the effectiveness of using the D-Star Lite algorithm in real-time applications for autonomous vehicles, even with paths containing obstacles. However, it is limited to simple obstacles.

Thus, further studies are needed to improve the algorithm's handling of different types of obstacles along the vehicle's path.

ACKNOWLEDGMENT

The research/publication of this article was funded by DIPA of Public Service Agency of Universitas Sriwijaya 2024. No SP DIPA 023.17.2.677515/2024, On November 24, 2023. In accordance with the Rector's Decree Number: 0013/UN9/LP2M.PT/2024, On May 20, 2024.

REFERENCES

- [1] R. Hussain and S. Zeadally, "Autonomous Cars: Research Results, Issues, and Future Challenges," *IEEE Commun. Surv. Tutorials*, vol. 21, no. 2, pp. 1275–1313, 2019, doi: 10.1109/COMST.2018.2869360.
- [2] J. Wang, Y. Yan, K. Zhang, Y. Chen, M. Cao, and G. Yin, "Path planning on large curvature roads using driver-vehicle-road system based on the kinematic vehicle model," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 311–325, 2021.
- [3] C. Jung, D. Lee, B. Kim, and D. H. Shim, "Lane level path planning for urban autonomous driving using vector map," in *2020 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, 2020, pp. 1–4.
- [4] J. Yu, J. Hou, and G. Chen, "Improved Safety-First A-Star Algorithm for Autonomous Vehicles," in *2020 5th International Conference on Advanced Robotics and Mechatronics (ICARM)*, Dec. 2020, pp. 706–710, doi: 10.1109/ICARM49381.2020.9195318.
- [5] J. Chen et al., "Path Planning for Autonomous Vehicle Based on a Two - Layered Planning Model in Complex Environment," *J. Adv. Transp.*, vol. 2020, no. 1, p. 6649867, 2020.
- [6] A. H. Ahmad, O. Zahwe, A. Nasser, and B. Clement, "Path Planning Algorithms For Unmanned Aerial Vehicle: Classification, Performance, and Implementation," in *2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME)*, 2023, pp. 1–6.
- [7] S. Sundarraj, R. V. K. Reddy, M. B. Basam, G. H. Lokesh, F. Flammini, and R. Natarajan, "Route planning for an autonomous robotic vehicle employing a weight-controlled particle swarm-optimized Dijkstra algorithm," *IEEE Access*, vol. 11, pp. 92433–92442, 2023.
- [8] S. Kadry, G. Alferov, and V. Fedorov, "D-Star Algorithm Modification," *Int. J. Online Biomed. Eng.*, vol. 16, no. 8, 2020.
- [9] M. Aizat, A. Azmin, and W. Rahiman, "A survey on navigation approaches for automated guided vehicle robots in dynamic surrounding," *IEEE Access*, vol. 11, pp. 33934–33955, 2023.
- [10] L. S. Liu et al., "Path Planning for Smart Car Based on Dijkstra Algorithm and Dynamic Window Approach," *Wirel. Commun. Mob. Comput.*, vol. 2021, 2021, doi: 10.1155/2021/8881684.
- [11] S. Erke, D. Bin, N. Yiming, Z. Qi, X. Liang, and Z. Dawei, "An improved A-Star based path planning algorithm for autonomous land vehicles," *Int. J. Adv. Robot. Syst.*, vol. 17, no. 5, pp. 1–13, 2020, doi: 10.1177/1729881420962263.
- [12] X. Li, "Path planning of intelligent mobile robot based on Dijkstra algorithm," in *Journal of Physics: Conference Series*, 2021, vol. 2083, no. 4, p. 42034.
- [13] S. W. G. Abusalim, R. Ibrahim, M. Zainuri Saringat, S. Jamel, and J. Abdul Wahab, "Comparative Analysis between Dijkstra and Bellman-Ford Algorithms in Shortest Path Optimization," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 917, no. 1, 2020, doi: 10.1088/1757-899X/917/1/012077.
- [14] R. Chen, "Dijkstra's Shortest Path Algorithm and Its Application on Bus Routing," *Proc. 2022 Int. Conf. Urban Plan. Reg. Econ. (2022)*, vol. 654, no. Upre, pp. 321 – 325, 2022, doi: 10.2991/aebmr.k.220502.058.
- [15] A. Candra, M. A. Budiman, and K. Hartanto, "Dijkstra's and A-Star in Finding the Shortest Path: A Tutorial," *2020 Int. Conf. Data Sci. Artif. Intell. Bus. Anal. DATABIA 2020 - Proc.*, pp. 28–32, 2020, doi: 10.1109/DATABIA50434.2020.9190342.

- [16] Y. Yan, "Research on the A Star Algorithm for Finding Shortest Path," *Highlights Sci. Eng. Technol.*, vol. 46, pp. 154–161, 2023.
- [17] M. R. Wayahdi, S. H. N. Ginting, and D. Syahputra, "Greedy, A-Star, and Dijkstra's algorithms in finding shortest path," *Int. J. Adv. Data Inf. Syst.*, vol. 2, no. 1, pp. 45–52, 2021.
- [18] K. Xie, J. Qiang, and H. Yang, "Research and optimization of d-start lite algorithm in track planning," *IEEE Access*, vol. 8, pp. 161920–161928, 2020.
- [19] S. nyeong Heo, J. Chen, Y. chi Liao, and H. hyol Lee, "Auto-splitting D* lite path planning for large disaster area," *Intell. Serv. Robot.*, vol. 15, no. 3, pp. 289–306, 2022.
- [20] S. Koenig and M. Likhachev, "Incremental A*," *Adv. Neural Inf. Process. Syst.*, 2002.
- [21] P. Paliwal, "A survey of a-star algorithm family for motion planning of autonomous vehicles," in *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, 2023, pp. 1–6.
- [22] R. Chen, J. Hu, and W. Xu, "An RRT-Dijkstra-based path planning strategy for autonomous vehicles," *Appl. Sci.*, vol. 12, no. 23, p. 11982, 2022.
- [23] S. Heo, J. Chen, Y. Liao, and H. Lee, "Auto-splitting D* lite path planning for large disaster area," *Intell. Serv. Robot.*, vol. 15, no. 3, pp. 289–306, 2022.

Stacking Regressor Model for PM_{2.5} Concentration Prediction Based on Spatiotemporal Data

Mitra Unik¹, Imas Sukaesih Sitanggang², Lailan Syaufina³, I Nengah Surati Jaya⁴

Department of Computer Science, IPB University, Bogor, Indonesia^{1,2}

Department of Informatics Engineering, Universitas Muhammadiyah Riau, Pekanbaru, Indonesia²

Department of Silviculture, IPB University, Bogor, Indonesia³

Department of Forest Management, IPB University, Bogor, Indonesia⁴

Abstract—This study presents the development of a predictive model for PM_{2.5} concentrations resulting from forest and peatland fires in Riau Province, utilizing the stacking regressor technique within an ensemble learning framework. The model integrates spatiotemporal data from remote sensing and ground-based sensors at a resolution of 1 km x 1 km, demonstrating its effectiveness in capturing the intricate patterns of PM_{2.5} concentrations. By combining Random Forest, Gradient Boosting Machine (GBM), and XGBoost, with RidgeCV as a meta-learner, the model attained optimal performance, achieving $R^2 = 0.851$, $MAE = 0.045 \mu\text{g}/\text{m}^3$, and $MSE = 0.003 \mu\text{g}/\text{m}^3$. The incorporation of temporal feature engineering techniques, including lag and rolling window methods, significantly enhanced prediction accuracy, enabling the model to effectively capture seasonal variations and temporal dynamics. Key variables, such as air temperature, evapotranspiration, and Aerosol Optical Depth (AOD), were found to exhibit strong correlations with PM_{2.5} concentrations. The findings from this research contribute to the formulation of data-driven policies for air quality management and pollution mitigation, with the potential for broader application in regions encountering similar environmental challenges.

Keywords—Ensemble learning; PM_{2.5} prediction; remote sensing; stacking regressor; spatio-temporal data

I. INTRODUCTION

PM_{2.5} (Particulate Matter ≤ 2.5 micrometres per cubic metre), which mainly comes from biomass burning such as forest and land fires, vehicle emissions, and coal combustion, causes various serious health impacts [1]. The measurement of PM_{2.5} due to forest fires faces challenges such as the episodic nature of fires, limited monitoring stations, and limited data availability [2]. Measurement approaches include ground stations with high accuracy but limited coverage, as well as satellite remote sensing that has wide and continuous coverage [3]. Satellite technology is effective in detecting fires, exposure to air pollution, and concentrations of aerosol particles including PM_{2.5} [4].

This research analyzes the performance of various machine learning algorithms, namely Gradient Boosting Machine (GBM), eXtreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Neural Network (NN), Long Short-Term Memory (LSTM), and Recurrent Neural Network (RNN), with the evaluation metrics of Coefficient of Determination (R^2), Mean Absolute Error (MAE), and Mean Squared Error (MSE). To improve prediction accuracy, feature engineering is applied

through the creation of lag and rolling window features. Lag features are based on the concept that historical values of a variable, such as PM_{2.5} concentrations, can influence current or future values, especially in time series data [5], [6]. Variables such as aerosol concentration, relative humidity, ground surface temperature, and air temperature are lagged to capture temporal influences. In addition, rolling window statistics, such as mean, median, and standard deviation, are calculated to capture long-term trends and seasonal patterns, helping the model understand the dynamics of PM_{2.5} changes influenced by seasonal factors or other external events [7]. Riau Province - Indonesia was chosen as the research location because it has the largest peatland in Sumatra Island, which is 3.89 million hectares out of a total of 5.85 million hectares. This condition makes Riau Province an appropriate location for study the impact of forest and peatland fires on PM_{2.5} concentrations [8], [9]. The aim of this research is to develop a machine learning ensemble model with optimised regressor stacking, and to integrate temporal dynamics and trend patterns to predict PM_{2.5} concentrations using 1 km x 1 km spatial and daily temporal remote sensing and ground sensor data, thereby supporting environmental management and public health policy.

II. RELATED WORK

Research relevant to this study includes various PM_{2.5} prediction models that integrate remote sensing-based predictor data, meteorological parameters and land use. Simple regression models such as Linear Regression (LR) and Multiple Linear Regression (MLR) are often used due to their simplicity, but they fail to capture non-linear relationships in high-dimensional datasets [10], [11], [12]. In contrast, machine learning techniques such as Random Forest (RF), Gradient Boosting (GB), and XGBoost have shown better ability in handling complex data and producing more accurate predictions [13], [14].

Further performance improvements are achieved through ensemble learning methods, such as Bagging, Boosting, and Stacking, which combine multiple models to reduce their individual weaknesses and improve prediction reliability [15], [16]. For example, research by Chen [17] showed that the stacking regressor model with meta-learner was able to achieve a coefficient of determination (R^2) of 0.85 and a Root Mean Squared Error (RMSE) of $17.3 \mu\text{g}/\text{m}^3$, which was superior to the single model. In addition, model combinations such as RF, GB, and Linear Mixed Regression (LMR) by Matsuki [18] and

findings Li [19] demonstrating the importance of spatial resolution in improving the accuracy of PM_{2.5} predictions.

Recent studies have also shown the successful application of ensemble models in predicting PM_{2.5} concentrations in various regions, such as China [18], South Asia [4], United States of America [20], and Italy [14]. Stacking regressor, in particular, is becoming a highly relevant method due to its ability to integrate predictions from base models such as RF, GB, and XGBoost using a meta-learner, which optimises the combination of predictions to produce more accurate final results [21]. This approach has shown its effectiveness in capturing complex and non-linear data patterns, which are often unreachable by conventional regression models.

III. METHOD

A. Location, Period and Research Data

This research was conducted in Riau Province, Indonesia, during the period 1 March 2022 to 31 March 2024. Geographically, Riau Province is located between 01°05'00" N to 02°25'00" N and 100°00'00" E to 105°05'00" East. Riau is the part of Sumatra Island that has the largest area of peatland, with 3.89 million hectares out of a total of 5.85 million hectares. The province frequently experiences forest and peatland ecosystem fires, which have the potential to cause haze disasters with transnational impacts. In this study, the prediction of PM_{2.5} concentrations due to forest and peatland ecosystem fires uses meteorological, environmental and geospatial data. Data were obtained from the air quality sensor of the Meteorology, Climatology and Geophysics Agency (BMKG) at Sultan Syarif Kasim II Airport Pekanbaru (101.45° East, 0.46° LU) as well as through satellite remote sensing. Data collection was conducted with daily temporal and spatial resolution, using a 30,000-metre buffer from the ground sensor, and a spatial buffer every 1,000-metres within the 30,000-metre range according to the Area of Interest (AOI), as shown in Fig. 1.

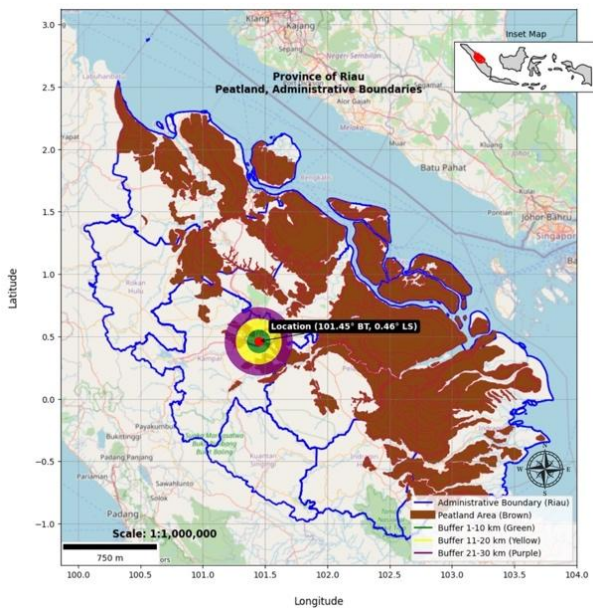


Fig. 1. Map of the study area and AOI.

B. Research Stage

The research utilizes machine learning algorithms as base models to enhance prediction accuracy through stacking, detailing the procedure, stacking architecture, and performance evaluation, as shown in Fig. 2.

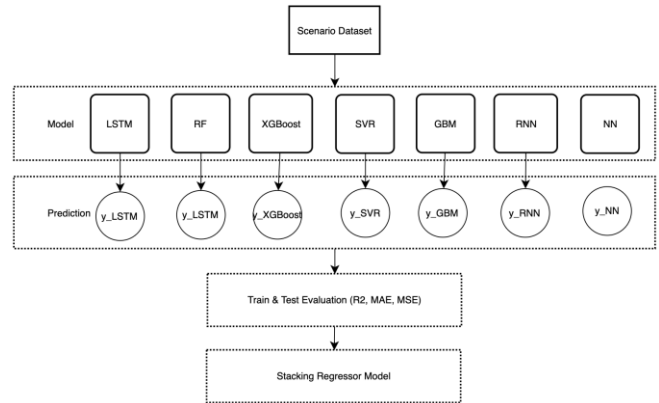


Fig. 2. General stages of modelling using the base model algorithm.

In the initial stage seven different machine learning algorithms as base models to get predictions from each model, namely LSTM, RF, XGBoost, SVR, GBM, NN, and RNN. Each base model generates predictions for the test data, which are referred to as (y_{LSTM}, y_{RF}, y_{XGBoost}, y_{SVR}, y_{GBM}, y_{RNN}, and y_{NN}). These predictions are generated from the training process performed on the training data. Each base model is evaluated using several evaluation metrics such as R², MSE, and MAE. This evaluation aims to measure how well each base model performs against the test data. The model with the best performance on these evaluation metrics is used as the basis for the next stage, which is the development of the ensemble learning model - Attention stacking regressor Model. Furthermore, the research process involves several main stages in applying the stacking regressor method to predict PM_{2.5} concentrations. These stages include dataset preparation, feature engineering, dataset sharing, basic model development, stacking regressor-meta learner modelling, model evaluation and result interpretation as visualised in Fig. 3.

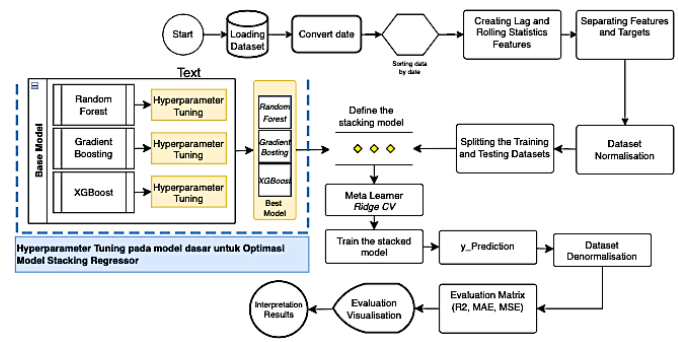


Fig. 3. Research stages of ensemble learning model - stacking regressor.

IV. RESULTS AND DISCUSSION

A. Research Dataset

In general, the predictors used as features of the prediction model for PM_{2.5} concentrations resulting from forest and

peatland fires, taken from ground and remote measurement sensor stations are as shown in below.

B. PM_{2.5} Concentration in the Study Period

During the study, PM_{2.5} concentrations were analysed through two graphs (Fig. 4): PM_{2.5} Level Distribution and PM_{2.5} Trend over Time. Fig. 4(a) shows that most of the PM_{2.5} concentrations were in the range of 15-30 µg/m³, falling into the Good to Moderate category, with concentrations above 55 µg/m³ rarely occurring, signalling generally safe air quality. Fig. 4(b) shows the daily trend of PM_{2.5} from 2022 to 2024, where 74.83% of days are in the Moderate category, 22.16% in the Good category, and 3.01% in the Unhealthy category for the Sensitive Group. Overall, the graph shows that while most days have safe to moderate air quality, there are certain periods where PM_{2.5} concentrations increase to potentially dangerous levels, especially for vulnerable groups. This emphasises the importance of continuous air quality monitoring to anticipate health risks, particularly during periods of increased pollution.

A pattern of fluctuations in PM_{2.5} concentrations was seen throughout the year, with a significant peak occurring at the end of 2023, which was most likely related to forest and peatland fires in Riau, covering more than 2,000 hectares in October 2023 [22].

C. Feature Correlation with PM_{2.5} Concentration

Feature correlation analysis aims to identify the variables that have the strongest relationship with PM_{2.5} concentrations in the dataset. Results in Fig. 5 shows the correlation heatmap for all features in the dataset against the PM_{2.5} target. Based on the results of the correlation analysis of PM_{2.5} in Fig. 6, the red colour represents a strong positive correlation, while the blue colour shows a significant negative correlation.

The feature with the greatest influence is TEMP (air temperature), which has a significant positive correlation, indicating that an increase in temperature tends to increase PM_{2.5} concentrations. In addition, ET (Evapotranspiration) features at certain radii, such as ET30, ET27, and ET28, also show strong positive correlations, signalling that areas with high evapotranspiration rates tend to have greater PM_{2.5} concentrations. AOD (Aerosol Optical Depth), especially at large radii such as *max_AOD*, also showed a significant relationship with PM_{2.5}, reinforcing the link between aerosol particles in the atmosphere and PM_{2.5} concentrations. These features were identified as the most relevant and influential variables in the air quality prediction model based on their strong relationship with PM_{2.5}.

TABLE I. COMMON PREDICTORS USED IN THE STUDY

Predictor	Description	Source	Unit	Temporal Resolution	Spatial Resolution
PM _{2.5} ground	Particulate Matter ≤ 2.5 µg/m ³	BMKG	µg/m ³	Daily	30 Km
TEMP	Relative Temperature	BMKG	°C	Daily	30 Km
PRS	Air pressure	BMKG	hPa	Daily	30 Km
PRE	Rainfall	BMKG	mm	Daily	30 Km
RHU	Relative Humidity	BMKG	%	Daily	30 Km
SSD	Sunlight hours	BMKG	Hours	Daily	30 Km
WIN	Wind speed	BMKG	m/s	Daily	30 Km
Min/Max_NDVI_bufer 1 to NDVI_30	NDVI	MODIS/061/MYD13A1	Unitless	16 Days	1 Km
Min/Max_AOD_bufer 1 to NDVI_30	Aerosol Optical Depth	MODIS (Terra & Aqua MAIAC MCD19A2.061)	Unitless	Daily	1 Km
Min/Max_ET_bufer 1 to NDVI_30	Evapotranspirasi	MODIS/061/MOD16A2	kg/m ²	8 Days	1 Km
Min/Max_LSTDay_bufer 1 to NDVI_30	Daytime surface temperature	MODIS/061/MOD11A1	°C	Daily	1 Km
Min/Max_LSTNight_bufer 1 to NDVI_30	Nighttime surface temperature	MODIS/061/MOD11A1	°C	Daily	1 Km

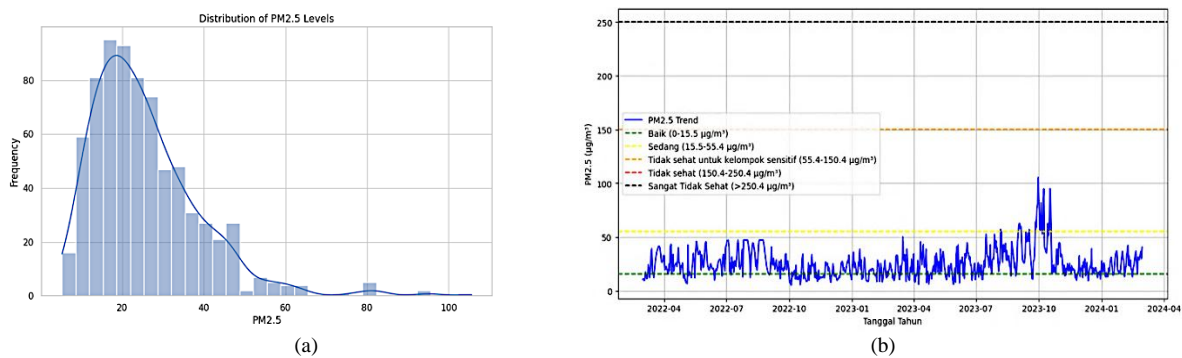


Fig. 4. PM_{2.5} concentration in the study period (a) Level distribution (b) Concentration trends.

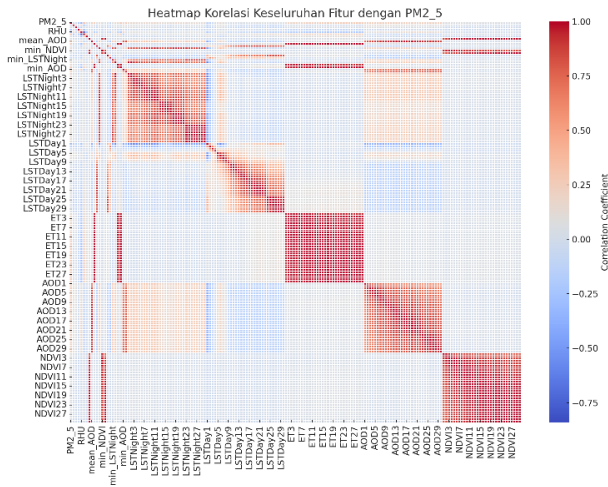


Fig. 5. Heatmap of feature correlation with PM_{2.5} concentration.

D. Evaluate the Performance of the base Model Algorithm

Table II shows the model performance evaluation results, for PM_{2.5} concentration prediction. The XGBoost model performed best on the training data with R² of 1.00, MAE of 0.07 μg/m³, and MSE of 0.01 (μg/m³)², indicating an almost perfect fit. However, on the test data, the performance decreased with an R² of 0.40, MAE of 7.18 μg/m³, and MSE of 109.65 (μg/m³)². The Random Forest model also showed good performance on training (R² 0.92, MAE 2.81 μg/m³, MSE 13.38 (μg/m³)²) but decreased on testing (R² 0.36, MAE 7.16 μg/m³, MSE 116.71 (μg/m³)²). The Gradient Boosting Machine, and Neural Network models had moderate performance with training R² of 0.84 and 0.81, and testing R² of 0.41 and 0.42, respectively. Meanwhile, the Support Vector Regression, LSTM and RNN models showed lower performance, with training R² ranging from 0.17 to 0.38 and testing R² between 0.14 and 0.27.

TABLE II. PERFORMANCE EVALUATION OF TRAINING AND TESTING MODELS

Model	Dataset Training Performance			Dataset Testing Performance		
	R ²	MAE	MSE	R ²	MAE	MSE
Random Forest	0.92	2.81	13.38	0.36	7.16	116.71
XGBoost	1.00	0.07	0.01	0.40	7.18	109.65
Support Vector Reg.	0.17	8.82	147.42	0.14	8.35	157.03
GBM	0.84	4.12	28.05	0.41	6.99	108.56
Neural Network	0.81	4.38	34.22	0.42	7.67	106.85
LSTM	0.38	7.81	109.17	0.27	8.36	133.65
RNN	0.33	8.07	118.26	0.23	8.39	141.53

E. Improving PM_{2.5} Predictions by Capturing Temporal Dynamics and Trend Patterns

This research applies feature engineering techniques by creating lag and rolling window features that allow the model to capture dynamics and temporal trend patterns in time series data.

1) *Lag creation*: The lag feature is based on the concept that the historical value of a variable may affect the current or future

value, especially in time series data. In the context of air pollution, PM_{2.5} concentration on a particular day can be influenced by meteorological conditions, especially AOD [16], [17], [18], [19], [20], [21] and the environment on previous days. Therefore, the variables that were considered to have significant influence and lag features were created include: Representation of aerosol concentration in the atmosphere, which is correlated with PM_{2.5} particles, Relative humidity of the air, which affects the formation and dispersion of pollutant particles, Ground surface temperature during the day, which can affect chemical and physical activity in the atmosphere, and Air temperature, an important factor in atmospheric processes.

The lag feature in time series data is calculated using a shift function, which represents the value of a variable in the previous time period. Conceptually, the lag-n value describes the value of a variable at a given time that has been shifted by n time steps backwards, with n representing the number of lag periods taken into account. In Python programming, the lag feature is created by shifting the data 4 time steps back using the .shift() method.

The 4-day lag was selected based on exploration to find the optimal value. The dynamic characteristics of PM_{2.5} that can persist and be influenced by atmospheric processes make this lag important in the model, allowing the utilisation of historical information to improve the accuracy of predicting concentration changes.

2) *Statistics rolling window*: Variables for which rolling statistics are calculated, such as max_AOD, mean_AOD, min_AOD, RHU, max_LSTDay, and TEMP, use specific time windows to apply statistical functions. The rolling mean provides a measure of the general trend by calculating the average of the values within that window, helping to understand the overall data pattern. The rolling mean calculation follows Eq. (1).

$$Rolling\ Mean = \frac{1}{n} \sum_{i=1}^n x_i \quad (1)$$

Where:

n : Total number of values in the window.

x_i : Individual values in the window.

$\sum_{i=1}^n x_i$: Sum of all values in the window.

Rolling median, If the number of data is even, the median is calculated as the average of the two middle values. If it is odd, the median is the centre value itself. The median is more resistant to outliers, so it gives a better idea of the centre of the data when there are extreme values. The even rolling median is calculated with Eq. (2) and the odd rolling median is calculated with Eq. (3).

$$Rolling\ Median_{even} = \frac{x_n + x_{n+1}}{2} \quad (2)$$

$$Rolling\ Median_{odd} = x_{\frac{n+1}{2}} \quad (3)$$

Where:

n : Total number of values in the window.

Rolling Standard Deviation (Std) measures the spread of data; the larger the standard deviation value, the greater the variation in the data. This is important for understanding how stable or volatile PM_{2.5} concentrations are. Rolling Standard Deviation (Std) is calculated with Eq. (4).

$$Std = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - Mean)^2} \quad (4)$$

Where:

n : Total number of values in the window.

x_i : Individual values in the window.

$Mean$: Average of the values in the window.

$(x_i - Mean)^2$: The squared difference between each value and the mean, which measures the deviation of each value from its centre.

Determination of the best rolling window size in modelling PM_{2.5} concentrations was done by utilising the XGBoost Regressor model. The tested rolling windows varied from size 3 to 20. For each rolling window size, the XGBoost model was trained and evaluated to obtain R². The model was trained using normalised data to ensure the data was in a comparable range. The results of the rolling window evaluation can be seen in Fig. 6.

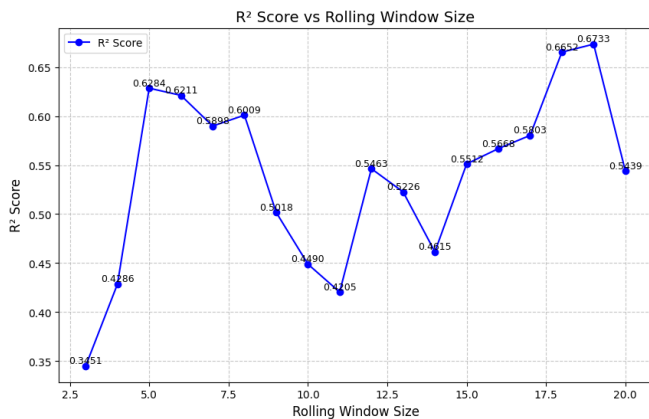


Fig. 6. Rolling window size evaluation results.

The analysis shows that the optimal rolling window size for PM_{2.5} prediction is 19 days with an R² Score of 0.6733. Rolling window sizes that are too small or too large tend to produce suboptimal performance, with the second peak at 5 days (R² = 0.6284) and the lowest performance at 10 days (R² = 0.4490). A larger rolling window is able to capture more historical information, thus improving the model's ability to predict PM_{2.5} dynamics. However, the application of the rolling and lag features led to the appearance of NaN values at the beginning of the data (e.g., the first 18 rows for a 19-day rolling window), which were removed using data.dropna() after the addition of the features. A summary comparison of the datasets before and after feature addition can be seen in Table III. The effect of data transformation with lag and rolling window features on data representation is shown in Fig. 7. The original variables (e.g., mean_AOD, RHU, max_LSTDay, min_LSTDay, and TEMP) shown in the left graph (blue) do not reflect the temporal dynamics clearly. In contrast, the transformed variables with a

lag period of 4 and a rolling window on the right graph (red) show a clearer and more dynamic historical pattern. Features such as mean_AOD_lag4 capture the influence of previous conditions on current values, thus improving the model's ability to understand temporal relationships. This transformation significantly improves the model's ability to capture complex patterns, which in turn is expected to improve the accuracy of PM_{2.5} predictions.

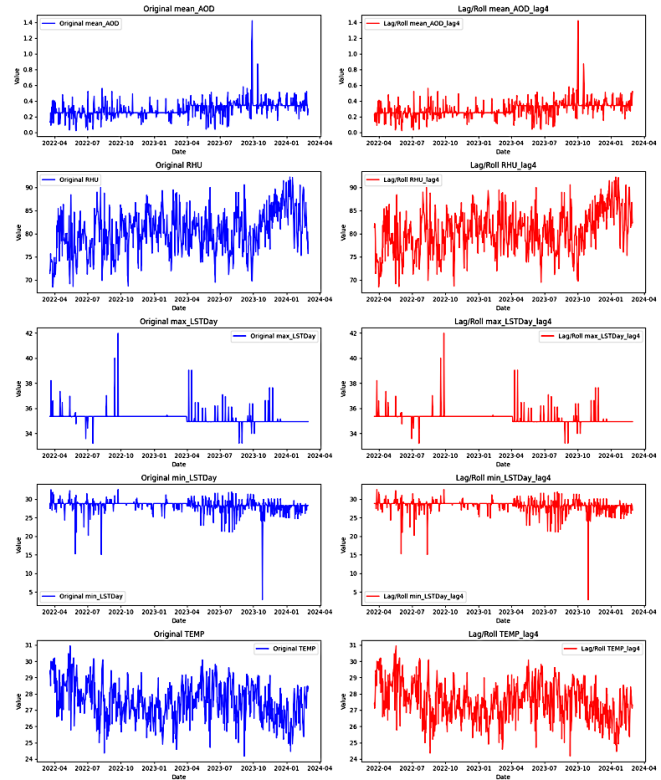


Fig. 7. Comparison of datasets before and after adding lag and rolling window features.

TABLE III. SUMMARY OF DATASETS BEFORE AND AFTER LEG AND ROLLING WINDOWS PROCESSING

Criteria	Before Lag & Rolling Feature Addition	After Addition of Lag & Rolling Feature
Number of Rows	731	713
Number of Columns	173	175
Average mean_AOD	0.2915	0.2934
Average RHU	80.43	80.45
Average max_LSTDay	35.23	35.23
Average min_LSTDay	28.22	28.23
Mean TEMP	27.54	27.53
Standard Deviation of mean_AOD	0.1076	0.1076
RHU Standard Deviation	4.59	4.63
Standard Deviation of max_LSTDay	0.53	0.32
min_LSTDay Standard Deviation	1.89	1.47
TEMP Standard Deviation	1.18	1.18

F. Performance of PM_{2.5} Prediction Model with and without Temporal Features

Before the temporal features were applied, the top three basic models-Random Forest, Gradient Boosting Machine, and XGBoost-had relatively low R² values of 0.36, 0.41, and 0.40, and high MAE and MSE. However, after the temporal features were included, the performance of the models improved significantly. Random Forest recorded an R² of 0.761, Gradient Boosting Machine achieved an R² of 0.767, and XGBoost recorded the highest R² of 0.798, with lower MAE and MSE. This shows that the application of temporal features can substantially improve the accuracy of PM_{2.5} prediction. Table 4 presents the performance evaluation of PM_{2.5} prediction models before and after the addition of temporal features (lag and rolling window).

TABLE IV. EVALUATION OF PM_{2.5} PREDICTION MODEL PERFORMANCE BEFORE AND AFTER INCORPORATING TEMPORAL FEATURES ON THE TEST DATASET

Model	Before			After		
	R ²	MAE	MSE	R ²	MAE	MSE
RF	0.36	7.16	116.71	0,761	0,058	0,005
GBM	0.41	6.99	108.56	0,767	0,058	0,005
XGBoost	0.40	7.18	109.65	0,798	0,053	0,005

G. Development of Ensemble Learning Model - Stacking Regressor

An ensemble learning model is applied using the Stacking Regressor approach to predict PM_{2.5} concentrations due to forest and land fires along with the use of lag and rolling window features. The stacking approach combines multiple machine learning models (base learners) to improve prediction accuracy by utilising three strengths of each base model (RF, GBM and XGBoost). The results from these base models are then fed into a meta-learner, which in this case is RidgeCV. RidgeCV was selected as the meta-learner in this study for several technical reasons. First, RidgeCV employs L2 regularization to prevent overfitting and enhance model stability by reducing excessive model complexity. Second, RidgeCV is effective in addressing multicollinearity among the predictions from base models. Third, it automatically performs cross-validation to select the optimal alpha parameter, ensuring an appropriate balance between bias and variance. Additionally, RidgeCV is computationally efficient compared to other meta-learners and is versatile in integrating predictions from various base models with different characteristics (e.g., Random Forest, which tends to be more robust with non-linear data, and XGBoost, which is more sensitive to structured data) [23].

We used the best alpha value (0.1) from the search results on a logarithmic scale from 10⁻⁶ to 10⁶ to effectively combine the predictions from the base model. Once trained, the stacking regressor model using RidgeCV as a meta-learner gave excellent results. The evaluation results on the test dataset (see Table V) showed an R² value of 0.845, with an MAE of 0.044 µg/m³ and MSE of 0.003 (µg/m³)².

H. Hyperparameter Tuning for Base Model Optimisation and Stacking Regressor via Grid Search

Grid Search with Cross-Validation (GSCV) is a robust method for optimizing hyperparameters in deep learning models, where cross-validation plays a critical role in enhancing model accuracy by systematically using different subsets of the training data for both training and testing [24], [25]. This approach evaluates the performance of hyperparameters across all potential configurations, making it a thorough and exhaustive search technique [26]. In this study, hyperparameter tuning was conducted using Grid Search to enhance the performance of each base model based on neg_mean_squared_error, with five-fold cross-validation (cv=5) ensuring the stability of performance, and n_jobs=-1 utilized to fully leverage all available processors. The optimal parameters identified through Grid Search were subsequently employed for the base learners, as detailed in Table VI.

TABLE V. EVALUATION OF ENSEMBLE LEARNING MODEL - STACKING REGRESSOR

Model	R ²	MAE	MSE
Stacking Regressor	0,845	0,044	0,003

TABLE VI. INITIAL PARAMETER RESULTS AND HYPERPARAMETER TUNING RESULTS WITH GRID SEARCH FOR EACH MODEL

Model	Parameters	Initial Parameters values	Parameter value after tuning
GBM	n_estimators	100	200
	learning_rate	0.1	0.2
	max_depth	3	3
	min_samples_split	2	5
	min_samples_leaf	1	1
	random_state	42	42
XGBoost	n_estimators	200	100
	learning_rate	0.1	0.1
	max_depth	5	6
	subsample	0.8	1.0
	min_child_weight	1	1
	colsample_bytree	-	1.0
	objective	'reg'	'reg'
	random_state	42	42
RF	n_estimators	100	300
	max_depth	None (unlimited)	None (unlimited)
	min_samples_split	2	2
	min_samples_leaf	1	1
	random_state	42	42
RidgeCV (meta-learner)	alphas	np.logspace(-6, 6, 13)	np.logspace(-6, 6, 13)
	store_cv_values	True (opsional)	True (opsional)

After tuning, significant improvements were observed in the models (Table VII). For Gradient Boosting, the number of estimators ($n_estimators$) increased from 100 to 200, and the learning rate ($learning_rate$) from 0.1 to 0.2, enhancing learning detail at the risk of overfitting. In XGBoost, $n_estimators$ decreased from 200 to 100, but max_depth and $subsample$ increased, balancing tree depth and data sampling efficiency. For Random Forest, $n_estimators$ increased from 100 to 300, improving model stability and accuracy by averaging more tree predictions.

I. Performance Evaluation of the Stacking Regressor Model

After optimisation, the three base models were combined using the stacked regressor model, where the predictions from each base model became the input for the meta-learner (RidgeCV). Table VII shows the evaluation of the stacked regressor model before and after hyperparameter tuning. Fig. 8 displays the scatter plot between the actual and predicted values for each tuned base model as well as the meta-learner. The points on the stacking regressor are closer to the reference line ($y = x$), indicating higher prediction accuracy compared to the base model.

TABLE VII. MODEL EVALUATION BEFORE AND AFTER HYPERPARAMETER TUNNING

Model	Before hyperparameter tuning	After hyperparameter tuning	Increase R ²	MAE decrease	MSE Decrease
RF	R ² = 0,761, MAE = 0,058, MSE = 0,005	R ² = 0,776, MAE = 0,057, MSE = 0,005	+0,015	-0,001	0,000
GBM	R ² = 0,767, MAE = 0,058, MSE = 0,005	R ² = 0,781, MAE = 0,055, MSE = 0,005	+0,014	-0,003	0,000
XGBoost	R ² = 0,798, MAE = 0,053, MSE = 0,005	R ² = 0,835, MAE = 0,048, MSE = 0,004	+0,037	-0,005	-0,001
Stacking Regressor	R ² = 0,845, MAE = 0,044, MSE = 0,003	R ² = 0,851, MAE = 0,045, MSE = 0,003	+0,006	+0,001	0,000

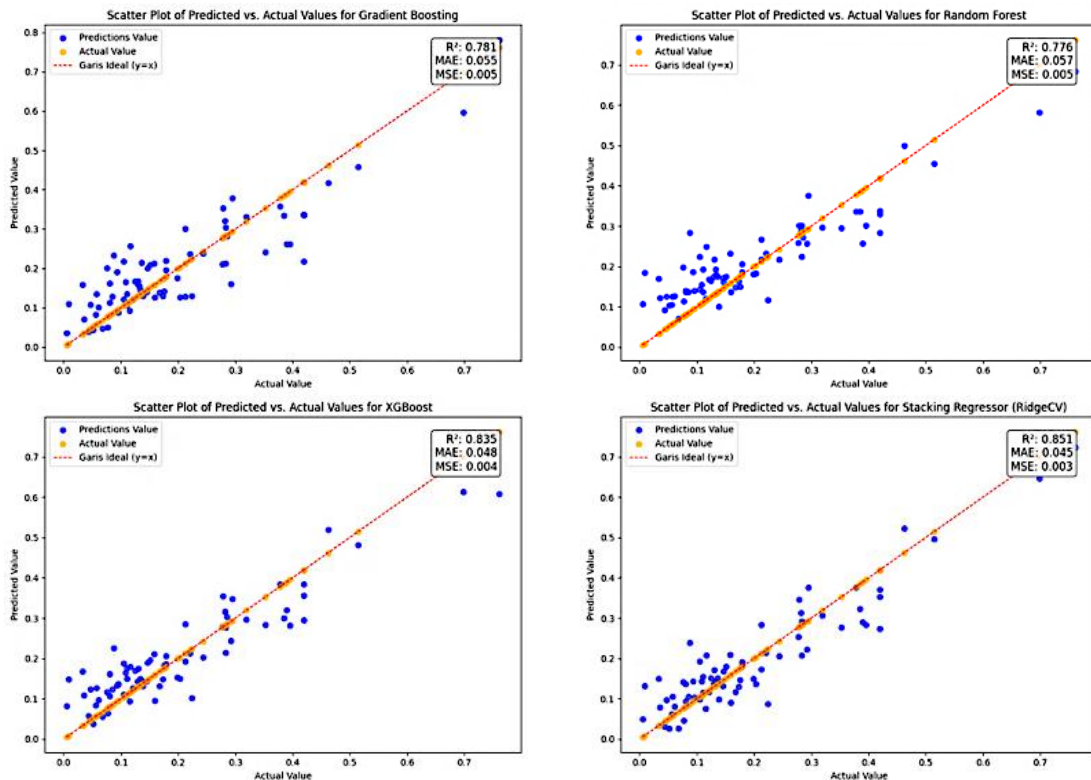


Fig. 8. Scatter plot of hyperparameter tuning performance of prediction model versus actual value.

V. CONCLUSION

This study successfully developed an effective prediction model for PM_{2.5} concentrations caused by forest and peatland fires in Riau Province, employing an ensemble learning approach through the stacking regressor method. The model outperforms other methods, demonstrating superior prediction performance due to the integration of spatiotemporal data from remote sensing and ground sensors. By combining base models such as Random Forest, Gradient Boosting Machine (GBM),

and XGBoost, optimized with RidgeCV as a meta-learner, the model achieved optimal performance with R² = 0.851, MAE = 0.045 μg/m³, and MSE = 0.003 μg/m³. The application of temporal feature engineering techniques, including lag and rolling window, significantly enhanced the model's accuracy, enabling a better understanding of seasonal patterns and temporal dynamics in PM_{2.5} concentrations. Key variables such as air temperature, evapotranspiration, and Aerosol Optical Depth (AOD) were found to have strong correlations with PM_{2.5}

concentrations, highlighting the critical role of atmospheric conditions in influencing air pollution levels. This research makes a significant contribution to the development of data-driven air pollution mitigation policies and holds potential for global application in regions facing similar pollution challenges, supporting efforts for more responsive and evidence-based air quality policy planning and public health management.

ACKNOWLEDGMENT

We would like to thank BMKG Sultan Syarif Kasim II Pekanbaru-Riau for its valuable cooperation and contribution in collecting the data on which this research is based. We also express our appreciation to the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for providing financial support for the Doctoral Dissertation Research Grant, so that this research can be carried out properly.

REFERENCES

- [1] P. Thangavel, D. Park, and Y. C. Lee, "Recent Insights into Particulate Matter (PM_{2.5})-Mediated Toxicity in Humans: An Overview," Jun. 01, 2022, MDPI. doi: 10.3390/ijerph19127511.
- [2] Ditppu-KLHK, "Kondisi Kualitas Udara Di Beberapa Kota Besar Tahun 2019," Direktorat Pengendalian Pencemaran Udara- KLHK. Accessed: Feb. 06, 2022. [Online]. Available: <https://ditppu.menlhk.go.id/portal/kontak-kami/?token=E7fKNFZqQzWdteaDKXW>
- [3] N. Islam, S. K. Khan, A. Rehman, U. Aftab, and D. Syed, "Stock Prediction for ARGAM Companies Dataset," KIET Journal of Computing and Information Sciences, vol. 6, no. 2, pp. 1–13, 2023.
- [4] A. Mhawish et al., "Estimation of High-Resolution PM_{2.5} over the Indo-Gangetic Plain by Fusion of Satellite Data, Meteorology, and Land Use Variables," Environ Sci Technol, vol. 54, no. 13, pp. 7891–7900, 2020, doi: 10.1021/acs.est.0c01769.
- [5] Z. Y. Chen et al., "Extreme gradient boosting model to estimate PM_{2.5} concentrations with missing-filled satellite data in China," Atmos Environ, vol. 202, pp. 180–189, 2019, doi: 10.1016/j.atmosenv.2019.01.027.
- [6] J. Wei et al., "Estimating 1-km-resolution PM_{2.5} concentrations across China using the space-time random forest approach," Remote Sens Environ, vol. 231, no. April, p. 111221, 2019, doi: 10.1016/j.rse.2019.111221.
- [7] C. Li, N. C. Hsu, and S. C. Tsay, "A study on the potential applications of satellite data in air quality monitoring and forecasting," Atmos Environ, vol. 45, no. 22, pp. 3663–3675, 2011, doi: 10.1016/j.atmosenv.2011.04.032.
- [8] PRIMS, "Data Kondisi Lahan Gambut dan Perkembangan Terkini Terkait Restorasi di Tujuh Provinsi Prioritas BRGM.," Badan Restorasi Gambut. Accessed: Feb. 01, 2022. [Online]. Available: <https://prims.brg.go.id/peta>
- [9] S. Ritung, "Sosialisasi Peta Gambut BBSDLP 2019," in Perubahan Luasan Lahan Gambut Dari Hasil Pemutakhiran Pemetaan Lahan Gambut, Bogor, 2020. [Online]. Available: http://sawitwatch.or.id/wp-content/uploads/2020/12/TSVOL27_Gambut_BBSDLP_021220.pdf
- [10] M. Sorek-Hamer, A. W. Strawa, R. B. Chatfield, R. Esswein, A. Cohen, and D. M. Broday, "Improved retrieval of PM_{2.5} from satellite data products using non-linear methods," Environmental Pollution, vol. 182, pp. 417–423, 2013, doi: <https://doi.org/10.1016/j.envpol.2013.08.002>.
- [11] R. O. Saunders, J. D. W. Kahl, and J. K. Ghorai, "Improved estimation of PM_{2.5} using Lagrangian satellite-measured aerosol optical depth," Atmos Environ, vol. 91, pp. 146–153, 2014, doi: <https://doi.org/10.1016/j.atmosenv.2014.03.060>.
- [12] J. Zhong et al., "Robust prediction of hourly PM_{2.5} from meteorological data using LightGBM," Natl Sci Rev, vol. 8, no. 10, 2021, doi: 10.1093/nsr/nwaa307.
- [13] M. Unik, I. S. Sitanggang, L. Syaufina, and I. N. S. Jaya, "PM_{2.5} Estimation using Machine Learning Models and Satellite Data: A Literature Review," vol. 14, no. 5, pp. 359–370, 2023.
- [14] A. Shtein et al., "Estimating Daily PM_{2.5} and PM₁₀ over Italy Using an Ensemble Model," Environ Sci Technol, vol. 54, no. 1, pp. 120–128, Jan. 2020, doi: 10.1021/acs.est.9b04279.
- [15] L. K. Hansen and P. Salamon, "Neural network ensembles," IEEE Trans Pattern Anal Mach Intell, vol. 12, no. 10, pp. 993–1001, 1990.
- [16] H. Tian, H. Kong, and C. Wong, "A Novel Stacking Ensemble Learning Approach for Predicting PM_{2.5} Levels in Dense Urban Environments Using Meteorological Variables: A Case Study in Macau," Applied Sciences, vol. 14, no. 12, p. 5062, Jun. 2024, doi: 10.3390/app14125062.
- [17] J. Chen, J. Yin, L. Zang, T. Zhang, and M. Zhao, "Stacking machine learning model for estimating hourly PM_{2.5} in China based on Himawari 8 aerosol optical depth data," Science of the Total Environment, vol. 697, p. 134021, 2019, doi: 10.1016/j.scitotenv.2019.134021.
- [18] K. Matsuki, V. Kuperman, and J. A. Van Dyke, "The Random Forests statistical technique: An examination of its value for the study of reading," Sci Stud Read, vol. 20, no. 1, pp. 20–33, 2016, doi: 10.1080/10888438.2015.1107073.
- [19] X. Li, Y. Li, Q. Ma, and S. Wang, "Random Forest Model for PM_{2.5} Concentration in China Using Himawari-8 Hourly AOD Product," in 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, 2021, pp. 1935–1938. doi: 10.1109/IGARSS47720.2021.9554364.
- [20] G. Geng, X. Meng, K. He, and Y. Liu, "Random forest models for PM_{2.5} speciation concentrations using MISR fractional AODs," Environmental Research Letters, vol. 15, no. 3, p. 34056, 2020, doi: 10.1088/1748-9326/ab76df.
- [21] D. H. Wolpert, "Stacked generalization," Neural Networks, vol. 5, no. 2, pp. 241–259, 1992, doi: [https://doi.org/10.1016/S0893-6080\(05\)80023-1](https://doi.org/10.1016/S0893-6080(05)80023-1).
- [22] Frislidia, "Lebih dari 2.000 hektare lahan terbakar di Riau hingga 8 Oktober 2023," ANTARA, 2023. Accessed: Nov. 19, 2024. [Online]. Available: <https://www.antaraneews.com/berita/3765324/lebih-dari-2000-hektare-lahan-terbakar-di-riau-hingga-8-oktober-2023>
- [23] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning Data Mining, Inference, and Prediction, 2nd ed., vol. 2. Springer New York, NY, 2017. doi: <https://doi.org/10.1007/978-0-387-84858-7>.
- [24] A. Chadha and B. Kaushik, "A Hybrid Deep Learning Model Using Grid Search and Cross-Validation for Effective Classification and Prediction of Suicidal Ideation from Social Network Data," New Gener Comput, vol. 40, no. 4, pp. 889–914, Dec. 2022, doi: 10.1007/s00354-022-00191-1.
- [25] S. M. Malakouti, M. B. Menhaj, and A. A. Suratgar, "The usage of 10-fold cross-validation and grid search to enhance ML methods performance in solar farm power generation prediction," Clean Eng Technol, vol. 15, Aug. 2023, doi: 10.1016/j.clet.2023.100664.
- [26] G. M. Habtemariam, S. K. Mohapatra, and H. W. Seid, "Software reliability prediction using ensemble learning with random hyperparameter optimization," Review of Computer Engineering Research, vol. 11, no. 1, pp. 1–15, Jan. 2024, doi: 10.18488/76.v11i1.3597.

Feature Substitution Using Latent Dirichlet Allocation for Text Classification

Norsyela Muhammad Noor Mathivanan¹, Roziyah Mohd Janor², Shukor Abd Razak³, Nor Azura Md. Ghani^{4*}
College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia^{1, 2, 4}
School of Computing and Creative Media, University of Wollongong Malaysia, Shah Alam, Malaysia¹
Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Malaysia³

Abstract—Text classification plays a pivotal role in natural language processing, enabling applications such as product categorization, sentiment analysis, spam detection, and document organization. Traditional methods, including bag-of-words and TF-IDF, often lead to high-dimensional feature spaces, increasing computational complexity and susceptibility to overfitting. This study introduces a novel Feature Substitution technique using Latent Dirichlet Allocation (FS-LDA), which enhances text representation by replacing non-overlapping high-probability topic words. FS-LDA effectively reduces dimensionality while retaining essential semantic features, optimizing classification accuracy and efficiency. Experimental evaluations on five e-commerce datasets and an SMS spam dataset demonstrated that FS-LDA, combined with Hidden Markov Models (HMMs), achieved up to 95% classification accuracy in binary tasks and significant improvements in macro and weighted F1-scores for multiclass tasks. The innovative approach lies in FS-LDA's ability to seamlessly integrate dimensionality reduction with feature substitution, while its predictive advantage is demonstrated through consistent performance enhancement across diverse datasets. Future work will explore its application to other classification models and domains, such as social media analysis and medical document categorization, to further validate its scalability and robustness.

Keywords—Feature extraction; feature selection; Latent Dirichlet Allocation; text classification; Hidden Markov Model; dimensionality reduction

I. INTRODUCTION

The exponential growth of online content has transformed digital platforms into key sources for global information acquisition and dissemination. With the rise of unstructured text data from these platforms, there is an increasing need for efficient techniques to analyze and manage large-scale text data, which often surpasses numeric data in volume and complexity [1], [2]. Text mining has emerged as a crucial tool for processing unstructured data, supporting decision-making through tasks like classification, clustering, summarization, association rule mining, and topic detection [2]. Among these tasks, text classification plays a vital role in organizing diverse textual data, including e-commerce products, tweets, news articles, and customer reviews, into structured groups [3]. This process has been widely adopted in various fields, such as product categorization [4], [5], sentiment analysis [6], spam detection [7], news classification [8], and medical document classification [9].

Effective text classification relies on noise-free features that capture the essential semantic meaning of the data [2]. However, large-scale text corpora are often high-dimensional, posing challenges for computational efficiency and model accuracy. Input data preparation, particularly through pre-processing, feature extraction, and feature selection, is essential to ensure the performance of classification models [10]. Pre-processing techniques, such as tokenization, stop-word removal, and stemming, reduce the data's complexity and improve model accuracy by eliminating noise. Feature extraction creates a compact feature space by transforming the original data, while feature selection identifies a subset of relevant features that distinguish different categories [11]. These techniques have a profound impact on model accuracy and efficiency but often struggle with the high dimensionality inherent in text data [12].

Traditional dimensionality reduction methods, such as k-means clustering [13], two-stage feature selection [14], and hybrid approaches combining ReliefF and principal component analysis [15], aim to address these challenges. However, these methods may not fully integrate semantic context into the feature representation, limiting their impact on classification performance. To address these limitations, this study introduces Feature Substitution using Latent Dirichlet Allocation (FS-LDA), a novel technique that combines dimensionality reduction with semantic feature grouping.

FS-LDA leverages the topic modeling capabilities of Latent Dirichlet Allocation (LDA) to group and substitute high-probability topic words into unified representations, reducing dimensionality while preserving meaningful textual features [16]. Unlike feature selection, which eliminates irrelevant features, FS-LDA substitutes related features based on topic modeling, enhancing the representation of the data for classification tasks. A new term called feature substitution is introduced mainly to replace related features according to defined groups from a topic modelling technique. Previous studies have demonstrated the effectiveness of LDA in dimensionality reduction and topic clustering, but its application in feature substitution remains unexplored [16]. By integrating FS-LDA into the pre-processing phase, this study seeks to evaluate its effectiveness in improving classification accuracy and efficiency across various datasets.

The FS-LDA technique offers a significant advantage by reducing feature complexity while maintaining the semantic integrity of the data. This novel approach simplifies input data preparation and enhances the performance of classification

models, as demonstrated through experiments in this study. The findings highlight FS-LDA's potential as a scalable, efficient, and effective method for text classification tasks in real-world applications.

II. LITERATURE REVIEW

The current technological advancement and new research on machine learning over the years contribute tools to deal with a high volume of documents using algorithms that extract information from their original texts. One possible approach to simplify high-volume data is to apply some form of dimensionality reduction. Methods like feature extraction and feature selection offer distinct benefits; feature extraction transforms the original data into a compact feature space, while feature selection retains only the most relevant features, potentially improving model efficiency. Commonly, researchers used n-grams models such as unigram, bigram, and trigram to extract features. Linguistic pattern methods, statistical methods, or a combination of both can enhance the extraction process. Hybridization between a linguistic approach and a statistical method efficiently provides reliable features while improving accuracy, especially in classifying Arabic text [17].

Meanwhile, some researchers preferred to enhance the feature selection technique used in their study to improve classification rates. For instance, previous researchers used collaborative feature-weighted multi-view fuzzy c-means clustering [18] and hybrid binary grey wolf with harris hawks optimizer [19]. The utilization of both techniques accordingly provides a better data pre-processing process. However, these methods often lack a semantic perspective, which is addressed by techniques like Latent Dirichlet Allocation (LDA). Over the years, LDA has been widely used to explore features using a hidden topic analysis [20]. It is known as a classical statistical model for topic mining in natural language processing, and it was proposed by Blei et al. [21]. This model discovers various topics in many documents and builds to model text data subject information. Many domain retrievals involving machine learning models applied the LDA model to help deal with text-related problems [22]. Besides the LDA model, researchers often used another topic modelling approach, Latent Semantic Analysis (LSA) [23]. The model's weaknesses are its dependency on annotated training data and its tendency to overfit. Hence, LDA is often preferred over LSA due to its ability to handle sparse data and its probabilistic nature, which provides a more robust representation of text semantics. This advantage aligns with the study's objective to enhance text classification through semantically enriched feature substitution.

The LDA structure resembles the probabilistic variation of LSA known as Probabilistic Latent Semantic Analysis (PLSA) [24]. While LDA and its predecessor, Probabilistic Latent Semantic Analysis (PLSA), share probabilistic foundations, LDA's use of Dirichlet priors enables better generalization for unseen documents, addressing a critical limitation of PLSA and advancing its utility in text classification tasks [24]. The model learns a distribution over the topic for each document in training, but it is only applicable for training sets with the known topic distribution. The model cannot generate topics

from previously unseen documents. Meanwhile, the LDA model learns topic distribution as a random parameter vector and models based on Dirichlet prior. Researchers use symmetric Dirichlet distribution involving a similar value for all parameters in the LDA. The derivation methods commonly acquire the distributions are a variational inference [17] and Gibbs sampling [25].

Previous studies have successfully proved the efficiency and benefits of practicing this model. From the beginning, Blei et al. [21] discovered that LDA slightly decreases text classification performance but improves overall efficiency because of its dimensionality reduction characteristic. Researchers invented an LDA-based model known as Dual Latent Dirichlet Allocation (DLDA) to extract topics for short texts with knowledge obtained from long text data [26]. The improved model utilizes two sets of LDA topics where "target" and "auxiliary" represent short and long texts. The DLDA model performs better than the LDA model, primarily in clustering short text data based on entropy, purity and normalized mutual information as the evaluation criterion.

A previous study can merge the document's representation based on the LDA by applying labels to enhance text classifier performances [27]. The modified LDA works as a semi-supervised learning model where the model includes partial expert knowledge at word and document levels. There is accuracy rate improvisation as more documents are labelled. The modified LDA is feasible for real-world applications with many unlabeled data with few labelled data for training purposes. On the other hand, Cheng et al. [28] combine the idea of using the LDA and word co-occurrence patterns in the corpus to detect topics for a document. It addresses co-occurrence, such as bi-term individually as a semantic unit representing a single topic for recognizing the words most likely to be together. The LDA with word co-occurrence patterns combination improves the topic selection consistency for each document.

A study also merged the LDA with clustering through a Self-Aggregation based topic model (SATM) [29]. The proposed model helps detect relevant topics in short text data. A Multi-CoTraining (MCT) system implementation through LDA combination with Term Frequency-Inverse Document Frequency (TF-IDF) and Doc2Vec provides various feature sets for document classification [30]. The proposed model is robust when dealing with parameter changes. The performance of MCT is superior compared to other benchmark methods. Instead of using Doc2Vec, another study presents the combination of Word2Vec as the word embedding technique with the LDA [31]. The experimental result shows the proposed model outperforms the basic LDA. It can solve problems created by a Bag-of-Word (BOW) model related to high dimensionality and sparsity data.

An automatic text mining framework based on the LDA is proposed in the financial sector to analyze texts as financial disclosures from firms [32]. The topic model aims to find a firm's strengths and weaknesses through business units, activities, and processes depending on its risk. The proposed framework helps to improve the existing business management tools regardless of any business level. The LDA is also an

alternative representation model for BOW because it reduces the feature numbers for text classification [33]. The WEKA package has included the framework to provide a feasible option for other researchers to select features from their data sets.

LDA is used as a feature selection technique in Celard et al. [33] to create a new text representation model utilizing the probability of a document belonging to each topic. However, the probability is not yet used to substitute existing features extracted from classic representation models such as unigram and bigram models. The utilization of LDA topics in the feature selection process can greatly reduce the input data dimensionality while improving the classification model performance. Hence, this study's main objective is to assess the LDA model's efficiency in text classification as a feature substitution technique.

III. METHODOLOGY

This section briefly describes the proposed framework used in this study. The detailed description of the feature substitution technique provides a better understanding of the proposed technique for data preparation related to features. This study used HMM as the text classification model.

A. Proposed Framework

The study involves several steps before classifying the data, as shown in Fig. 1. The typical steps are data extraction, data pre-processing, feature extraction, and feature selection. These are the necessary steps in data preparation related to text classification. After data extraction, three pre-processing steps involve tokenization, stop word removal, and stemming [34]. Data pre-processing is vital to ensure the data is standardized and in proper form. The standardized way is achieved after applying the three pre-processing steps, where each observation is tokenized into words at first. Then, stop words are removed from the word list, and the remaining words are stemmed to ensure the words follow the root word forms.

Feature extraction and selection are essential to ensure the data are well transformed into significant and functional features before performing the classification process [35]. The choice of features may affect the classification model accuracy. Thus, the study compares two feature representation models, i.e., unigram and bigram, to observe their effect on classification performance. The feature selection used in the study was the filter method known as correlation-based feature selection (CFS). The study also compares the classification model before and after applying the proposed feature substitution technique. Then, the chosen features are used as inputs to perform the classification model. All the input data preparation steps were computed using R-Programming software. The classification step is done using Python programming software.

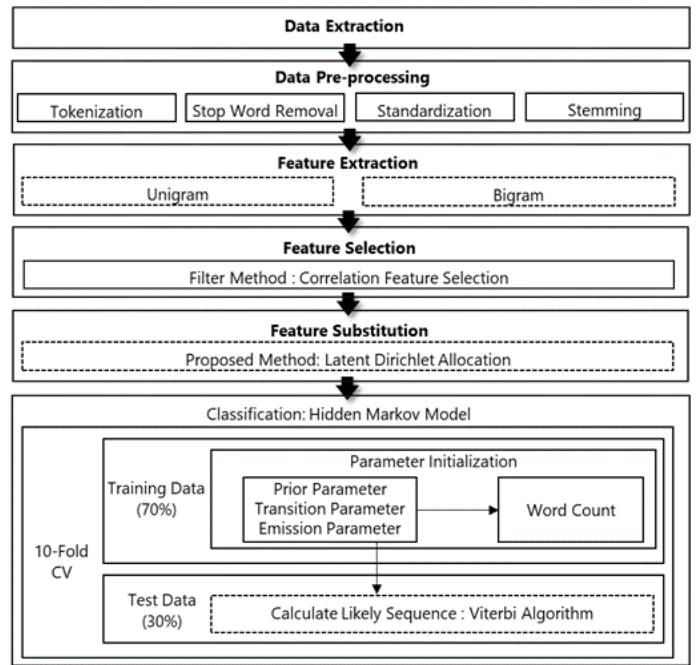


Fig. 1. Proposed study framework.

B. Feature Substitution using Latent Dirichlet Allocation

All the steps involved are standard procedures in text classification before training the classifier, except for applying the LDA model to perform the feature substitution. It is a generative probabilistic model of a document collection [15]. LDA searches for these latent semantic topics in the corpus [35], and it considers each document as a topic collection where each topic is a keyword collection. The topics are a collection of dominant keywords. These topics express an approach to quantitatively describe the document and describe the document content [36]. The critical factors in obtaining adequate keyword segregations are the text processing quality, topic diversity in the text, algorithm selection, and algorithm tuning.

The LDA algorithm input is basic units of discrete data, i.e., words in the text documents. The output of the LDA algorithm is a set of topics. For instance, each document's category belongs to an extensive collection of words, and documents can be observed by checking the words' occurrence in the documents. However, this method is costly and inefficient. Instead of checking every word in the document, another layer is initiated with a set of topics. The collection of words is mapped to the topics, and the topics are mapped to the documents. Hence, this action will reduce costs while increasing efficiency.

The study used LDA to group and substitute the features before applying the classification model. This model involves a generative process assuming that documents consist of a mixture of topics. Then, words from the typical vocabulary of each selected topic are drawn from each document. In the study, the document in topic modelling is represented by observation. Accordingly, LDA assumes that observations are described as a bag of words in a unigram or bigram model with different topics in different proportions. The pseudocode for the

proposed feature substitution technique using the LDA is shown in Algorithm 1.

Algorithm 1: Feature Substitution Technique using LDA

```

Initialize
O: Observations in the dataset
O0: The first observation
T: Topics in the observation
W: Word in the observation
P: Percentage of the highest probabilities in a topic

Compute
Assign each W in O0 a topic T

While (observation remain) do
  For each W in O0 do
    Assume the assigned topic T is wrong.
    Assume the assigned topic T for other W in O0 is correct.
    Update and analyze
      Calculate the probabilities to assign a topic T based on:
      Number of topics in the document.
      Number of times the same topic is assigned to the word across all document.
    End
  End
End
Repeat the process for all O
Remove overlap words with percentage P in all T
  For each T
  Assign a new topic name to W
  End
  
```

The calculation involves in LDA is to obtain the probability of words belonging to a topic where the procedure starts with randomly assigning each word in the observation *O* to one of *T* topics. Then, the required probabilities of each word, *W* can be computed after assuming the randomly assigned topic for that particular word is wrong. The computation of the first probability involves the proportion of words in observation *O* that are assigned to the topic *T*. This action is to observe how many words belong to the topic *T* for a given observation *O* excluding the current word *W*. If many words from observation *O* belongs to topic *T* it is more probable word *W* belongs to topic *T*.

The second probability involves the proportion of assignments to topic *T* out of all documents derived from the word *W*. This action is to observe how many observations are in topic *T* because of the word *W*. LDA represents documents as a collection of topics. A topic is also a collection of words. If a word has a high likelihood of appearing in a topic, all observations containing *W* will also be more strongly correlated with *T*. Similarly, if *W* is not very likely to be in *T*, documents including *W* will have a very low likelihood of being in *T*, because the rest of the words in *O* will belong to a different topic, giving *O* a higher probability for other topics. Even if *W* is added to *T*, it will not bring many of these observations to *T*. The probability that a *W* in observation *O* belongs to topic *T* is stated in Eq. (1).

$$P(T | W, O) = \frac{m \text{ of word } W \text{ in topic } T + \beta}{\text{total tokens in } T + \beta} \quad (1)$$

m represents the words in *O* that belong to *T*, adjusted by the hyperparameter α . The parameter α controls how topics are distributed in a document: a smaller α focuses the document on fewer topics, while a larger α mixes more topics evenly. Similarly, β manages the distribution of words within topics. A smaller β emphasizes a few dominant words, making topics more distinct, while a larger β spreads probabilities across many words, resulting in broader topics. Although each topic technically includes all words in the vocabulary, the most probable words define the topic, making it both meaningful and flexible.

After evaluating each word’s probability belonging to different topics based on the LDA model, the subsequent action is to substitute the non-overlap words with high probability in each topic. According to the LDA model analysis, these words become homogeneous by assigning the same name to represent the group of words that most probably belong to the topic. For example, Fig. 2 shows that Observation 1 is only about Topic 1.

In contrast, Observation 2 is a mix of Topic 1 and 2 because one of the words, “banana”, has a higher probability value in Topic 1 than in Topic 2. Specifically, each topic is represented as a probability distribution over a controlled vocabulary. Usually, all the words appear in the observation collection. In the example, Topic 1 has words such as “fresh” (3.41%), “drink” (2.35%), and “juice” (1.99%). Meanwhile, Topic 2 has words such as “biscuit” (2.73%), “mix” (2.21%), and “apple” (1.86%). These words are the three highest probabilities in each topic. Given this information, Topic 1 can be labeled as “drink” and Topic 2 as “food”. Consequently, Observation 1 is purely about “drink”, while Observation 2 is a mix of the “drink” and the “food” topics. The only observable variable is words from the observations, whereas all other variables, such as the topic distributions for each document and the word distributions for each topic, are hidden. Hence, LDA aims to infer these hidden distributions, given the observed words per observation.

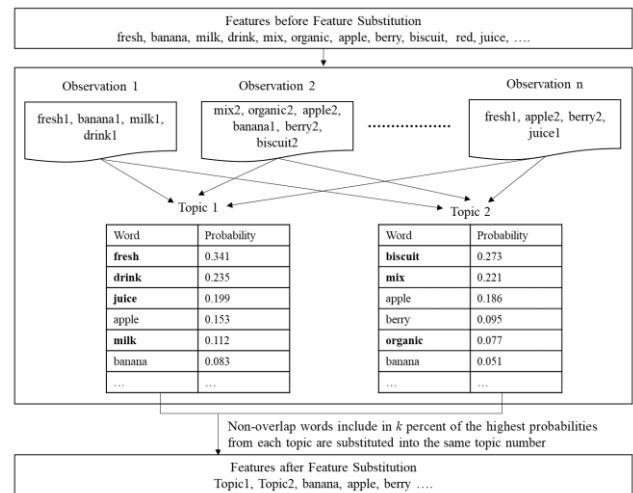


Fig. 2. Proposed feature substitution technique for input data preparation in text classification.

After applying LDA, each topic is represented by words with specific probabilities of belonging to that topic. The feature substitution technique replaces high-probability, non-overlapping words from each topic with a single constant term, such as “Topic1” or “Topic2,” ensuring that the selected words uniquely represent their topic. The study tested this substitution at different levels (10%, 20%, 30%, 40%, and 50% of the top words per topic). Fig. 3 illustrates how this technique represents data before applying the classification model, along with examples from a sample dataset.

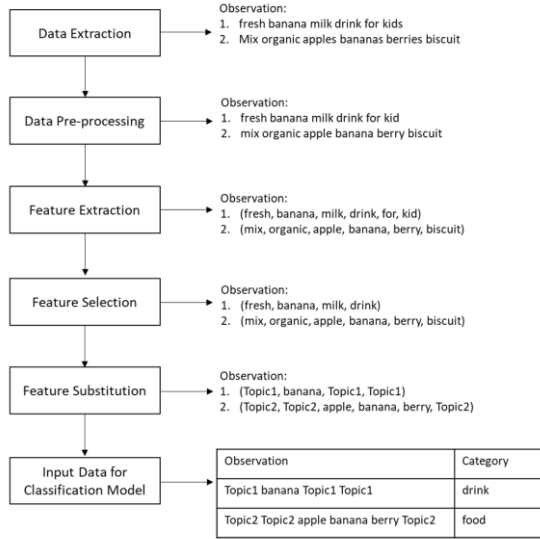


Fig. 3. Sample text representation with proposed feature substitution technique before applying classification model.

C. Classification

A Hidden Markov Model (HMM) is often applied to text classification as a supervised learning task. The application of HMM can be seen through various study areas related to text and language processing applications, e.g., text classification [37], text discretization [38], and information extraction [35]. The input data used for the supervised learning model is a corpus of words labeled with the correct category. Table I shows the components that specify an HMM.

TABLE I. HIDDEN MARKOV MODEL COMPONENTS

Symbol	Component	Description
Q	$q_1 q_2 \dots q_N$	A set of N states
A	$a_{11} \dots a_{ij} \dots a_{NN}$	A transition probability matrix A , each a_{ij} represents the moving probability from state i to state j
o	$o_1 o_2 \dots o_T$	A sequence of T observations, each one is drawn from a vocabulary $V = v_1, v_2, \dots, v_v$
B	$b_i(o_i)$	A sequence of observation likelihoods, also called emission probabilities, each expressing the probability of an observation o_i being generated from a state i
π	$\pi_1 \pi_2 \dots \pi_N$	An initial probability distribution over states. π_i is the probability that the Markov chain will start in the state i

HMM's decoding problem is finding the optimal state sequence given the observation sequence and the trained HMM. The Viterbi algorithm is commonly applied to find the most likely hidden state sequence based on every word sequence input. There are given the observation sequence for test data $\{o_t\}_{t=1}^N$ and trained HMM with parameters $\lambda = (\pi, A, B)$ to find the most likely sequence. The formula is presented in Eq. (2).

$$\underset{\{q_t\}_{t=1}^N}{\operatorname{argmax}} p(\{q_t\}_{t=1}^N | \{o_t\}_{t=1}^N) \quad (2)$$

The optimal hidden state sequence is produced for each word sequence of the test data using the Viterbi decoding algorithm. The prediction of the text data is based on the majority role, i.e., a product will be labeled as the drink category if the optimal hidden state sequence has more drink features than food features. Otherwise, the product is marked under the food category.

IV. DATASETS

This study utilizes two different text data to evaluate classification models' performance. The first data involves five e-commerce product data, which these datasets are crawled from an e-commerce website. Department of Statistics Malaysia (DOSM) has collected product information from one of the primary online store websites through the STATSBD A project known as Price Intelligence (PI) using its prototype web scraper. Another dataset is retrieved from the UCI repository. Table II presents a summary of all the datasets used in the study.

TABLE II. SUMMARY OF DATASETS

Data Name	Data Description	Class Number	Class Name (Instance Number per Class)	Instance Number
ECD01	E-Commerce pets products	2	food (265) and care & accessories (45)	310
ECD02	E-Commerce non-food products	2	cooking & dining (407) and party accessories (80)	487
ECD03	E-Commerce frozen food products	5	frozen food (291), yoghurt (162), ice cream (147), cheese (85), and juices (87)	772
ECD04	E-Commerce household products	6	laundry (370), air freshener (297), household kitchen cleaner (181), sundries (158), light bulbs (100), and toilet cleaner (100)	1206
ECD05	E-Commerce grocery products	14	cooking ingredients (677), chocolates & sweets (594), biscuits & cakes (491), snacks (440), sauces & dressings (364), canned food (331), pasta & instant noodles (294), baking (269), jam (220), cereals (208), dry condiments (206), sugar & flour (176), rice (138), and cooking oil (130)	4538
SPAM	SMS spam collection data set	2	ham (4827) and spam (747)	5574

A. Dataset's Characteristics

Each dataset's characteristics can be seen through its data distribution. Text length, word count, and class distribution can describe the data. The detailed characteristics are shown in Fig. 4 for each data set correspondingly. In class distribution for e-

commerce product datasets, ECD01 and ECD02 fall under binary classification problems. However, these two datasets have different text characteristics, as shown in Fig. 4. There are two dominant features in the ECD01 dataset, i.e., “food” and “cat”. Other features seem to have not much different frequent existences in the dataset compared to these two features. Meanwhile, there are six dominant features in ECD02, whereas other features are far less number of occurrences in the dataset. The variation of dominant features may affect a classification model, especially when using HMM because the parameter estimation is based on feature occurrences.

Three e-commerce product datasets, i.e., ECD03, ECD04, and ECD05, belong to multiclass classification problems. Usually, datasets with a higher number of classes tend to have a much lower classification model performance because of increased data complexity. ECD03 has a higher number of dominant features than the other two datasets. When a dataset has less prevalent features, such as features in ECD04, there is a tendency that is performing the proposed feature selection technique may not significantly reduce the number of features while improving the model performance. The reason is that the proposed model recognizes a group of features to be combined as one topic where the features must not belong to any pre-defined classes. The relatively similar number of occurrences for each feature in the dataset may emphasize that the features may have equal weight pertaining to any hidden topic created to reduce the features. Hence, there is an assumption that any dataset with a high number of dominant features may be beneficial for using the proposed feature substitution technique.

The text length plots represent the product description distributions for ECD01, ECD02, and ECD04, are appear to have an approximately normal distribution. Meanwhile, ECD03 and ECD05 have shorter product description lengths as their distribution is right-skewed. Typically, the term frequency distribution is based on the number of times a feature appears in a dataset divided by the total number of features in that dataset. Both axes are plotted on logarithmic scales in the term frequency distribution plot because the frequency of the most

frequent features is much higher than the frequency of the long tail of infrequent features that a figure of this size without a logarithmic transformation would look like the letter L.

The frequency distribution plots for all e-commerce product datasets illustrated the frequency curve decreases very steeply from the extremely high values corresponding to the most frequent features. They become progressively flattered until they reach an extensive level corresponding to the ranks assigned the tail of words occurring once. The same skewed shape is not specific to the datasets used in this study. Still, it often emerges in natural language texts, independently of tokenization or type mapping method, size, language, and textual typology [39]. The only difference is that the variation of inflected forms can be seen from the frequency distribution plots. Even though the overall pattern is the same, the number of very low-frequency forms in the three datasets, i.e. ECD01, ECD02, and ECD03, is lower than in the other two datasets.

The ordinary skewed structure of word frequency distributions was first comprehensively studied by Zipf [40]. The utilization of various datasets leads to frequency’s nonlinearly decreasing rank function. Theoretically, the high ranks fall more sharply than the low ranks. Fitting a straight line to the log-log curves is commonly rational and practicable. Fig. 5 visualizes the frequency distribution plot for each dataset according to Zipf’s law. The plots are generally not perfectly fitted, especially at the edges. The curve’s right edges represent features among the highest ranks with the lowest frequencies. The inconsistent patterns are because the increasingly more comprehensive horizontal lines, in accord with the rare words, are assigned different ranks but have the same frequency. The results may happen due to fitting a model consisting of many words with very near-continuous frequencies to an empirical curve, originally a discrete step function for high ranks.

Meanwhile, the left plot’s curved edges represent features among low ranks with high frequencies. Each plot portrayed a different degree of downward curves.

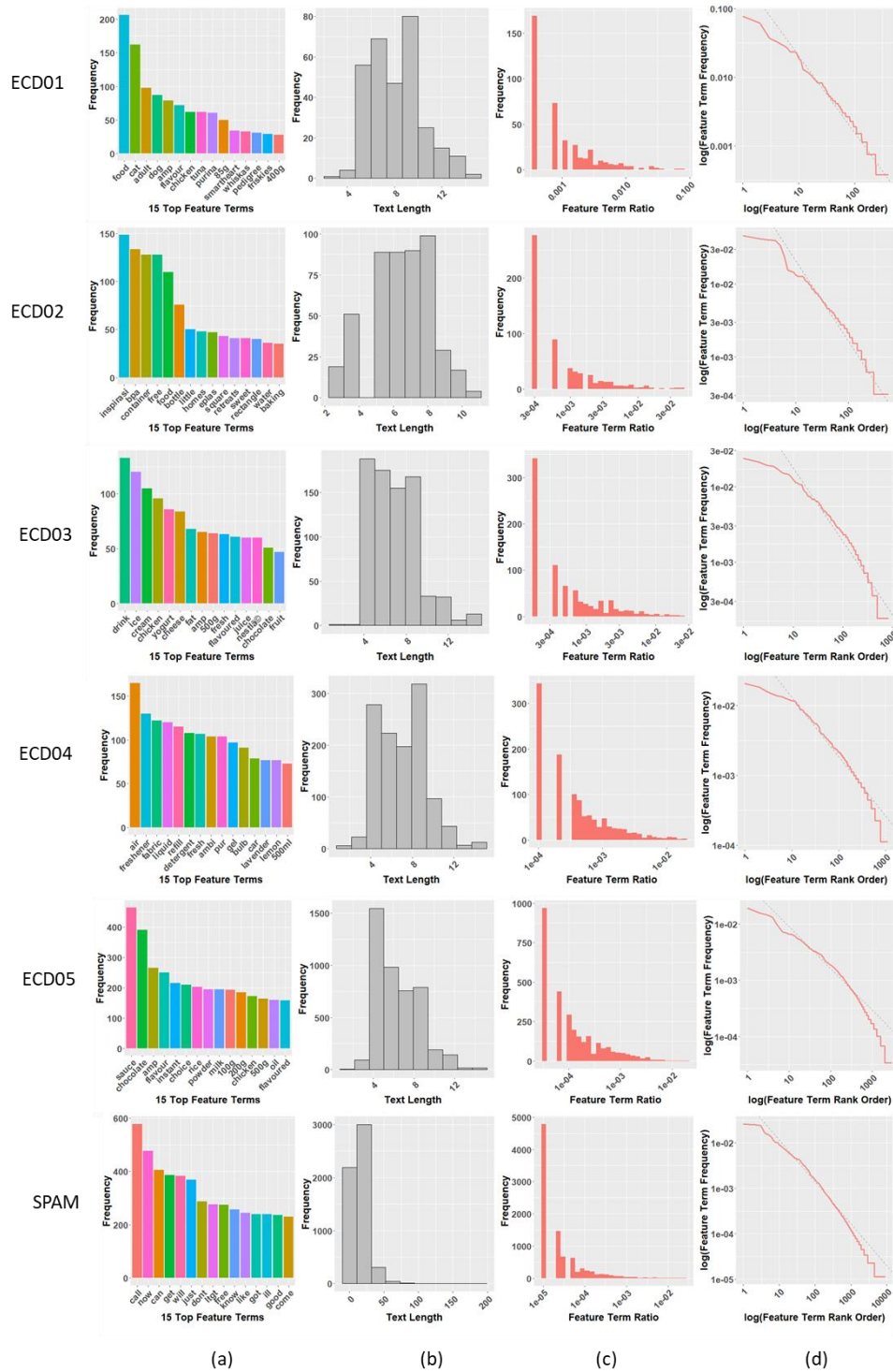


Fig. 4. Text characteristics and feature term distribution according to (a) 15 Top features, (b) Text length plot, (c) Term frequency distribution plot, and (d) Zipf's law distribution plot.

However, the curve falling under the fitted lines depicted features with high frequencies tend to be lower than predicted by their rank relative to Zipf's law. Natural language text distributions typically have similar overall patterns of a few very high-frequency types and long tails of infrequent words. The difference can be spotted through detailed observation of specific inconsistent parts in a frequency distribution plot. For

example, ECD01 and ECD02 may imply the same frequency distribution plot, but according to Zipf's law, the distribution varies, especially in explaining the features among low ranks with high frequencies. Hence, each e-commerce product dataset implied typical text distributions, yet they may encounter different classification performance results.

On the other hand, there is a noticeable difference between text characteristics for the SPAM dataset and e-commerce product datasets. The former dataset showed a right-skewed text length distribution because messages have longer texts than e-commerce product descriptions. Meanwhile, the frequency distribution plot illustrated that the frequency curve decreased more steeply and quickly flattered than frequency distribution plots for e-commerce product datasets. This pattern implied that many features in the dataset might not be frequently used. Some of the features only occurred once when processing text from messages.

In addition, Zipf's law distribution plot for the SPAM dataset closely follows the fitted line. The model predicts a very rapid decrease in frequency among the most frequent words, which becomes slower as the rank grows, leaving very long tails of words with similar low frequencies. Contrary, e-commerce product descriptions tend to utilize similar features across different categories and, at the same time, use particular features to describe products in a category. Hence, the text distributions for e-commerce products differed from the SPAM dataset. The study utilized both datasets to show the effectiveness of the proposed model.

B. Feature Reduction

Each dataset had been through all the pre-processing data procedures. Two feature extraction techniques, i.e., unigram and bigram, are used to extract the features. Then, the features from each set are selected using a correlation-based feature selection (CFS). It is a well-known filter method widely used in previous studies [10]. The features were also chosen using feature substitution by Latent Dirichlet Allocation (FS-LDA) with 10%, 20%, 30%, 40%, and 50% of each class's highest probability features. Table III shows the number of features used as the input data for HMM using different feature extraction techniques and feature substitution involvement in the model.

TABLE III. NUMBER OF FEATURES FOR EACH DATASET

Feature Extraction	Data	CFS	FS-LDA				
			10%	20%	30%	40%	50%
Unigram	ECD01	304	250	194	202	220	224
	ECD02	461	383	301	275	325	345
	ECD03	656	497	508	524	529	561
	ECD04	941	919	914	901	887	860
	ECD05	2630	2054	2189	2310	2393	2443
	SPAM	5903	4788	3641	3210	3847	4171
Bigram	ECD01	734	702	639	638	653	656
	ECD02	1072	1027	962	921	961	943
	ECD03	1934	1864	1832	1781	1738	1788
	ECD04	2789	2781	2779	2772	2756	2659
	ECD05	10852	10554	10407	10333	10341	10246
	SPAM	30349	29299	27531	26624	27102	26981

Fig. 5 and Fig. 6 show the number of features for some datasets does not decrease with the percentage increment of features from the highest probability in each class. The selected features to be substituted differ for each percentage where the overlap features are not replaced. The feature substitution by 10% shows features decreasing regardless of any datasets used

in the study. Then, the increment of 20% shows irregular patterns in unigram representation models.

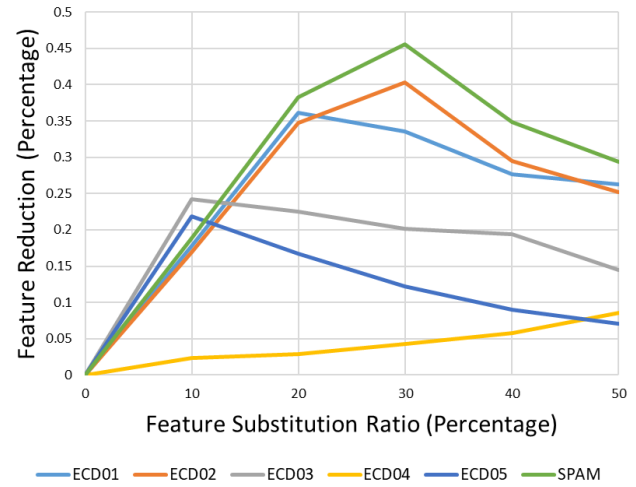


Fig. 5. Unigram feature reduction percentage for each dataset.

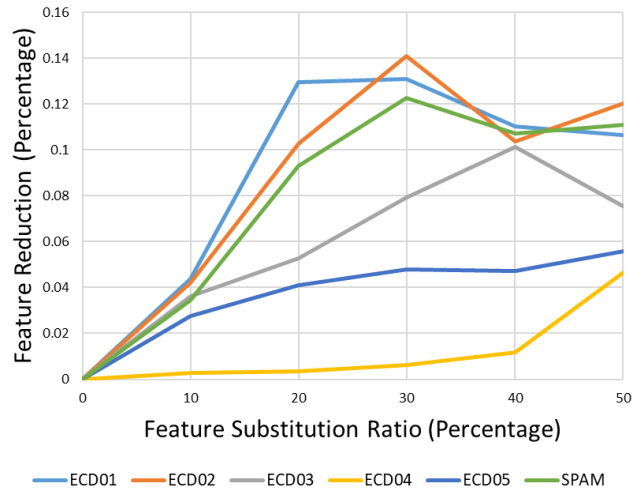


Fig. 6. Bigram feature reduction percentage for each dataset.

The feature number for ECD03 and ECD05 is greater than the feature number reduced by the 10% FS-LDA for each dataset. The irregular pattern for bigram models is only noticeable when the feature substitution by 30% is applied to the datasets. All datasets using the bigram model for feature extraction showed a lower performance increase than the unigram model. When using the bigram model, features extracted from a dataset become more specific, and each feature's representativeness differs from the unigram model. The same feature occurrences decrease drastically with the increase of unique features through the Bigram model. The inclusion of various features with minimal occurrences leads to poor LDA estimation on features belonging to particularly one hidden topic. Hence, the feature reduction percentage becomes smaller than expected while not being able to increase the model performance efficiently.

Nonetheless, using FS-LDA in preparation for classifying data using HMM did not jeopardize the model performance.

Regarding data reduction consistency, 10% of each topic's highest probabilities of non-overlap words seems like a good percentage to be used in general. However, the feature sets' performance was analyzed to prove that the proposed model is useful for reducing data dimensionality while improving a classifier's accuracy.

V. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the results of the experiments conducted and discusses the findings in the context of the proposed framework. The analysis evaluates the effectiveness of the feature substitution technique using FS-LDA in reducing dimensionality and its impact on text classification performance.

A. Feature Reduction

This study utilized two performance measurements, namely macro F1-score and weighted F1-score. The micro F1-score is not used in the study because all classification decisions in the dataset are considered without class discrimination when using this approach. Contrary, the macro F1-score is computed for each class within the dataset. Its average score calculation is based on the overall classes. In this way, class distributions in the training set are disregarded, and equal weight is assigned to each class. The formulas are presented in Eq. (3) - Eq. (7). S is the set of classes or states, TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

Meanwhile, the weighted F1-score is represented because this approach considers class imbalance [1]. Hence, the study observed the difference when the average score calculation for macro F1-score is based on each class's weight. The formula for the weighted F1-score is presented in Eq. (8).

$$accuracy_s = \frac{TP_s + TN_s}{TP_s + FP_s + TN_s + FN_s} \quad (3)$$

$$precision_s = \frac{TP_s}{TP_s + FP_s} \quad (4)$$

$$recall_s = \frac{TP_s}{TP_s + FN_s} \quad (5)$$

$$f_s = 2 \cdot \frac{precision_s \times recall_s}{(precision_s + recall_s)} \quad (6)$$

$$macro\ F1 - Score = \frac{\sum_{s \in S} f_s}{size\ (dataset)} \quad (7)$$

$$weighted\ F1 - Score = \frac{\sum_{s \in S} f_s \times size(s)}{size\ (dataset)} \quad (8)$$

B. Results for E-Commerce Product Data

The proposed model's effectiveness (FS-LDA) was observed based on its performance in classifying five e-commerce product data. The data involved binary and multiclass classification using HMM. Table IV presents the macro F1-scores for these datasets. According to the results, the unigram model application for extracting the features enhanced the HMM performance compared to the bigram model regardless of the feature substitution existence. The macro F1-

score for HMM with correlation-feature selection (CFS) seemed to increase when substituting 10% and 20% of the ECD01 and ECD04 data features with the unigram model. Meanwhile, the feature substitution worked the best when using 10% FS-LDA for ECD02, ECD03, and ECD05 data with the unigram model.

TABLE IV. MACRO F1-SCORES FOR E-COMMERCE DATASETS USING HMM

Feature Extraction	Data	CFS	FS-LDA				
			10%	20%	30%	40%	50%
Unigram	ECD01	0.6346	0.7866	0.7366	0.6947	0.6913	0.6531
	ECD02	0.8227	0.8695	0.8679	0.8632	0.8643	0.8437
	ECD03	0.6685	0.7402	0.7212	0.6733	0.6690	0.6469
	ECD04	0.6431	0.6454	0.6470	0.6449	0.6279	0.5994
	ECD05	0.5421	0.5935	0.5503	0.5259	0.5225	0.5097
Bigram	ECD01	0.4236	0.4650	0.4630	0.4306	0.4489	0.4560
	ECD02	0.4940	0.5038	0.4932	0.4928	0.4940	0.4940
	ECD03	0.2900	0.3663	0.3368	0.3573	0.3100	0.3412
	ECD04	0.2892	0.2902	0.2913	0.2909	0.2903	0.2985
	ECD05	0.2748	0.2856	0.2942	0.3022	0.2979	0.3053

On the other hand, Table V shows the weighted F1-scores for e-commerce product data. Like the macro F1-scores results, HMM with the unigram model was preferable rather than the bigram model to extract features for classifying these data. The HMM model for each data was similar to results obtained using macro F1-scores. However, the only difference is that the weighted F1-scores produced higher scores than macro F1-scores. A macro F1-score is most useful if there are many classes in the data and the researchers are interested in the average F1-score for each class.

Meanwhile, weighted F1-scores are influenced by the proportion for each class in the dataset. The score works well for observing the dataset's classification performance for unequal classes. Even though this score provides an alternative score for imbalanced dataset performance, a large weighted F1-score might be slightly misleading for a highly imbalanced dataset because the majority class overly influences it.

For example, the macro F1-score for ECD02 using CFS and 10% FS-LDA of the bigram HMM model was 0.5038 compared to the weighted F1-score value of 0.7805. The proportion of classes in Table II for ECD02 indicated that the dataset consists of 83.57% product descriptions for the cooking and dining category and only 16.43% product descriptions for party accessories. Hence, a noticeable difference in these two F1-scores was due to a highly imbalanced dataset. The inclusion of both scores was to observe the impact of the imbalanced dataset towards F1-scores as most of the datasets in the study are imbalanced datasets. However, both scores are equally acceptable according to the final goals of the study. The proposed feature substitution technique improves HMM performance according to both macro and weighted F1-scores.

TABLE V. WEIGHTED F1-SCORES FOR E-COMMERCE DATASETS USING HMM

Table with 8 columns: Feature Extraction, Data, CFS, and FS-LDA (10%, 20%, 30%, 40%, 50%). Rows include Unigram and Bigram models for datasets ECD01-ECD05.

The percentage of feature substitution that worked best for each dataset differed due to their text characteristics and distributions. The results encountered two situations: the HMM model performance suddenly dropped at a certain percentage of FS-LDA, or the model performance did not show any promising result throughout the FS-LDA.

Meanwhile, the second situation can be described through model performance for ECD04. The model did not show promising performance improvement regardless of any percentage of FS-LDA due to a highly similar number of features existing in the dataset, as shown in Fig. 5, compared to other e-commerce datasets.

Despite showing macro or weighted F1-scores, Table VI presents the straight-forward model performance evaluation using model accuracy between several text classifiers, including HMM, HMM with 10% FS-LDA, Naïve Bayes, and Support Vector Machine.

Support Vector Machine and Naïve Bayes outperformed the proposed model performance for ECD01. These two classifiers are known for their excellent performances in solving binary

classification problems without interfering with uncommon feature distributions such as ECD01. When dealing with data such as ECD01, the proposed model seemed to improve the performance of standard HMM. However, combining improvisation from the feature substitution technique presented in the study with enhancing theory in developing a better HMM model may outperform the other two classifiers.

TABLE VI. ACCURACY RATE COMPARISON BETWEEN HMM, HMM (10% FS-LDA), NAÏVE BAYES AND SUPPORT VECTOR MACHINE

Table with 6 columns: Feature Extraction, Data, HMM, HMM (10% FS-LDA), Naive Bayes, Support Vector Machine. Rows include Unigram and Bigram models for datasets ECD01-ECD05.

C. Results for Spam Data

The proposed technique presented in this paper can be applied to other kinds of text data. The study utilized a well-known benchmark data, SMS spam data collection, to evaluate its performance in a different text data application.

TABLE VII. PERFORMANCE RESULTS FOR SPAM DATASET

Table with 8 columns: Metrics, Feature Extraction, CFS, and FS-LDA (10%, 20%, 30%, 40%, 50%). Rows include Macro F1-score, Weighted F1-score, and Accuracy for Unigram and Bigram models.

The HMM model performed the best by securing an accuracy of 90.56% to classify spam and ham SMS considering 10% FS-LDA as the optimum HMM model across different datasets. Both model precision and recall increased when applying the proposed technique. This improvement leads to finer F1-scores for the HMM. The result implied the effectiveness of FS-LDA not only for e-commerce product classification but also for spam detection. The model accuracy was superior compared to the LDA result obtained by Nagwani and Shara [42]. The proposed model outperformed the Naïve Bayes model.

However, when the proposed model is compared with J48 and multi-layer perceptron classifiers, it seems not to be better, as shown in Renuka et al. [43]. Although, there is a slight difference between the accuracy of these models and the proposed model. The HMM model is a reliable and good classifier for classifying text datasets, especially when applying the FS-LDA technique. An HMM model itself may need some modification to achieve better performance. Yet, this feature substitution technique using the LDA model proposed in this study is relatively helpful, simple, and easy to implement. Hence, it is beneficial for commercial uses related to text classification.

VI. CONCLUSION

This study introduces FS-LDA, a novel technique integrating LDA into the preprocessing phase of text data classification. The results highlight the effectiveness of FS-LDA when applied with HMMs, demonstrating superior performance compared to using feature selection alone. By substituting non-overlapping words in high-probability topic groups identified by LDA, FS-LDA significantly reduces data dimensionality while enhancing the accuracy and efficiency of classification models.

The study also highlights the advantage of using a unigram model over a bigram model for feature extraction. Unigrams simplify the feature space while retaining important semantic information, making them more effective for accurate classification. This aligns with findings that simpler models often perform better in text classification by focusing on key features efficiently.

Overall, the integration of FS-LDA with HMMs and the adoption of unigram-based feature extraction represent robust strategies for improving the practical utility of text classification systems, paving the way for enhanced performance in various applications such as e-commerce product classification, spam detection, sentiment analysis, and document categorization. However, the fixed substitution percentage of FS-LDA could be tested on more datasets or through simulations to confirm its reliability. While this study focused on HMMs, trying FS-LDA with other machine learning models could offer more insights.

ACKNOWLEDGMENT

This research was financially supported by Universiti Teknologi MARA and the Institute of Postgraduate Studies, UiTM. It forms part of a study under the Grant Scheme (FRGS/1/2018/STG06/UITM/01/1). The authors would like to

express their deepest gratitude to the Department of Statistics Malaysia for their knowledge and data support.

REFERENCES

- [1] D. D. Le Nguyen, Y. C. Huang, and Y. C. Chang, "Discriminative features fusion with bert for social sentiment analysis," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12144 LNAI, pp. 30–35, 2020, doi: 10.1007/978-3-030-55789-8_3.
- [2] B. S. Kumar and V. Ravi, "LDA based feature selection for document clustering," *ACM Int. Conf. Proceeding Ser.*, pp. 125–130, 2017, doi: 10.1145/3140107.3140129.
- [3] S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao, "Deep Learning Based Text Classification: A Comprehensive Review," *arXiv*, vol. 1, no. 1, pp. 1–43, 2020, [Online]. Available: <http://arxiv.org/abs/2004.03705>
- [4] M. Gupta, R. Kumar, C. Ved, and S. Taneja, "Hybrid deep learning approach for product categorization in e-commerce," *AIP Conf. Proc.*, vol. 3072, no. 1, 2024, doi: 10.1063/5.0198666.
- [5] D. Pakpahan, V. Siallagan, and S. Siregar, "Classification of E-Commerce Product Descriptions with The Tf-Idf and Svm Methods," *Sinkron*, vol. 8, no. 4, pp. 2130–2137, 2023, doi: 10.33395/sinkron.v8i4.12779.
- [6] M. T. Alrefaie, N. E. Morsy, and N. Samir, "Exploring Tokenization Strategies and Vocabulary Sizes for Enhanced Arabic Language Models," *Mar. 2024*, Accessed: May 02, 2024. [Online]. Available: <http://arxiv.org/abs/2403.11130>
- [7] M. Adnan, M. O. Imam, M. F. Javed, and I. Murtza, "Improving spam email classification accuracy using ensemble techniques: a stacking approach," *Int. J. Inf. Secur.*, vol. 23, no. 1, pp. 505–517, 2024, doi: 10.1007/s10207-023-00756-1.
- [8] J. Singh, D. Pandey, and A. K. Singh, "Event detection from real-time twitter streaming data using community detection algorithm," *Multimed. Tools Appl.*, vol. 83, no. 8, 2024, doi: 10.1007/s11042-023-16263-3.
- [9] A. M. Ali et al., "Explainable Machine Learning Approach for Hepatitis C Diagnosis Using SFS Feature Selection," *Machines*, vol. 11, no. 3, 2023, doi: 10.3390/machines11030391.
- [10] N. M. N. Mathivanan, N. A. M. Ghani, and R. M. Janor, "Improving Classification Accuracy Using Clustering Technique," *Bull. Electr. Eng. Informatics*, vol. 7, no. 3, pp. 465–470, 2018, doi: 10.11591/eei.v7i3.1272.
- [11] N. M. N. Mathivanan, N. A. M. Ghani, and R. M. Janor, "Performance analysis of supervised learning models for product title classification," *IAES Int. J. Artif. Intell.*, vol. 8, no. 3, pp. 299–306, 2019, doi: 10.11591/ijai.v8.i3.pp299-306.
- [12] H. Liu and H. Motoda, *Computational Methods of Feature Selection*, vol. 198, no. 1. 2007. doi: 10.1201/9781584888796.
- [13] N. M. N. Mathivanan, N. A. M. Ghani, and R. M. Janor, "A comparative study on dimensionality reduction between principal component analysis and k-means clustering," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 16, no. 2, pp. 752–758, 2019, doi: 10.11591/ijeecs.v16.i2.pp752-758.
- [14] J. Meng, H. Lin, and Y. Yu, "A two-stage feature selection method for text categorization," *Comput. Math. with Appl.*, vol. 62, no. 7, pp. 2793–2800, 2011, doi: 10.1016/j.camwa.2011.07.045.
- [15] D. Jain and V. Singh, "An Efficient Hybrid Feature Selection model for Dimensionality Reduction," *Procedia Comput. Sci.*, vol. 132, no. Iccids, pp. 333–341, 2018, doi: 10.1016/j.procs.2018.05.188.
- [16] M. Shao and L. Qin, "Text Similarity Computing Based on LDA Topic Model and Word Co-occurrence," no. Sekeie, pp. 199–203, 2014, doi: 10.2991/sekeie-14.2014.47.
- [17] N. Omar and Q. Al-Tashi, "Arabic nested noun compound extraction based on linguistic features and statistical measures," *GEMA Online J. Lang. Stud.*, vol. 18, no. 2, 2018, doi: 10.17576/gema-2018-1802-07.
- [18] M. S. Yang and K. P. Sinaga, "Collaborative feature-weighted multi-view fuzzy c-means clustering," *Pattern Recognit.*, vol. 119, 2021, doi: 10.1016/j.patcog.2021.108064.
- [19] R. Al-Wajih, S. J. Abdulkadir, N. Aziz, Q. Al-Tashi, and N. Talpur, "Hybrid binary grey Wolf with Harris hawks optimizer for feature selection," *IEEE Access*, vol. 9, 2021, doi:

- 10.1109/ACCESS.2021.3060096.
- [20] A. Christy, A. Praveena, and J. Shabu, "A hybrid model for topic modeling using latent dirichlet allocation and feature selection method," *J. Comput. Theor. Nanosci.*, vol. 16, no. 8, 2019, doi: 10.1166/jctn.2019.8234.
- [21] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.*, vol. 1, no. 4–5, 2003.
- [22] J. S. Su, B. F. Zhang, and X. Xu, "Advances in machine learning based text categorization," *Ruan Jian Xue Bao/Journal Softw.*, vol. 17, no. 9, 2006, doi: 10.1360/jos171848.
- [23] Q. Wang, R. Peng, J. Wang, Y. Xie, and Y. Zhou, "Research on Text Classification Method of LDA- SVM Based on PSO optimization," in *Proceedings - 2019 Chinese Automation Congress, CAC 2019*, 2019, doi: 10.1109/CAC48633.2019.8996952.
- [24] Y. Lu, Q. Mei, and C. X. Zhai, "Investigating task performance of probabilistic topic models: An empirical study of PLSA and LDA," *Inf. Retr. Boston.*, vol. 14, no. 2, 2011, doi: 10.1007/s10791-010-9141-9.
- [25] T. L. Griffiths and M. Steyvers, "Finding scientific topics," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, no. SUPPL. 1, 2004, doi: 10.1073/pnas.0307752101.
- [26] O. Jin, N. N. Liu, K. Zhao, Y. Yu, and Q. Yang, "Transferring topical knowledge from auxiliary long texts for short text clustering," in *International Conference on Information and Knowledge Management, Proceedings*, 2011, doi: 10.1145/2063576.2063689.
- [27] D. Wang, M. Thint, and A. Al-Rubaie, "Semi-supervised latent Dirichlet allocation and its application for document classification," in *Proceedings of the 2012 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology Workshops, WI-IAT 2012*, 2012, doi: 10.1109/WI-IAT.2012.211.
- [28] X. Cheng, X. Yan, Y. Lan, and J. Guo, "BTM: Topic modeling over short texts," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 12, 2014, doi: 10.1109/TKDE.2014.2313872.
- [29] X. Quan, C. Kit, Y. Ge, and S. J. Pan, "Short and sparse text topic modeling via self-aggregation," in *IJCAI International Joint Conference on Artificial Intelligence*, 2015.
- [30] D. Kim, D. Seo, S. Cho, and P. Kang, "Multi-co-training for document classification using various document representations: TF-IDF, LDA, and Doc2Vec," *Inf. Sci. (Ny.)*, vol. 477, 2019, doi: 10.1016/j.ins.2018.10.006.
- [31] W. Zhou, H. Wang, H. Sun, and T. Sun, "A method of short text representation based on the feature probability embedded vector," *Sensors (Switzerland)*, vol. 19, no. 17, 2019, doi: 10.3390/s19173728.
- [32] N. Pröllochs and S. Feuerriegel, "Business analytics for strategic management: Identifying and assessing corporate challenges via topic modeling," *Inf. Manag.*, vol. 57, no. 1, 2020, doi: 10.1016/j.im.2018.05.003.
- [33] P. Celard, A. S. Vieira, E. L. Iglesias, and L. Borrajo, "LDA filter: A Latent Dirichlet Allocation preprocess method for Weka," *PLoS One*, vol. 15, no. 11 November, pp. 1–14, 2020, doi: 10.1371/journal.pone.0241701.
- [34] G. R. Venkataraman et al., "FasTag: Automatic text classification of unstructured medical narratives," *PLoS One*, vol. 15, no. 6 June, pp. 1–18, 2020, doi: 10.1371/journal.pone.0234647.
- [35] D. Freitag and A. K. McCallum, "Information Extraction with HMMs and Shrinkage," in *Proceedings of Workshop on Machine Learning for Information Extraction*, 1999.
- [36] D. R. H. Miller, T. Leek, and R. M. Schwartz, "A hidden Markov model information retrieval system," *Proc. 22nd Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retrieval, SIGIR 1999*, pp. 214–221, 1999, doi: 10.1145/312624.312680.
- [37] N. M. N. Mathivanan, N. A. M. Ghani, and R. M. Janor, "Text classification of E-commerce product via Hidden Markov model," *Adv. Technol. Ind. through Intell. Softw. Methodol. Tools Tech. Proc. 18th SoMeT_19*, vol. 318, pp. 310–318, 2019, doi: 10.3233/FAIA190058.
- [38] M. S. Khorsheed, "Diacritizing Arabic text using a single hidden markov model," *IEEE Access*, vol. 6, pp. 36522–36529, 2018, doi: 10.1109/ACCESS.2018.2852619.
- [39] W. De Gruyter, *Corpus Linguistics: An International Handbook*, Volume 2, vol. 2, no. 1. 2008.
- [40] P. S. Florence and G. K. Zipf, "Human Behaviour and the Principle of Least Effort.," *Econ. J.*, vol. 60, no. 240, 1950, doi: 10.2307/2226729.
- [41] M. Zrigui, R. Ayadi, M. Mars, and M. Maraoui, "Based on Latent Dirichlet Allocation," *J. Comput. Inf. Syst.*, vol. 20, no. 2, pp. 125–140, 2012.
- [42] N. K. Nagwani and A. Sharaff, "SMS spam filtering and thread identification using bi-level text classification and clustering techniques," *J. Inf. Sci.*, vol. 43, no. 1, pp. 75–87, 2017, doi: 10.1177/0165551515616310.
- [43] D. Karthika Renuka, T. Hamsapriya, M. Raja Chakkaravarthi, and P. Lakshmi Surya, "Spam classification based on supervised learning using machine learning techniques," in *Proceedings of 2011 International Conference on Process Automation, Control and Computing, PACC 2011*, 2011, doi: 10.1109/PACC.2011.5979035.

Multilabel Classification of Bilingual Patents Using OneVsRestClassifier: A Semiautomated Approach

Slamet Widodo¹, Ermatita^{2*}, Deris Stiawan³

Doctoral Program in Engineering Science, University Sriwijaya Palembang, Indonesia^{1, 2, 3}

Department of Computer Engineering, Politeknik Negeri Sriwijaya Palembang, Indonesia¹

Faculty of Computer Science, University Sriwijaya Palembang, Indonesia^{2, 3}

Abstract—In response to the increasing complexity and volume of patent applications, this research introduces a semiautomated system to streamline the literature review process for Indonesian patent data. The proposed system employs a synthesis of multilabel classification techniques based on natural language processing (NLP) algorithms. This methodology focuses on developing an iterative and modular system, with each step visualised in detailed flowcharts. The system design incorporates data collection and preprocessing, multilabel classification model development, model optimisation, query and prediction, and results presentation modules. Experimental results demonstrate the promising potential of the multilabel classification model, achieving a micro F1 score of 0.6723 and a macro F1 score of 0.6009. The OneVsRestClassifier model with LinearSVC as the base classifier shows reasonably good performance in handling a bilingual dataset comprising 15,097 patent documents. The optimal model configuration uses TfidfVectorizer with 20,000 features, including bigrams, and an optimal C parameter of 0.1 for LinearSVC. Performance analysis reveals variations across IPC classes, indicating areas for further improvement. The discussion highlights the implications of the proposed system for researchers, patent examiners and industry professionals by facilitating efficient searches within patent databases. This study acknowledges the potential of semiautomated systems to enhance the efficiency of patent analysis while emphasising the need for further research to address identified challenges, such as class imbalance and performance variations across patent categories. This research paves the way for further developments in the field of automated patent classification, aiming to improve efficiency and accuracy in international patent systems while recognising the crucial role of human experts in the patent classification process.

Keywords—Multilabel patent classification; Natural Language Processing (NLP); OneVsRestClassifier; TF-IDF vectorisation; bilingual patent analysis

I. INTRODUCTION

At present, conducting manual patent literature reviews involves a relatively challenging level of difficulty. The continuous influx of submissions adds complexity, which demands efficient analysis for intellectual property management and strategic innovation tracking [1]. The intricate technical and legal language in these documents also contributes to the complexity of manual processing [2]. Traditional methods, although widely used, are time consuming, resource intensive and prone to human error and bias, which can lead to inconsistent and unreliable results [1], [2].

While recent advances in natural language processing (NLP) have automated aspects of patent analysis [3], [4], critical gaps remain. First, most systems have focused on monolingual datasets (e.g. English only [5] or Indonesian only [6]), neglecting the bilingual nature of patents in countries such as Indonesia, where filings combine local and international languages. Second, existing methods have often failed to address class imbalance in the International Patent Classification (IPC) system, leading to poor performance in underrepresented technology categories (e.g. Y02A) [7]. Third, few studies have integrated local patent databases (e.g. Indonesian Patent Database) with global repositories (e.g. Google Patents), limiting their applicability to multinational innovation ecosystems. Our work bridges these gaps by proposing a bilingual framework that combines Indonesian and English patents, addresses class imbalance through weighted learning and validates utility across diverse IPC categories.

This study addresses the following research questions:

RQ1: How effective is the OneVsRestClassifier with LinearSVC for bilingual (Indonesian–English) patent classification compared to monolingual approaches?

RQ2: What feature engineering strategies (e.g. TF-IDF with bigrams and class weighting) optimise multilabel IPC classification performance in imbalanced datasets?

RQ3: How does class imbalance affect model performance across different IPC categories, and what mitigation strategies are most effective?

Recognising these limitations, this research seeks to refine the semiautomated process for reviewing Indonesian patent literature by using data from local and international repositories. Our approach uses web-scraping techniques to obtain datasets, followed by preprocessing to clean and structure the data for processing using machine learning algorithms. We use the IPC to train multilabel classification models, which allow for categorisation that represents the diverse nature of patent data [3], [4].

The proposed solution links multilabel classification algorithms to increase efficiency and reduce the resources required for a comprehensive review [3], [4]. This process aims to optimise patent analysis by leveraging computational power. The application of these techniques is intended to address the vast amount of data and complex patent language. By using an approach based on machine learning, the proposed system

seeks to simplify this complex task and make it more manageable [5].

The significance of this research is its significant potential to develop and advance patent classification techniques by substantially improving the accuracy and precision of analysis, as well as accelerating systematic, structured and data-driven decision-making processes [1], [4]. This research has high value and strong relevance to patent examiners, research and development institutions and companies that rely heavily on accurate and efficient patent analysis, with much broader implications for innovation tracking, in-depth competitive analysis and future technology forecasting [6], [7]. Ultimately, this study aims to lay a strong foundation and solid groundwork for visionary strategic planning and well-informed policymaking in the dynamic field of intellectual property [8].

Paper Overview. The remainder of this paper is organised as follows. Section II reviews key studies on multilabel patent classification, emphasising the bilingual and imbalanced data contexts. Section III outlines the proposed methodology, detailing the dataset collection, feature engineering and model training processes. Section IV presents the experimental setup, along with the results and discussion of the findings. Finally, Section V concludes the paper, summarising the main contributions, acknowledging current limitations and suggesting avenues for future research.

II. RELATED WORK

The increasing volume of patent applications worldwide has triggered a critical need for advancements in patent analysis methodologies. Traditional manual reviews, characterised by their meticulous yet cumbersome nature, have become unsustainable in the face of rapid technological innovation; the corresponding increase in intellectual property documentation [1] highlights the intrinsic limitations of manual reviews, particularly their vulnerability to human error and the inherent subjectivity in interpreting complex legal and technical terms.

Patent classification using the k-nearest neighbours (KNN) and fastText classifier algorithms individually performs worse than when they are combined by a meta-classifier. The former approach is based on a linguistically supported KNN algorithm using a method of searching for topically similar documents based on comparisons of lexical descriptor vectors. The latter approach employs fastText based on word embeddings, in which sentence (or document) vectors are obtained by averaging n-gram embeddings, and then vectors are used as features in multinomial logistic regression [9].

To address challenges, the field of NLP has emerged as a beacon of innovation. The ability of NLP to parse and interpret complex language structures makes it a powerful tool for the semiautomated analysis of patent documents. The study in [2] underlined the transformative impact of NLP in the domain of summarisation, simplification and generation of patent texts, indicating an urgent need for research specifically tailored to the nuanced demands of patent documentation.

At the forefront of this domain, multilabel classification has been identified as a crucial component for effective patent categorisation, often encapsulating the convergence of various

technological domains. The complexity involved in accurately classifying multifaceted documents is further exacerbated in fields such as artificial intelligence, in which the intersection of technology and legal language demands sophisticated computational techniques for precise analysis [3] [4].

The integration of NLP techniques into semiautomated systems for patent analysis signifies a substantial leap from manual review processes, promising enhanced accuracy and efficiency in patent analysis. However, this integration is not without challenges. The need for comprehensive and well-annotated datasets for training and testing NLP models remains an ongoing hurdle, alongside the development of models that can adeptly navigate the intricacies of patent language and accurately reflect the evolving landscape of technological innovation [6], [10].

Three critical gaps persist in the literature, which are as follows:

1) *Monolingual bias:* Most studies [9], [10] have focused on monolingual patent datasets, overlooking the bilingual complexity inherent in countries such as Indonesia. For instance, [9] combined KNN and fastText but only tested on English patents, neglecting cross-lingual term alignment.

2) *Class imbalance:* Prior works [3], [11] have often assumed balanced IPC label distributions, leading to poor performance in rare categories (e.g. Y02A). For example, [11] reported high accuracy overall but did not address label skew.

3) *Local–global integration:* Existing frameworks [12], [13] have rarely combined local patent databases (e.g. Indonesian Patent Database) with international repositories (e.g. Google Patents), limiting their ability to capture region-specific innovations.

Our work directly addresses these gaps by (1) designing a bilingual (Indonesian–English) classification pipeline, (2) optimising for class imbalance via `class_weight='balanced'` in LinearSVC and (3) integrating local and global patent data to enhance coverage and relevance.

As the discipline evolves, ethical considerations and data sharing become increasingly important. Unbiased data representation in training sets is crucial to mitigating biases that might be perpetuated in patent analysis. Additionally, sharing open-source tools and datasets to catalyse innovation through collaborative efforts underscores the importance of interdisciplinary cooperation in advancing the capabilities of NLP systems in patent informatics. [11], [8] emphasised the importance of collaboration in data sharing and of ethical implications in developing NLP tools for scientific research.

Natural language processing technology has made significant strides in transforming patent informatics, and the field is ripe for further exploration and development. The research in [12], [13] provided evidence of the effectiveness of semiautomated approaches in machine learning-based literature reviews, which can be applied in patent data analysis. Further research is needed to refine NLP models, enhance the understanding and processing of patent data and drive systematic and data-driven approaches to intellectual property management [10], [14]. One approach to handling NLP is the

classification chain (CC), which links these binary classifiers in a certain sequential order so that each classifier includes labels predicted by the previous classifier as additional features. Despite the simplicity of this approach, recent comprehensive empirical studies have shown that CC is among the best-performing algorithms [15].

III. METHODOLOGY

A. Dataset

This study combines 7,298 patents from the Indonesian Patent Database and 7,801 patents from Google Patents, forming a bilingual corpus of 15,097 documents. This hybrid dataset was strategically selected to address the following three critical requirements for robust multilabel patent classification:

1) *Bilingual representation*: The Indonesian Patent Database provides local language coverage (Indonesian), while Google Patents ensures international relevance (English). This combination reflects real-world patent ecosystems in multilingual jurisdictions, such as Indonesia.

2) *Class diversity*: Google Patents broadens the scope of IPC codes beyond region-specific innovations, ensuring the coverage of emerging global technologies (e.g. Y02A for climate adaptation).

3) *Imbalanced IPC mitigation*: Merging datasets diversifies label distributions, reducing bias towards dominant classes (e.g. A61K) while retaining rare categories for comprehensive analysis.

The dataset includes four key features: patent_id, patent_title, patent_abstract and ipc_code. Table I summarises the dataset composition.

TABLE I. DATASET OF INDOONESIAN PATENTS AND GOOGLE PATENTS

No	Dataset	Jumlah Record
1	Indonesia_Patents	7298
2	Google_Patents	7801
	Total	15099

B. Proposed Framework

The framework depicted in Fig. 1 is the basis of this research. Data were taken from Google Patents and the Indonesian Patent Database [10]. We begin by collecting patent data from these two sources, ensuring a comprehensive dataset that covers various innovations. Once collected, these data are then preprocessed, which includes text cleaning, stopword removal and preparation for efficient machine learning classification [1], [15]–[17].

The research methodology is iterative and modular [18], focused on developing a semiautomated system for reviewing Indonesian patent data literature [19]. Each step is visualised in a detailed flowchart, which serves as a guide through various stages of data collection, processing and analysis. Fig. 2 explains the flowchart of this patent classification system research, which uses machine learning techniques to classify patent documents into IPC codes [9]. This system is designed to process and analyse patent data from Google Patents and the Indonesian Patent Database. The feature structure consists of

the ID as a unique patent identification, the patent title, the patent abstract or summary and related IPC codes. Next, data loading and cleaning are performed. The clean_text() function performs text cleaning by removing HTML tags and non-alphanumeric characters and digits and converting text to lowercase [20], [21]. Text processing involves tokenisation and stopword removal using a combination of English and Indonesian stopwords [1], [22], [23]. Feature engineering and data splitting combine datasets, convert IPC codes into multilabel formats and split data into training and testing sets. Model training and evaluation conduct the experiments with various parameter configurations, train the OneVsRestClassifier model with LinearSVC as the base classifier and calculates the evaluation metrics for each configuration. The OneVsRest (OVR) model can provide informative hidden representations for unknown examples, and in open-set classification scenarios, the proposed probability model is better than modern approaches [15], [24].

This research begins by collecting patent data, followed by preprocessing procedures to prepare the data for classification. The processed data are then used to train and evaluate multilabel classification models, specifically the OneVsRestClassifier algorithm, to assign multiple IPC labels to each patent document [19]. This research also performs experiments to optimise the model by varying parameter values, such as n-gram range and maximum features for TF-IDF, as well as the C parameter for LinearSVC. The performance of each configuration is assessed using evaluation metrics, such as the F1 score (micro and macro), as well as cross-validation to determine the optimal model configuration [25], [26].

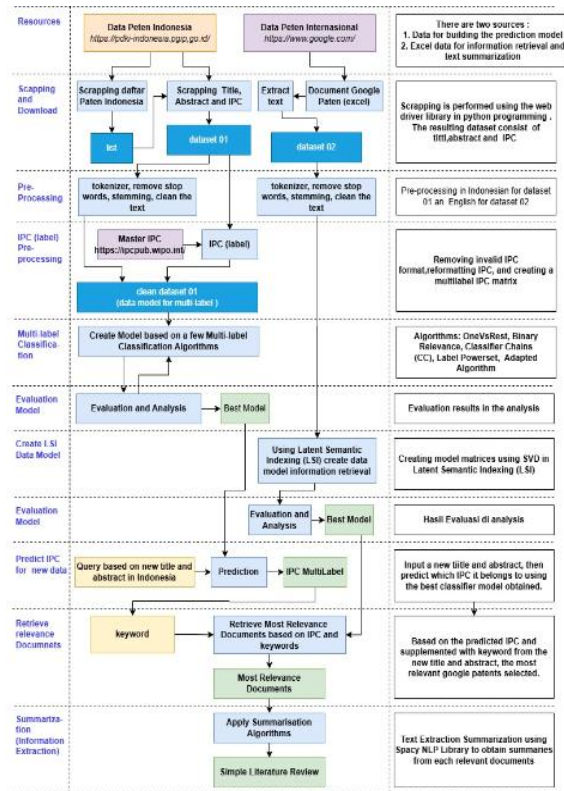


Fig. 1. Proposed framework architecture.

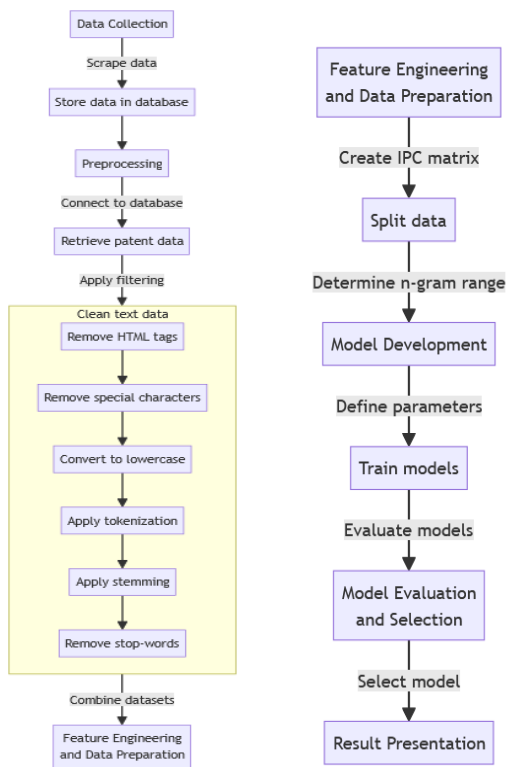


Fig. 2. Research flowchart.

C. OneVsRestClassifier

The OVR method is a strategy used in multiclass classification, in which separate binary classifiers are trained for each class to distinguish a particular class from all other classes [24], [27]. In this approach, for a particular class, samples belonging to it are treated as a positive class, and all other samples are treated as a negative class. This results in the need for only K binary classifiers for K classes, which is a smaller number than that in the one-versus-one method [28].

In this implementation, we use LinearSVC as the base classifier in the OVR framework. The LinearSVC configuration includes the following parameters. 1) `class_weight='balanced'` is used to address class imbalance by assigning appropriate weights to each class. 2) `max_iter=5000` increases the maximum number of iterations to ensure model convergence. 3) `dual=False` uses the primal formulation of SVM, which is more efficient for cases in which the number of samples is larger than the number of features. 4) `tol=1e-4` indicates tolerance for stopping criteria.

The main challenge with the OVR method is the imbalance between positive and negative classes, especially as the number of classes increases. This imbalance can lead to biased classifiers that favour the majority class, resulting in poor classification performance for minority classes [29]. To address this issue, we use the `class_weight='balanced'` parameter in LinearSVC. We also apply GridSearchCV to search for the optimal value of the C parameter in LinearSVC, with a range of values [0.01, 0.1, 1, 10]. The C parameter controls the trade-off between achieving a low margin and minimising classification errors.

To further optimise model performance, we apply threshold optimisation techniques. This process involves searching for the optimal threshold to convert the output of the decision function into binary predictions. The threshold is optimised in the range of 0.1 to 0.9 to obtain the best F1 score, allowing flexibility in balancing precision and recall [30], [31]. This approach allows the model to handle the complexity of multilabel classification in patent data effectively while maintaining computational efficiency and model interpretability.

D. Data Collection and Processing Module

The data collection and processing module is responsible for collecting and processing patent data from Google Patents and the Indonesian Patent Database, with a total of 15,099 patent documents. This process involves a series of comprehensive preprocessing steps. Text cleaning is performed by removing HTML tags and non-alphanumeric characters and digits, as well as converting all text to lowercase. Stopwords are removed using a combination of 936 English and Indonesian stopwords from NLTK. International Patent Classification codes are processed by extracting sections, classes and subclasses, as well as filtering codes with a minimum of 200 samples. The processed titles and abstracts are combined into a single 'preprocessed_text' field for further analysis. This approach ensures that the data used have been cleaned, standardised and optimised for multilabel classification, increasing the potential for model accuracy and reliability [25], [32].

E. Multilabel Classification Model Development Module

The multilabel classification model development module focuses on converting IPC codes into a multilabel format using MultiLabelBinarizer and developing classification models. The processed data are split into training and validation sets. Then, the TF-IDF vectoriser is used with the parameters `max_features=20000` and `ngram_range=(1, 2)` for feature extraction [9]. The main model used is OneVsRestClassifier with LinearSVC as the base classifier. LinearSVC is configured with `class_weight='balanced'`, `max_iter=5000`, `dual=False` and `tol=1e-4`. Cross-validation with GridSearchCV is used for hyperparameter optimisation, with the F1 score (micro and macro averages) as the main evaluation metric. This approach allows the model to effectively handle the complexity of multilabel classification in patent data while maintaining computational efficiency and model interpretability [30], [31].

F. Model Optimisation Module

The model optimisation module focuses on improving the performance of multilabel classification models through experiments with various parameter combinations [33]. This module uses GridSearchCV to search for the optimal value of the C parameter in LinearSVC, with a range of values [0.01, 0.1, 1, 10]. Additionally, threshold optimisation is performed to convert the output of the decision function into binary predictions, with the threshold optimised in the range of 0.1 to 0.9. Threefold cross-validation is used to assess the effectiveness of each configuration. Evaluation results, especially the F1 scores (micro and macro), are saved and analysed for each parameter combination. The optimal model

configuration is selected based on the balance between model performance and computational efficiency [24], [34]. This approach allows for better model adjustment to the specific characteristics of the patent dataset, thereby improving overall classification accuracy.

G. Feature Extraction Using TF-IDF

The method that determines how often each word appears in one document component, called term frequency (TF), and how rarely it occurs in all document components, called inverse document frequency (IDF), is the inverse of the TF document [12]. To calculate weights, the TF-IDF method combines two ideas: the frequency of a word appearing in a particular document and the inverse frequency of documents containing that word. The tf value is divided by the frequency of the most frequently occurring words in the document. This process ensures that the most frequently occurring words obtain the highest if value, which is 1, and that the least frequently occurring words obtain values between 0.5 and 1 [35].

$$if = 0,5 + 0,5 \times \frac{tf}{\max(tf)}. \quad (1)$$

Weighting is used with the TF-IDF formula in research conducted with the equation formula from several previous research sources [22].

$$W_{t,d} = TF_{t,d} \times IDF_{t,d} = TF_{t,d} \times \left(\log\left(\frac{N}{dft}\right)\right) \quad (2)$$

The TF-IDF formula is very important for document analysis because it gives higher values to words that appear frequently in one document but rarely appear in other documents. Eq. (2), representing the change in IDF using $\log(1 + N/dft)$, prevents division by zero problems or negative logarithms when dft approaches or equals N. This change ensures that the IDF weight remains well defined, even if a word appears in all documents (preventing IDF from becoming zero or negative), providing stability to TF-IDF weights in real applications.

$$W_{t,d} = TF_{t,d} \times IDF_{t,d} = TF_{t,d} \times \left(\log\left(1 + \frac{N}{dft}\right)\right). \quad (3)$$

In Eq. (3), which represents the change in IDF using $\log(1 + N/dft)$, prevents division by zero problems or negative logarithms when dft approaches or equals N. This change ensures that the IDF weight remains well defined, even if a word appears in all documents (preventing IDF from becoming zero or negative), providing stability to TF-IDF weights in real applications.

$$W_{t,d} = TF_{t,d} \times IDF_{t,d} = TF_{t,d} \times \left(\log\left(\frac{N}{1+dft}\right)\right). \quad (4)$$

To generate a new score, the code-mixed relevance score modifies the TF-IDF score, and weighting and normalisation are applied to obtain the final feature vector EF [36].

H. Model Evaluation

In the new implementation, model evaluation uses LinearSVC as the base classifier in the OneVsRestClassifier framework, replacing the previously used random forest. This method is effective for multilabel classification, in which each instance can have more than one label. Model evaluation is

performed using several main metrics, which are as follows. 1) The F1 score (micro and macro averages) is the harmonic mean of precision and recall, providing an overall picture of model performance. $F1 = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$. 2) The classification report provides a summary of the precision, recall and F1 scores for each class. 3) Threshold optimisation optimises the threshold to convert the output of the decision function into binary predictions [37].

The evaluation process also involves GridSearchCV for hyperparameter tuning, specifically the C parameter of LinearSVC. Threefold cross-validation is used to assess model reliability across different subsets of the data. The main evaluation metrics used are as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}, \quad (5)$$

$$Precision = \frac{TP+TN}{TP+FP+TN+FN}, \quad (6)$$

$$Recall = \frac{TP}{TN+FN}, \quad (7)$$

$$F1 - Measure = 2x\left(\frac{prec \times rec}{prec+ewc}\right), \quad (8)$$

Where TP = true positive, TN = true negative, FP = false positive and FN = false negative.

This evaluation approach allows for a comprehensive assessment of model performance in the context of multilabel classification of patent documents, focusing on the balance between precision and recall represented by the F1 score [38].

I. Query and Prediction Module

The query and prediction module provides an interface for new patent input and performs IPC code predictions. Input data undergo preprocessing consistent with the previous module. The trained OneVsRestClassifier model with LinearSVC is applied for prediction, involving TF-IDF transformation, model application, conversion to probabilities and application of the optimal threshold. Relevant documents are retrieved based on the predicted IPC codes and user keywords, allowing for efficient searching in the patent database [26], [34].

J. Presentation of Results Module

The results presentation module presents a concise overview of relevant patent literature. This module displays related patent documents, key information, predicted IPC codes with confidence levels, matching keywords and visualisation of the IPC code distribution. Using automatic summarisation techniques, this module generates brief but informative summaries of each relevant patent document, facilitating an efficient literature review process and enabling quick identification of the most relevant patents [25], [10].

IV. RESULTS AND DISCUSSION

The implementation of OneVsRestClassifier with LinearSVC as the base classifier for multilabel patent classification has yielded promising results. The model achieved a micro F1 score of 0.6723 and a macro F1 score of 0.6009, indicating reasonably good overall performance across various patent categories. These scores suggest that the model has a balanced capability in handling both frequent and rare patent classes, although there is still room for improvement.

Such balanced performance aligns with the broader literature on patent classification complexities [7], in which heterogeneous technology domains often require the careful handling of imbalanced labels.

In comparison to earlier approaches, hybrid methods (e.g. KNN+fastText) [9] and fine-tuned transformer-based models (e.g. BERT and XLNet) [3] have been explored by prior work on monolingual patent classification. While these studies report competitive or even state-of-the-art F1 metrics on single-language datasets, they do not address bilingual corpora (e.g. Indonesian–English). By contrast, our approach handles cross-lingual patent data and addresses class imbalance, thereby filling a gap not extensively covered in previous work.

The hyperparameter optimisation process, using GridSearchCV, identified an optimal C parameter of 0.1 for LinearSVC. This relatively low value indicates that the model prefers a large margin, potentially enhancing its generalisation capability. Interestingly, the threshold optimisation process found that the default threshold of 0.5 was optimal for converting probabilities into binary predictions, suggesting that the raw predictions of the model are well calibrated.

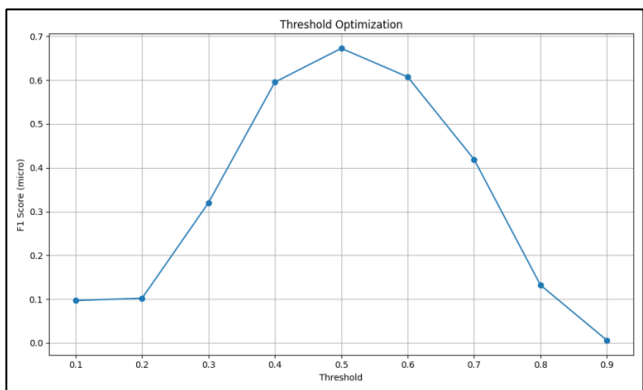
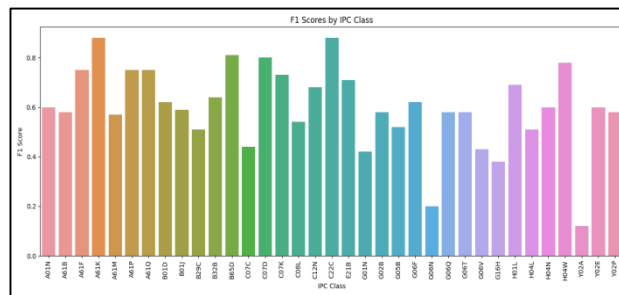


Fig. 3. Performance analysis of IPC patents.

Performance analysis by class revealed significant variations among the different IPC classes. Some categories, such as C22C (Alloys) and A61K (Preparations for medical, dental or toilet purposes), showed very good performance, with an F1 score of 0.88. This result suggests that these categories may have distinct features or terminology that the model can effectively identify. Conversely, categories such as Y02A (Technologies for adaptation to climate change) and G06N (Computer systems based on specific computational models) showed lower performance, with F1 scores of 0.12 and 0.20, respectively. These differences highlight the challenges in handling the inherent complexity and potential imbalances in patent data across various technology domains.

The feature extraction approach, using TfidfVectorizer with 20,000 features and including bigrams, appears to have captured important nuances in the patent texts. The decision to focus on thorough text cleaning and stopword removal, rather than stemming, seems effective, as evidenced by the overall model performance. However, the varying performance across classes suggests that there might be room for further refinement of the feature extraction process for certain technology domains.



process can involve exploring more advanced NLP techniques, such as BERT or domain-specific language models pretrained on patent data. Additionally, investigating techniques to improve performance in low-scoring classes, such as oversampling or developing class-specific features, could yield further improvements. Such strategies align with contemporary research calling for data augmentation and specialised embeddings to enhance multilabel patent classification [14].

In conclusion, while the current methodology demonstrates good potential in tackling the complex task of multilabel patent classification across languages, there remains room for improvement. The performance of the model suggests that it could be a valuable tool in assisting patent classification processes, potentially enhancing efficiency and consistency in international patent classifications. However, further research and refinement are needed to address the challenges identified, particularly in handling the diverse and evolving nature of technological innovations reflected in patent documents.

V. CONCLUSION

This research developed and evaluated a multilabel classification model for patent documents using a machine learning approach. The OneVsRestClassifier model with LinearSVC as the base classifier demonstrated competitive performance, achieving a micro F1 score of 0.6723 and a macro F1 score of 0.6009. These results indicate the potential of the model to handle the complexity of multilabel and multilingual patent classification.

In contrast to the hybrid KNN–fastText approach proposed by Yadrintsev and Sochenkov [9], which showed improved classification results on Russian and English texts through a stacking meta-classifier, our work specifically addressed bilingual data (Indonesian–English) and class imbalance in the IPC. Similarly, Haghghian Roudsari et al. [3] leveraged BERT, XLNet and other transformer-based models for multilabel patent classification but focused on monolingual English corpora. Our framework addressed this gap by targeting cross-lingual challenges and imbalanced labels within a single methodology, allowing for the robust handling of diverse patents.

The use of TfidfVectorizer with 20,000 features, including bigrams, proved effective in capturing important nuances in patent texts, although there is still room for refinement. Performance analysis revealed variations across IPC classes, indicating the need for targeted improvements in lower-performing categories (e.g. Y02A). Nevertheless, several limitations remain, which are as follows:

1) *Vocabulary coverage*: The TF–IDF approach, while effective, may not fully capture deep contextual or semantic relationships.

2) *Data scope*: This study focuses on Indonesian–English patents. Extending to additional languages or specialised subfields may require further adaptation.

3) *Class imbalance handling*: Although weighted learning helps mitigate skew, advanced sampling or data augmentation strategies could further improve performance for rare IPC codes.

Despite these limitations, this research contributes to the development of an automated patent classification system that has the potential to increase efficiency in patent analysis. Although the results are promising, it is important to remember the crucial role of human experts, especially for highly specialised IPC classes. With further refinements, the methodology outlined here can become a valuable supporting tool in the patent classification process, facilitating effective intellectual property management. This work paves the way for further progress in automated patent classification, addressing the multilingual and imbalanced data challenges inherent in international patent systems.

REFERENCES

- [1] E. Sharma, C. Li, and L. Wang, “BigPatent: A large-scale dataset for abstractive and coherent summarization,” *ACL 2019 - 57th Annu. Meet. Assoc. Comput. Linguist. Proc. Conf.*, pp. 2204–2213, 2020, doi: 10.18653/v1/p19-1212.
- [2] S. Casola and A. Lavelli, “Summarization, simplification, and generation: The case of patents,” *Expert Syst. Appl.*, vol. 205, 2022, doi: 10.1016/j.eswa.2022.117627.
- [3] A. Haghghian Roudsari, J. Afshar, W. Lee, and S. Lee, “PatentNet: Multi-label classification of patent documents using deep learning based language understanding,” *Scientometrics*, vol. 127, no. 1, pp. 207–231, 2022, doi: 10.1007/s11192-021-04179-4.
- [4] Y. Yoo, T.-S. Heo, D. Lim, and D. Seo, “Multi label classification of artificial intelligence related patents using modified D2SBERT and sentence attention mechanism,” 2023, <https://arxiv.org/abs/2303.03165>
- [5] B. S. Haney, “Patents for NLP software: An empirical review,” *SSRN Electron. J.*, 2020, doi: 10.2139/ssrn.3594515.
- [6] H. S. Al-Khalifa, T. AlOmar, and G. AlOlyyan, “Natural language processing patents landscape analysis,” *Data*, vol. 9, no. 4, 2024, doi: 10.3390/data9040052.
- [7] A. Abbas, L. Zhang, and S. U. Khan, “A literature review on the state-of-the-art in patent analysis,” *World Pat. Inf.*, vol. 37, pp. 3–13, 2014, doi: 10.1016/j.wpi.2013.12.006.
- [8] R. S. Eisenberg, “Patents and data-sharing in public science,” *Ind. Corp. Chang.*, vol. 15, no. 6, pp. 1013–1031, 2006, doi: 10.1093/icc/dt025.
- [9] V. V. Yadrintsev and I. V. Sochenkov, “The hybrid method for accurate patent classification,” *Lobachevskii J. Math.*, 2019. [Online]. Available: <https://link.springer.com/article/10.1134/S1995080219110325>
- [10] H. Zhu, C. He, Y. Fang, B. Ge, M. Xing, and W. Xiao, “Patent automatic classification based on symmetric hierarchical convolution neural network,” *Symmetry (Basel)*, vol. 12, no. 2, pp. 1–12, 2020, doi: 10.3390/sym12020186.
- [11] C. Diaz-Asper, M. K. Hauglid, C. Chandler, A. S. Cohen, P. W. Foltz, and B. Elvevåg, “A framework for language technologies in behavioral research and clinical applications: Ethical challenges, implications, and solutions,” *Am. Psychol.*, vol. 79, no. 1, pp. 79–91, 2024, doi: 10.1037/amp0001195.
- [12] F. Bacinger, I. Boticki, and D. Mlinaric, “System for semi-automated literature review based on machine learning,” *Electron.*, vol. 11, no. 24, 2022, doi: 10.3390/electronics11244124.
- [13] P. H. Santoso, E. Istiyono, Haryanto, and W. Hidayatulloh, “Literature using machine learning,” *Data*, vol. 7, pp. 1–41, 2022.
- [14] Y. Zhang and Z. Lu, “Exploring semi-supervised variational autoencoders for biomedical relation extraction,” *Methods*, vol. 166, no. November 2018, pp. 112–119, 2019, doi: 10.1016/j.ymeth.2019.02.021.
- [15] W. Weng, D. H. Wang, C. L. Chen, J. Wen, and S. X. Wu, “Label specific features-based classifier chains for multi-label classification,” *IEEE Access*, vol. 8, pp. 51265–51275, 2020, doi: 10.1109/ACCESS.2020.2980551.
- [16] A. Kravets, N. Shumeiko, B. Lempert, N. Salmikova, and N. Shcherbakova, ““Smart queue” approach for new technical solutions discovery in patent applications,” *Commun. Comput. Inf. Sci.*, vol. 754, pp. 37–47, 2017, doi: 10.1007/978-3-319-65551-2_3.

- [17] X. Yu and B. Zhang, "Obtaining advantages from technology revolution: A patent roadmap for competition analysis and strategy planning," *Technol. Forecast. Soc. Change*, vol. 145, no. October, pp. 273–283, 2019, doi: 10.1016/j.techfore.2017.10.008.
- [18] A. Khurana and V. Bhatnagar, "Investigating entropy for extractive document summarization," *Expert Syst. Appl.*, vol. 187, 2022, doi: 10.1016/j.eswa.2021.115820.
- [19] F. Zhu, X. Wang, D. Zhu, and Y. Liu, "A supervised requirement-oriented patent classification scheme based on the combination of metadata and citation information," *Int. J. Comput. Intell. Syst.*, vol. 8, no. 3, pp. 502–516, 2015, doi: 10.1080/18756891.2015.1023588.
- [20] S. Sarica, J. Luo, and K. L. Wood, "TechNet: Technology semantic network based on patent data," *Expert Syst. Appl.*, vol. 142, p. 112995, 2020, doi: 10.1016/j.eswa.2019.112995.
- [21] N. Shibayama, R. Cao, J. Bai, W. Ma, and H. Shinnou, "Evaluation of pretrained {BERT} model by using sentence clustering," *Proc. 34th Pacific Asia Conf. Lang. Inf. Comput.*, pp. 279–285, 2020. [Online]. Available: <https://aclanthology.org/2020.paclic-1.32>
- [22] N. Febriyanti, D. Palupi, and O. Arsalan, "Text similarity detection between documents using case based reasoning method with cosine similarity measure (case study SIMNG LPPM Universitas Sriwijaya)," vol. 3, no. 2, pp. 36–45, 2022.
- [23] I. O. Suzanti, A. Jauhari, N. Hidayanti, I. Y. Harianti, and F. A. Mufarroha, "Comparison of stemming and similarity algorithms in Indonesian translated Al-Qur'an text search," *J. Ilm. Kursor*, vol. 11, no. 2, pp. 91–91, 2022. [Online]. Available: <http://kursorjournal.org/index.php/kursor/article/view/280>
- [24] J. Jang and C. O. Kim, "One-vs-Rest network-based deep probability model for open set recognition," 2020, <http://arxiv.org/abs/2004.08067>
- [25] X. Chen and N. Deng, "A semi-supervised machine learning method for chinese patent effect annotation," *Proc. - 2015 Int. Conf. Cyber-Enabled Distrib. Comput. Knowl. Discov. CyberC 2015*, pp. 243–250, 2015, doi: 10.1109/CyberC.2015.99.
- [26] R. Ros, E. Bjamason, and P. Runeson, "A machine learning approach for semi-automated search and selection in literature studies," *ACM Int. Conf. Proceeding Ser.*, vol. Part F1286, pp. 118–127, 2017, doi: 10.1145/3084226.3084243.
- [27] M. Abazar, P. Masjedi, and M. Taheri, "A binary relevance adaptive model-selection for ensemble steganalysis," *ISeCure*, vol. 14, no. 1, pp. 105–113, 2022, doi: 10.22042/isecure.2021.262990.596.
- [28] Y. Liu, "Yang Liu (刘洋)," p. 86, 2010.
- [29] H. Sasaki and I. Sakata, "Identifying potential technological spin-offs using hierarchical information in international patent classification," *Technovation*, vol. 100, no. September 2019, p. 102192, 2021, doi: 10.1016/j.technovation.2020.102192.
- [30] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains: A review and perspectives," *J. Artif. Intell. Res.*, vol. 70, pp. 683–718, 2021, doi: 10.1613/JAIR.1.12376.
- [31] W. Chmielnicki and K. Stapor, "Using the one-versus-rest strategy with samples balancing to improve pairwise coupling classification," *Int. J. Appl. Math. Comput. Sci.*, vol. 26, no. 1, pp. 191–201, 2016, doi: 10.1515/amcs-2016-0013.
- [32] M. Suzgun, L. Melas-kyriazi, and S. K. Sarkar, "The Harvard USPTO Patent Dataset: Corpus of patent applications," no. MI, pp. 1–38, 2020.
- [33] Z. Wang, T. Wang, B. Wan, and M. Han, "Partial classifier chains with feature selection by exploiting label correlation in multi-label classification," *Entropy*, vol. 22, no. 10, pp. 1–22, 2020, doi: 10.3390/e22101143.
- [34] M. S. Hajmohammadi, R. Ibrahim, and A. Selamat, "Bi-view semi-supervised active learning for cross-lingual sentiment classification," *Inf. Process. Manag.*, vol. 50, no. 5, pp. 718–732, 2014, doi: 10.1016/j.ipm.2014.03.005.
- [35] R. T. Wahyuni, D. Prastiyanto, and E. Suprptono, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi," vol. 9, no. 1, 2017.
- [36] R. Sharma and P. Shrinath, "Ensemble of weighted code mixed feature engineering and machine learning-based multiclass classification for enhanced opinion mining on unstructured Data," vol. 15, no. 10, pp. 1220–1230, 2024.
- [37] A. Alshammari, F. Alotaibi, and S. Alnafrani, "Prediction of outpatient no-show appointments using machine learning algorithms for pediatric patients in Saudi Arabia," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 8, pp. 108–116, 2024, doi: 10.14569/IJACSA.2024.0150812.
- [38] Dafid, Ermatita, and Samsuryadi, "A framework for predicting academic success using classification method through filter-based feature selection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 9, pp. 435–444, 2023, doi: 10.14569/IJACSA.2023.0140947.

Dolphin Inspired Optimization for Feature Extraction in Augmented Reality Tracking

Indhumathi S, Christopher Clement J

School of Electronics Engineering, Vellore Institute of Technology, Vellore, India 632014

Abstract—Feature extraction has the prominent role in Augmented Reality (AR) tracking. AR tracking monitor the position and orientation to overlay the 3D model in real-world environment. This approach of AR tracking, encouraged to propose the optimum feature extraction model by embedding the dolphin grouping system. We implemented dolphin grouping algorithm to extract the features effectively without compromising the accuracy. In addition, to prove the stability of the proposed model, we have included the affine transformation images such as rotation, blur image and light variation for the analysis. The Dolphin model obtained the average precision of 0.92 and recall score of 0.84. Whereas, the computation time of dolphin model is identified as 2ms which is faster than the other algorithm. The comparative result analysis reveals that accuracy and the efficiency of the proposed model surpasses the existing descriptors.

Keywords—Feature descriptor; dolphin optimization; feature extraction; augmented reality tracking

I. INTRODUCTION

Feature extraction is the function of Machine Learning (ML) algorithm, which identify the significant features present in the image. Feature descriptor or extractor is the model to extract the features in the form of edge, corner, texture, contour, color and shape of the image. These extracted features have to be robust and efficient in affine transformation of the image and it can be done by adopting the handcrafted and learning based model. Feature extraction is used in many applications: (i) Autonomous vehicle: To recognize the objects and predict the distance of vehicle. (ii) Augmented Reality: Feature extraction applied to superimpose the augmented model in real world. (iii) Manufacturing Industry: To identify the defects and ensure the safety of the products. (iv) Medical applications: Feature extraction aids to identify the early detection and diagnosis of the critical diseases. Therefore, many algorithm have been published for various feature extraction applications in recent years. Speeded Up Robust Features (SURF) [1] deploys box filters for the feature extraction in object recognition and tracking. Scale Invariant Feature Transform (SIFT) [2] is an image matching algorithm, used to extract the features present in multiple scale images, the difference of gaussian is utilized for the feature prediction of multiscale images in SIFT. The Binary Robust Invariant Scalable Keypoints (BRISK) descriptor embedded with Bee colony algorithm, adopts the sampling pattern to extract the robust keypoints present in image matching [3]. Moreover, the Histogram of Oriented Gradients (HOG) [4] aid the histogram technique for the feature extraction of image, which enhances the human detection process. In [5] author proposed learning based feature descriptor model for the anomaly detection. They examined the

model with 32 datasets and its result provides the accuracy of the model. However, the model encountered the challenge as computation complexity to process the large size of data with affine transformation. To address the above problem, a Rotation Invariant and Globally Aware Descriptor (RIGA) [6] is proposed. RIGA, extract the feature correspondences of the rotation transformation image. This model enhances the rotation in-variant property of point cloud by deploying the Point-Net architecture which consumes the input from a rotated traditional descriptors. Vision transformer embrace the global awareness geometry in RIGA. Therefore, RIGA performs well in both rotation in-variance and global awareness of the descriptor. Nevertheless, its feature prediction lags in additional transformation properties such as scale, light and occlusion transformation. To deal these challenges, Fencher Multiscale Local Descriptor (FMLD) is proposed. FMLD extract the features from light illumination image. The model uses magnitude and angle fusion for feature prediction. The FMLD performs well in occlusion and light variation. However, this model has the limitation in computation complexity [7]. The computation complexity problem has been addressed in Superpixel-based Brownian Descriptor (SBD). Integration of superpixel with brownian model provides the internal structures of the Hyper Spectral Image. This method extract the efficient spectral spatial features. This hybrid model reduces the computation complexity [8]. The artificial bee colony algorithm is implemented to extract only five features, which achieve 98.8 % of accuracy to detect cyberattacks [9].

Outlier detection using projection pursuit is one of the techniques, which has not been used so far in the feature extraction. However, the authors of [10] have developed four novel feature selection techniques using the concept of outlier detection and projection pursuit by exploiting the bio-inspired algorithms. The method seemed to outperform the state-of-the-art techniques with an improvement rate ranging between 0.76% and 36%.

Hence, the descriptor is processing with more number of images and datasets so it consumes more computation. There is a trade-off between accuracy and computation in feature descriptor. So researchers are working in this challenge to improve the feature extraction model. However, with respect to the application we can modify the trade-off. Since, the feature can be in any form as mentioned earlier in this section, it is important to normalise or optimize the model.

One of the main challenge of feature extraction is to diminish the data complexity. We proposed the new model dolphin optimization to extract the essential feature with effective computation. The main contribution of the paper is:

- The two filters are proposed to calculate the gradient

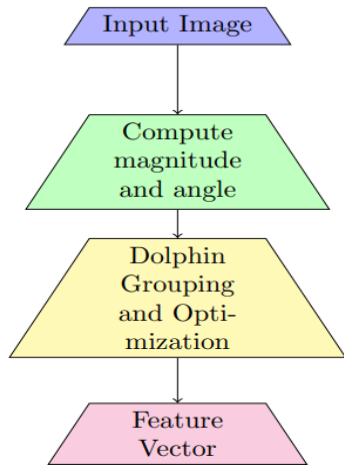


Fig. 1. Dolphin model process flow diagram.

magnitude and orientation of the pixels.

- The spatial location along with magnitude and angle are minimized to measure the similarity in groups.
- In each group feature is formed based on dolphin inspired model.
- The feature is extracted using dolphin optimization.

A. Organization of the Research

The paper is organised as follows in Section II, the related bio inspired optimization models are discussed. In Section III, the dolphin inspired feature extraction methodology is included. Section IV discusses the results and validation of the model followed by a conclusion in Section V.

II. RELATED WORK

This proposed work discuss about the optimization of feature extraction model. Wherefore, here we discuss the recent work related to the subject. In [11] author proposed Principal Component Analysis (PCA) in edge feature extraction for 3D point cloud. This model uses the covariance matrix to predict the features. The PCA is embedded for the optimization of the feature selection. The accuracy of the PCA is compared with traditional method. The PCA model surpasses the traditional model in feature selection. Besides the PCA model a dual correlate, course fine optimization technique is also involved in the feature extraction. This dual correlate model extract the feature in two level such as course and fine level which refines the feature selection process [12]. [13] author designed a locality based approach for the feature selection. It can create the local topology structures to identify the robust features. It identifies the features by matching the similar objects between two images and it removes the mismatch to obtain the robust matching. In addition to that, recently many bio-inspired models are proposed for the feature optimization which we discuss in next section.

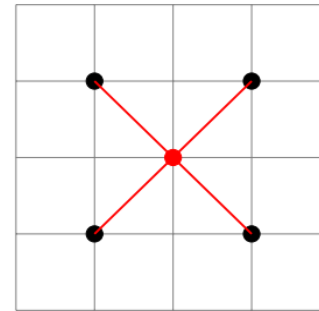


Fig. 2. Filter design to compute gradient in vertical image plane.

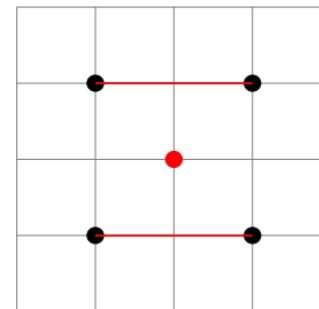


Fig. 3. Filter design to compute gradient in horizontal image plane.

A. Bio-Inspired Optimization in Feature Extraction

Many Bio-inspired optimization model have been implemented to enhance the accuracy of the feature extraction. In this paper [14], author published support vector machine algorithm by incorporating the ant colony optimization which identifies the early stage of cancer detection. The abnormal cells or features are recognized by gray level co-occurrence matrix then ant colony optimization is applied to extract the significant features. In [15] author applied the Binary whale optimization algorithm to enhance the accuracy of the feature selection in molecular descriptor. This molecular descriptor contains all the information about molecule in drug market. So the prediction of prominent features is necessary to avoid the heavy computation of the descriptor. This can be achieved by the innovation of non-linear time varying sigmoid function in whale optimization. Therefore, this whale optimization improves the feature selection in molecular descriptor. Further, to optimize the texture features, the Binary particle swarm optimization technique [16] is implemented to select the desired texture features from an optical character recognition system. This system accepts only the text, so to automate the model we need to suppress the non text from the text, it can be done with the help of swarm optimization model. Similarly, in [17] the author proposed a Crow search algorithm to select the essential feature from face image using the neural network model. The Local ternary pattern with SURF based hybrid descriptor is used to predict the features in the model. Crow optimization act as a prominent role to achieve the extraction of optimized features with the accuracy of 95% for face recognition models.

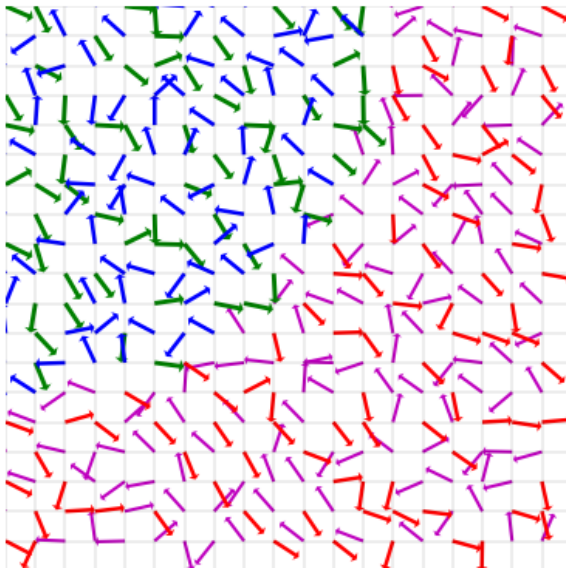


Fig. 4. The lines with different color intensities describe the dolphin grouping - 4 groups - based on spatial distances, angle and magnitude of pixel intensities in image plane.

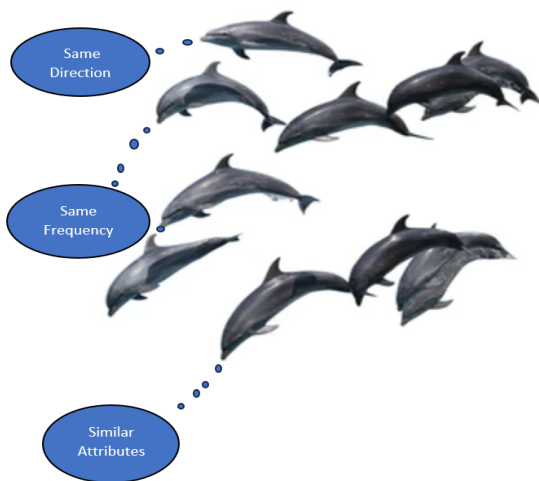


Fig. 5. Dolphin attributes mentioned in image to form a group.

III. METHODOLOGY

This section presents the detail of the proposed model. Our methodology consist of three stages: Social secrets of Dolphin, AR tracking and Dolphin dynamic optimization. The secret behaviour of dolphins are discussed in Section III-A for the better understanding of the dolphin optimization. In Section III-B provides the detail of the AR tracking system along with filter design implementation and finally Section III-C describes the dolphin implementation in feature extraction.

A. The Social Secrets of Dolphins: How These Clever Creatures Form the Groups

Dolphin and human have many similar characteristics, which motivated to adopt the dolphin behaviour in feature extraction. Dolphin is a social animal like human so it can communicate, eat and stay together as a group. How it forms the group is really an impressive and unique nature of

dolphin. After many studies about the dolphins we come to the conclusion that the dolphin can form a group according to the factors such as species, spatial proximity, behaviour, food habits, age and family association [18]. A single member in a group is the sample to understand the behaviour of the group. Moreover, dolphin can interact with each other through their signature sound [19] and body languages. This way of communication is helpful to share their thoughts between them. The dolphin group size, is different with respect to the food availability region. Usually, each group has the maximum limit of 30 members however, mega-group consist of 1000 members. This mega-group can form instantly where it is attracted by the abundant food. Dolphin studies says, it has preferences to meet the individuals and it can remember and identify them even after long period of separation. The reason of group formation is for their safety and growth.

From the understanding of the dolphin behaviour it narrates few points with respect to the creation of group:

- Dolphin can form a group to protect themselves.
- The signature whistles are used for communication between individuals.
- The whistle sound is composed from the hereditary and each dolphin has its unique vocals with respect to the certain range of frequencies.
- This sound helps to identify their location.
- The similarity in behaviours are attracted to be a member in the groups.

These behaviour of dolphin is implemented in our model to extract the features and the dolphin group is shown in Fig. 5. In the upcoming section, we will discuss the process flow of AR tracking with filter design and the necessity of dolphin dynamic optimization in feature extraction.

B. Augmented Reality Tracking

AR Tracking system immerse the 3D model in physical world. The Tracking process consist of five steps:

- 1) *Pre-processing*: Re-size the image to form a uniformity in the analysis.
- 2) *Feature detection*: Detector can detect the necessary information as a keypoints from the image.
- 3) *Feature description*: The feature vector is manifested by the inclusion of neighbouring pixel surrounded by the keypoint.
- 4) *Feature matching*: The feature vector of reference and test image is compared to find the matches in the image.
- 5) *3D Model*: Once the feature is matched in the above stage then in this stage it creates a 3D model.

These above mentioned five stages are the process flow of AR tracking. We focus our work in development of the feature descriptor for AR tracking and the flow diagram of our design is illustrated in Fig. 1. According to the flow diagram, the first stage contains the input image. The image matrix of the input is represented as $I_{M \times N}$. Moreover, we designed two filters to measure the gradient changes in vertical and horizontal plane

of the image. The pixel point of interest is considered as $I_{m,n}$. The m is varied from 0 to M similarly, the n changes from 0 to N in image. The vertical and horizontal filters are shown in Fig. 2 and Fig. 3 respectively. From the figure, the pixel point of interest is shown in red color and the four neighbouring pixels are indicated by black for the visualization. The red line reveals the relation between those pixels. The gradient changes of two plane is measured as \mathbf{P}_x and \mathbf{P}_y from Eq. (1) and Eq. (2).

$$\mathbf{P}_x = I(m-1,n-1) - I(m-1,n+1) + I(m+1,n-1) - I(m+1,n+1) \quad (1)$$

$$\mathbf{P}_y = I(m-1,n-1) - I(m+1,n+1) + I(m-1,n+1) - I(m+1,n-1) \quad (2)$$

The Eq. (1) and Eq. (2) are used to obtain the gradient magnitude and orientation as per the Eq. (3) and Eq. (4).

$$G = \sqrt{(\mathbf{P}_x)^2 + (\mathbf{P}_y)^2} \quad (3)$$

$$\Phi = \arctan\left(\frac{\mathbf{P}_y}{\mathbf{P}_x}\right) \quad (4)$$

C. The Proposed Model-Dolphin Dynamic Optimization

The robust feature extraction is the fundamental task of AR Tracking. Hence, the significance of the feature descriptor model is growing, we need to solve the challenge arises to the feature descriptor, that is to identify the stable feature with efficient computation time. In this paper, we provide the design of optimum feature descriptor by incorporating the dolphin optimization model. The main concept of Dolphin Dynamic Optimization is transforming the behaviour of dolphin into image plane for the effective feature extraction using the grouping techniques. The reason for the implementation of dolphin algorithm is how the individuals can combine with its neighbours to form a group is similar to the group of pixels along with its neighbour tends to create a feature. The dolphin key parameters mentioned in the Section III-A is used for the feature prediction. The goal of this study is to group the features precisely. This feature grouping leads to retrieve the shape of the image.

Dolphin model is innovated to predict the shape of the image in a 2-Dimensional space. The dolphin model divide the data points into K groups. The core plan of the model utilize few parameters to form a group such as, image gradient, orientation and spatial location respectively. The image gradient and orientation is obtained from the vertical and horizontal filter mentioned in Fig. 2 and Fig. 3. The gradient and orientation measurements are discussed in Section III-B. Moreover, at this stage the keypoints are ready but its not fit into the shape to retrieve the image. This gap can be fulfilled by the implementation of the dolphin dynamic optimization. The dolphin algorithm works as follows:

- Randomly initialize the dolphin head from the image spatial location.
- The selected head spatial location, gradient and orientation is compared with other data points in the image plane.

- The minimization of the cost function is the criteria to join as a member into the group.
- To fix the head of the dolphin the process is repeats for the R number of iterations until it convergence.
- Then, finally we have the group of features which reflects the shape of the image.

Considering the 8 bit image representation, the gradient magnitude can vary from 0 to 366.6 and the orientation is in the range of 0 to 90°. Assuming that the image dimension is $M \times N$, then the shape normalization is ensured in such a way that M and N correspond to λ_K and δ_K respectively. At the beginning, K spatial locations of the dolphin heads are uniformly spread having the distribution ranging within M and N respectively. Let $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_K\}$ and $\delta = \{\delta_1, \delta_2, \dots, \delta_K\}$ indicate these initial locations. The cost function accounts the spatial distance, magnitude difference and orientation between dolphin heads and every other pixels as given in Eq. (5).

$$\mathbf{J} = \sum_{m=1}^M \sum_{n=1}^N \sum_{k=1}^K \mathbf{r}_k(m,n) \{ [m - \lambda_k]^2 + [n - \delta_k]^2 + [G(m,n) - G(\lambda_k, \delta_k)] + [\Phi(m,n) - \Phi(\lambda_k, \delta_k)] \} \quad (5)$$

The indicator function $\mathbf{r}_k(m,n) \in \{0,1\}$ is introduced to mention the spatial values m and n at which \mathbf{J} is minimum. As a result, we need to minimize \mathbf{J} by differentiating it with respect to $\mathbf{r}_k(m,n)$ in the first step followed by λ_k and δ_k partially. By doing this way, a group is formed with updated dolphin head position in each iteration. The continuous shift of dolphin head in each iteration leads to identify the head by its optimization function. The update takes place, until a stopping criterion is met. When the group can not identify a new head, it is assumed that stopping criterion is met. According to Eq. (6), \mathbf{J} attains minimum with the indices p and q , so that $\mathbf{r}(p,q)$ is 1.

$$\mathbf{r}_k(m,n) = \begin{cases} 1; & \text{if } m = \arg\min_p; n = \arg\min_q \\ & \left\{ [m - \lambda_p]^2 + [n - \delta_q]^2 + [G(m,n) - G(\lambda_p, \delta_q)] + [\Phi(m,n) - \Phi(\lambda_p, \delta_q)] \right\} \\ 0; & \text{otherwise} \end{cases} \quad (6)$$

In this way, the dolphin groups are formed to extract the image features. The updation of a new dolphin head positions are given by Eq. (7) and (8).

$$\lambda_k = \frac{\sum_m \sum_n r_{mn} \cdot m}{\sum_m \sum_n r_{mn}} \quad (7)$$

$$\delta_k = \frac{\sum_m \sum_n r_{mn} \cdot n}{\sum_m \sum_n r_{mn}} \quad (8)$$



(a) Graffiti image.

(b) House image.

Fig. 6. Input images for test without any transformation.

Algorithm 1: Feature Extraction of Dolphin Algorithm

```

1: Input:  $I_{M \times N}$ 
2: Output:  $C$ 
3:  $P_x = I(m-1, n-1) - I(m-1, n+1)$ 
    $+ I(m+1, n-1) - I(m+1, n+1)$ 
4:  $P_y = I(m-1, n-1) - I(m+1, n+1)$ 
    $+ I(m-1, n+1) - I(m+1, n-1)$ 
5: Calculate  $G = \sqrt{p_x^2 + p_y^2}$ 
6: Calculate  $\Phi = \arctan\left(\frac{P_y}{P_x}\right)$ 
7: while
   do
8:   for  $m = 1$  to  $M$  do
9:     for  $n = 1$  to  $N$  do
10:      for  $k = 1$  to  $K$  do
11:        Find  $J$  using (5)
12:        Minimize  $J$  to update
13:         $r_k(p, q)$  using (6).
14:        Update  $\lambda_k$  and  $\delta_k$  using (7)
15:      and (8)
16:    end for
17:     $C(m, n) = \min_{j \in \{1, 2, 3, \dots, K\}} J$ 
18:  end for
19: end while

```

The feature extraction of dolphin model is grouped based on Algorithm 1. The dolphin group formation with respect to the cost function minimization is illustrated in Fig. 4 which shows the line with arrow indicates the output of the cost function which is generated from the minimization of spatial location, gradient and orientation of the pixels. Each color in the plot indicates the different group of the dolphin which is directly proportional to the J function. Fig. 4 provides the

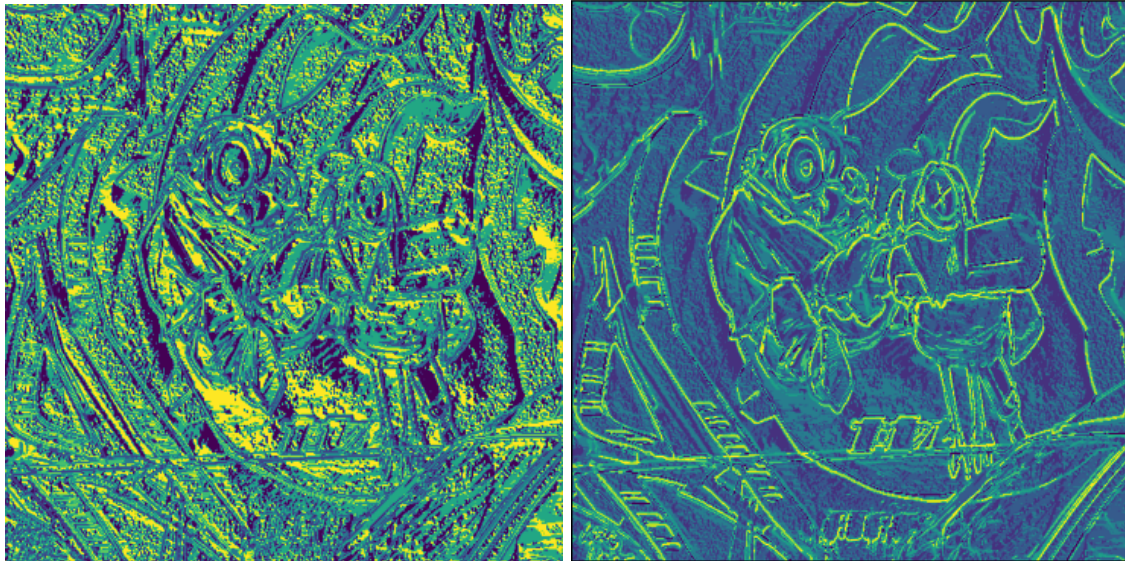
pictorial representation of the nearest spatial location with similar direction and magnitude belongs to one group. This way we can retrieve the shape of any image from the dolphin optimization model.

IV. RESULTS AND DISCUSSION

In this section, we discuss the simulation results of dolphin optimization algorithm with respect to feature extraction. The input images are adopted from two public datasets which is available in the following link [<https://www.robots.ox.ac.uk/>] and [<http://sipi.usc.edu/database/>]. We utilized two benchmark to evaluate the dolphin algorithm, namely accuracy and efficiency of the model. The accuracy of the feature prediction is measured using precision and recall score. The efficiency is defined from the processing time of the feature extraction model which is denoted as computation time. At initial stage, graffiti and house image both are tested to measure the accuracy of the features. These images are consist of different structures, hence its feature extraction is evaluated. In addition to that, the affine transformation image also included to validate the robustness of the design. The robustness of the feature extraction with respect to viewpoint variation, blurred image and light variation is measured using graffiti, bike, and car image respectively. Moreover, the transformed image such as graffiti is rotated with particular angle, bike image is tested with the noise, and car image light intensity is reduced for the validation. The experiments are simulated using python 3.10, with NVIDIA, 11th Generation i7processor.

A. Original Image Features

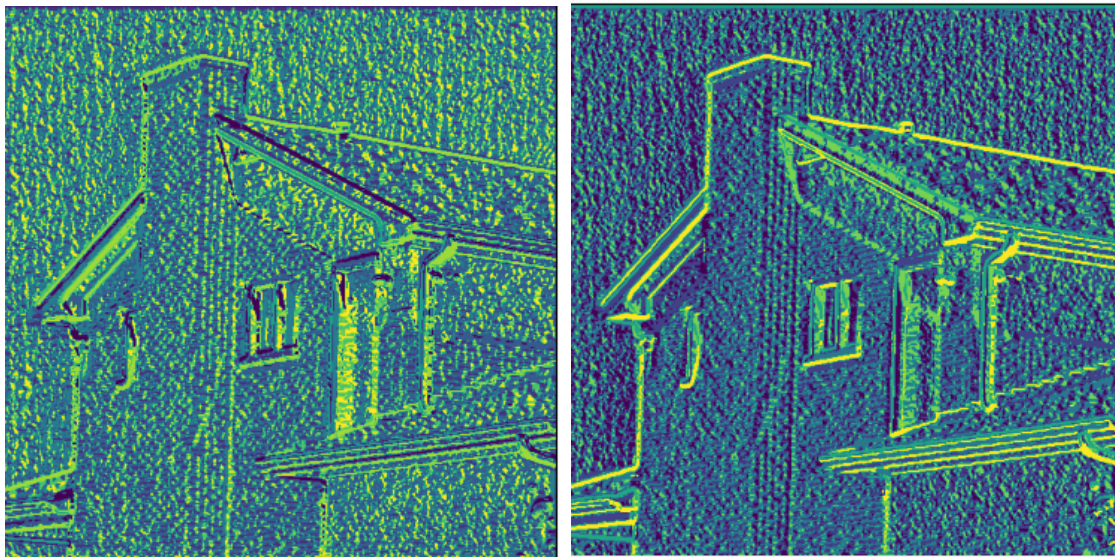
In Fig. 6 shows the original image of graffiti and house image which is considered as the reference because it is not given into any transformation. Graffiti and house image



(a) Feature extraction of dolphin for group size=6.

(b) Feature extraction of dolphin for group size=8.

Fig. 7. Feature extraction of graffiti images.



(a) Dolphin feature extraction of house image for group size=6.

(b) Dolphin feature extraction of house image for group size=8.

Fig. 8. Feature extraction of house images.

features are compared to provide the robustness of the dolphin model. Although, graffiti has more number of edges than the house image, dolphin model predicts the feature very well for both the images. For the detail analysis of feature extraction the graffiti and house image features are extracted with different group size. Initially, group size is low then we gradually increases the group size to verify the performance of the model. For the visualization of the feature, we have included two group sizes which is shown in Fig. 7a it carries the group size of 6 and Fig. 7b holds the group size of 8. The elbow method is aided to show the optimal group value of dolphin algorithm which is shown in Fig. 13. It indicates the optimum

value of the group is 8. Therefore, the optimum number of group size indicates the accurate feature prediction. From the results of Fig. 7b and 8b we can visualize the shape of the image, hence it proves all the edges are completely retrieved from the original image with respect to the cost function of J forms a group. Each group in image is illustrated with different color according to the intensity variation. The yellow color has high intensity, then the level of intensity is decreases with the different color representation such as green, blue and purple respectively. Each color has two different groups with the color deviation hence the total group is 8. Thus reflects in an image as an edge color from yellow to purple.



Fig. 9. Viewpoint and light variation of original images.

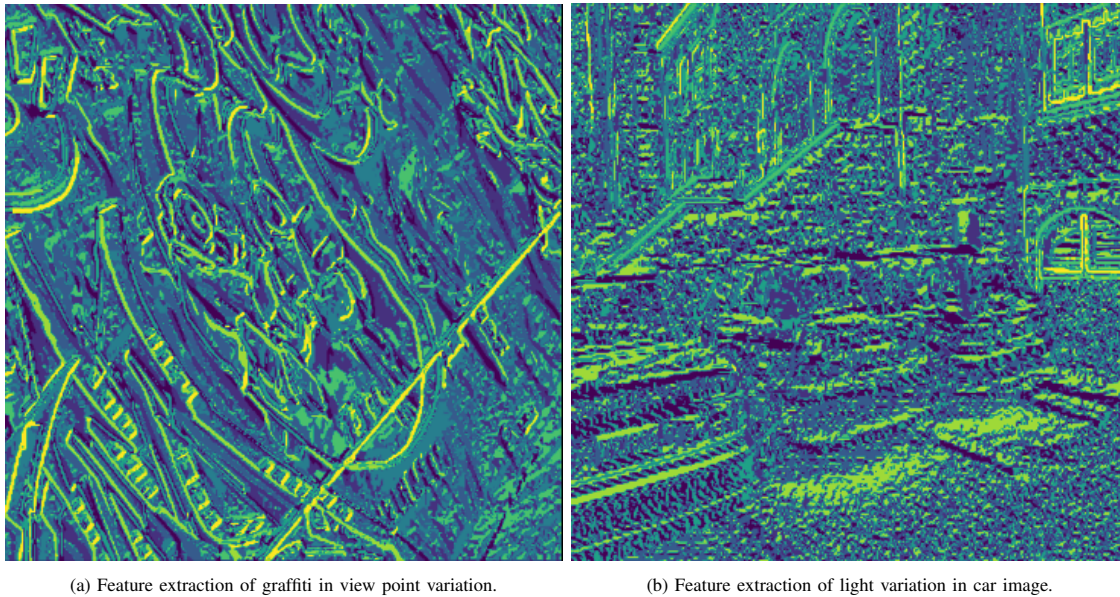


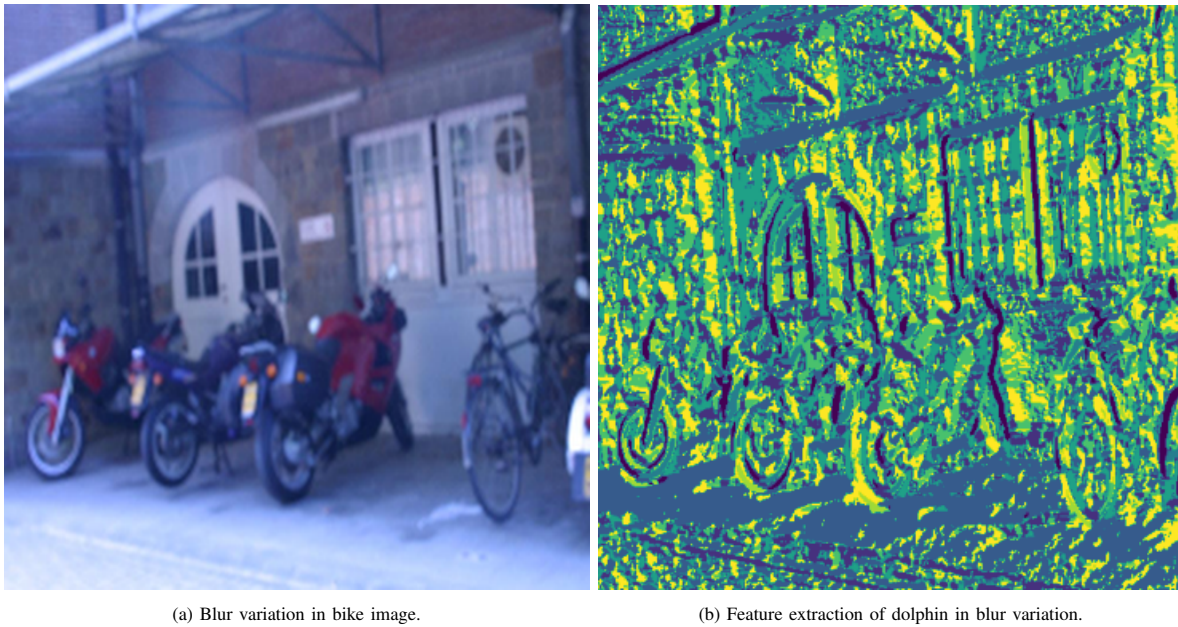
Fig. 10. Viewpoint and light variation of image features.

TABLE I. THE AVERAGE PRECISION AND AVERAGE RECALL: A COMPARISON OF PROPOSED ALGORITHM WITH OTHER ALGORITHMS

Descriptor	Average Precision	Average Recall
Dolphin	0.92	0.84
HOG	0.82	0.76
BRIEF	0.73	0.65
BRISK	0.63	0.59
SURF	0.57	0.44

TABLE II. COMPARISON ANALYSIS OF COMPUTATION TIME

Feature Descriptor	Computation Time(ms)
Dolphin	2.0
HOG	3.8
BRIEF	5.6
BRISK	13.7
SURF	18.6



(a) Blur variation in bike image.

(b) Feature extraction of dolphin in blur variation.

Fig. 11. Blur image and its feature extraction.

B. Transformation Image Features

The image transformation, indicates transformation of image from one form to other. For the analysis, we resized all the transformation test image into the size of 512×512 and the group size is chosen as 8. We used three transformation, namely, rotation, blur variation and light variation. The goal of this testing is to prove the robustness of the dolphin model in feature prediction with affine transformation. Fig. 9a shows the 40° rotated image of graffiti, hence it is created by view-point variation of the camera is given as 40° . In case of car image the light intensity is decreased from the original image to test the illumination in-variance of our proposed model refer Fig. 9b. In addition to that, the bike image which is shown in Fig. 11a is blurred due to the movement arises between scene and camera. The Fig. 10a shows the output of the rotated image which is reflected with the extraction of all the edge features present in image so our model outperforms in rotation in-variance. Moreover, the feature prediction in light intensity variation is very difficult to process whereas, our model reaches the success to regain the car image from light variation and its shown in Fig. 10b. However, the blur variation shown in Fig. 11b is lagging in accuracy of the feature prediction than the other transformation but still it can retrieve the shape of the image. Fig. 12 shows the feature extraction using dolphin method after compressing the original image. The 80% of compression is applied to the original image and then the features are extracted. The result verify that there is no compromise in extracting the features even after compression. Therefore, in accordance with affine transformation our model produces best results for rotation variation than light and blur variation. Even though our model provides with good accuracy of feature extraction, still there is a space for improvement of the model.

For the quantitative analysis, we have included the precision and recall value of the different descriptors to

evaluate the accuracy of the model. The validation image is taken as the graffiti with the size of 512×512 . The dolphin model is compared with recently proposed feature descriptors, namely, HOG, BRIEF, BRISK and SURF. The true prediction is measured using the precision and recall provides the correct identification of features in image the evaluation results are given in Table I. From the results dolphin model achieves good results than the existing models. The second largest value scored by BRIEF, its precision and recall value is better than BRISK and SIFT. Therefore, it concludes dolphin extract all the necessary features to retrieve the image.

In addition to that, to prove the efficiency of the model the computation time is measured. The computation time of the dolphin model is measured from $\mathcal{O}(M * N * K * R)$. The M and N are the size of the image and K indicates the group size then R is the iteration of the process. The system we used for the simulation is capable to run 5000 millions FLOPS per second. The model validated the results with the image size, number of groups and iteration value are assigned in such a way that, $M = 512$, $N = 512$, $K = 8$ and $R = 500$ then 1048576000 FLOPS are needed according to our model design. Therefore, from this validation, we can obtain the computation time of dolphin as 2.0ms which is faster than other algorithms as perceived from the Table II.

V. CONCLUSION

This article, proposed a optimized feature extraction model for AR tracking along with affine transformation such as rotation, blur and illumination variation using dolphin algorithm. Precisely, dolphin optimization contains two stages, namely, gradient computation and dolphin grouping.

We proposed two filters for the gradient measurement of the image pixels. Further, to measure the optimized grouping with respect to dolphin behaviour we tested the

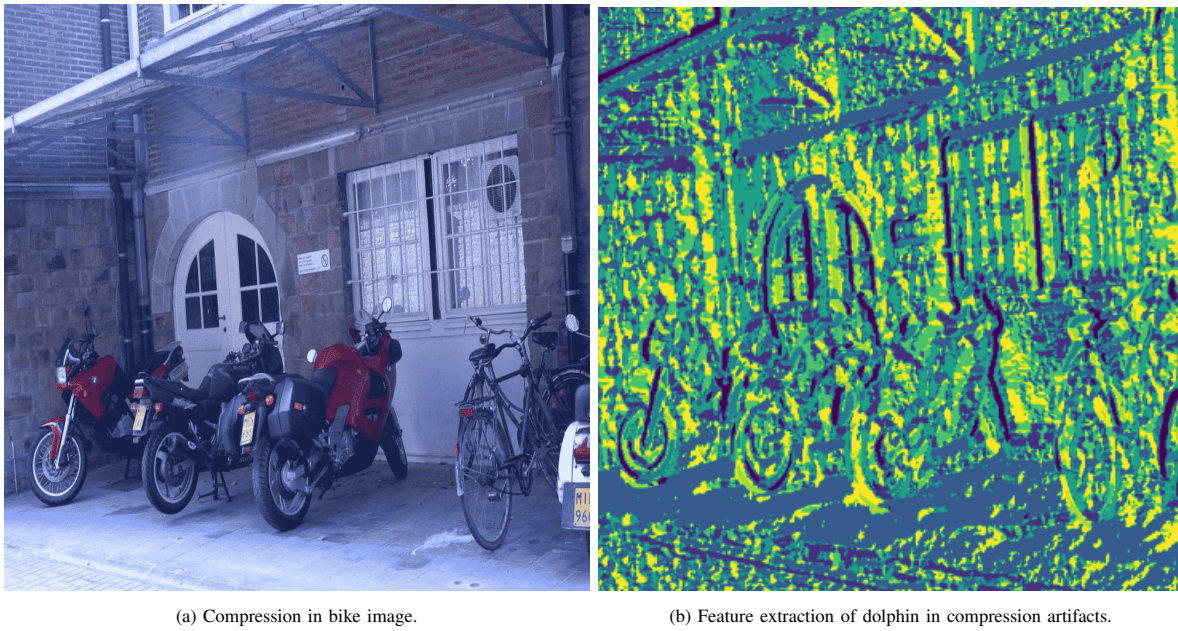


Fig. 12. Compressed image and its feature extraction.

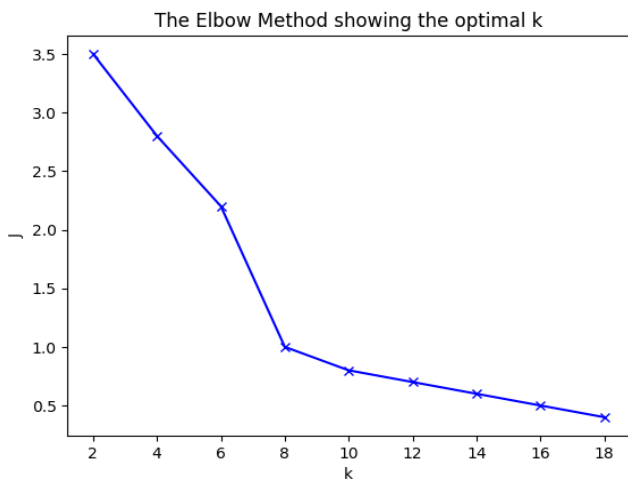


Fig. 13. Filter design to compute gradient in horizontal image plane.

image with several group size for the validation of the feature extraction. From the results we can conclude the optimum group size is identified as 8 for the robust feature prediction. The computation of the model surpasses the existing model is observed from the measurement. The accuracy of the image retrieval is measured in terms of precision and recall. Dolphin model outperforms other existing algorithm in terms of accuracy and efficiency. In future, the scale variation and partial occlusion can be included for the better development of feature extraction model in AR tracking.

ACKNOWLEDGMENT

The authors would like to thank god almighty for the support and thanks to our Vellore Institute for providing the facilities to finish our research.

REFERENCES

- [1] Z. F. Mohammed and D. J. Mussa, "Brain tumour classification using bof-surf with filter-based feature selection methods," *Multimedia Tools and Applications*, pp. 1–23, 2024.
- [2] N. Gupta and M. K. Rohil, "An elliptical sampling based fast and robust feature descriptor for image matching," *Multimedia Tools and Applications*, pp. 1–20, 2024.
- [3] A. Soualmi, A. Benhocine, and I. Midoun, "Artificial bee colony-based blind watermarking scheme for color images alter detection using brisk features and dct," *Arabian Journal for Science and Engineering*, pp. 1–14, 2023.
- [4] S. Karanwal, "Robust face descriptor in unconstrained environments," *Expert Systems with Applications*, p. 123302, 2024.
- [5] K. Batzner, L. Heckler, and R. König, "Efficientad: Accurate visual anomaly detection at millisecond-level latencies," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, pp. 128–138.
- [6] H. Yu, J. Hou, Z. Qin, M. Saleh, I. Shugurov, K. Wang, B. Busam, and S. Ilic, "Riga: Rotation-invariant and globally-aware descriptors for point cloud registration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
- [7] J. Feng, J. Xu, Y. Deng, and J. Gao, "A fechner multiscale local descriptor for face recognition," *The Journal of Supercomputing*, pp. 1–28, 2023.
- [8] S. Zhang, T. Lu, S. Li, and W. Fu, "Superpixel-based brownian descriptor for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2021.
- [9] H. Najafi Mohsenabad and M. A. Tut, "Optimizing cybersecurity attack detection in computer networks: A comparative analysis of bio-inspired optimization algorithms using the cse-cic-ids 2018 dataset," *Applied Sciences*, vol. 14, no. 3, p. 1044, 2024.
- [10] S. Larabi-Marie-Sainte, "Outlier detection based feature selection exploiting bio-inspired optimization algorithms," *Applied Sciences*, vol. 11, no. 15, p. 6769, 2021.
- [11] W. Zhao, D. Zhang, D. Li, Y. Zhang, and Q. Ling, "Optimized gicp registration algorithm based on principal component analysis for point cloud edge extraction," *Measurement and Control*, vol. 57, no. 1, pp. 77–89, 2024.

- [12] Z. Zhang, H. Song, J. Fan, T. Fu, Q. Li, D. Ai, D. Xiao, and J. Yang, "Dual-correlate optimized coarse-fine strategy for monocular laparoscopic videos feature matching via multilevel sequential coupling feature descriptor," *Computers in Biology and Medicine*, p. 107890, 2023.
- [13] Y. Xia and J. Ma, "Locality-guided global-preserving optimization for robust feature matching," *IEEE Transactions on Image Processing*, vol. 31, pp. 5093–5108, 2022.
- [14] V. Kalaiyarasi, S. Jain, S. Jain, R. Umapiya, R. Sarala, M. S. Ramkumar *et al.*, "Bio-inspired optimization technique for feature selection to enhance accuracy of bc detection," in *2023 International Conference on Inventive Computation Technologies (ICICT)*. IEEE, 2023, pp. 741–748.
- [15] N. Mohd Yusof, A. K. Muda, S. F. Pratama, and A. Abraham, "A novel nonlinear time-varying sigmoid transfer function in binary whale optimization algorithm for descriptors selection in drug classification," *Molecular diversity*, vol. 27, no. 1, pp. 71–80, 2023.
- [16] S. Ghosh, S. K. Hassan, A. H. Khan, A. Manna, S. Bhowmik, and R. Sarkar, "Application of texture-based features for text non-text classification in printed document images with novel feature selection algorithm," *Soft Computing*, pp. 1–19, 2022.
- [17] S. Veerashetty, Virupakshappa, and Ambika, "Face recognition with illumination, scale and rotation invariance using multiblock ltp-glm descriptor and adaptive ann," *International Journal of System Assurance Engineering and Management*, pp. 1–14, 2022.
- [18] J. Syme, J. J. Kiszka, and G. J. Parra, "How to define a dolphin "group"? need for consistency and justification based on objective criteria," *Ecology and Evolution*, vol. 12, no. 11, p. e9513, 2022.
- [19] H. J. Kriesell, S. H. Elwen, A. Nastasi, and T. Gridley, "Identification and characteristics of signature whistles in wild bottlenose dolphins (*tursiops truncatus*) from namibia," *PloS one*, vol. 9, no. 9, p. e106317, 2014.

Empirical Analysis of Variations of Matrix Factorization in Recommender Systems

Srilatha Tokala¹, Murali Krishna Enduri², T. Jaya Lakshmi³, Koduru Hajarathaiyah⁴, Hemlata Sharma⁵

Department of CSE, SRM University-AP, India^{1,2}

Department of Computing, Sheffield Hallam University, United Kingdom^{3,5}

School of Computer Science and Engineering, VIT-AP University, India⁴

Abstract—Recommender systems recommend products to users. Almost all businesses utilize recommender systems to suggest their products to customers based on the customer's previous actions. The primary inputs for recommendation algorithms are user preferences, product descriptions, and user ratings on products. Content-based recommendations and collaborative filtering are examples of traditional recommendation systems. One of the mathematical models frequently used in collaborative filtering is matrix factorization (MF). This work focuses on discussing five variants of MF namely Matrix Factorization, Probabilistic MF, Non-negative MF, Singular Value Decomposition (SVD), and SVD++. We empirically evaluate these MF variants on six benchmark datasets from the domains of movies, tourism, jokes, and e-commerce. MF is the least performing and SVD is the best-performing method among other MF variants in terms of Root Mean Square Error (RMSE).

Keywords—Recommendations; matrix factorization; content-based; collaborative filtering; RMSE

I. INTRODUCTION

A large number of websites offer products to their users. Users purchase products based on a variety of necessities and tastes. By giving customers the best products, one can help to accelerate the purchasing process and raise customer contentment. As the state of technology advances at a rapid pace, it becomes increasingly difficult to anticipate user preferences and meet their requirements. Recommendations are quite useful in many aspects of our daily lives. We employ some external features to learn and make choices about a user's preferences for a specific product [1]. The recommender system is developed to address this issue. These systems learn from user actions and preferences to predict what content may most likely catch the user's interest [2]. Many commercial sites like YouTube, Amazon, and Netflix are highly benefited by using highly sophisticated recommender systems. Potential applications include suggesting books on Amazon, movies on Netflix, products on Flipkart, and so on. These sites continuously monitor the user's watch/view/purchase history and attempt to make educated guesses about what other products the user might find interesting. Many times, systems ask users to provide explicit ratings on used products. This rating information is a significant input to the recommender systems [3].

In 1979, a computer-based librarian introduced the first iteration of the recommender system to offer customers advice on what books to read. Then it advanced in the 1990's with a lot of research achievements in various fields. A research lab Group Lens at the University of Minnesota in the United

States launched another recommender system implementation in the 1990's to assist people [4]. After that, they started calling it a Group Lens Recommender System. The use of recommender systems in advancing research across disciplines and sectors has grown in recent years. Recommender systems are essential components of many online platforms, providing users with tailored content and product suggestions. These systems significantly boost user engagement and satisfaction by forecasting user preferences through historical data and behavior analysis [5]. However, despite their extensive use, traditional recommender systems encounter issues with accurately modeling preferences, ensuring fairness, mitigating bias, and maintaining transparency. Researchers have been exploring advanced methodologies to overcome these obstacles and enhance the effectiveness and dependability of recommender systems.

Causal inference methods are essential for uncovering the fundamental causes of user preferences and behaviors, thereby improving the precision and dependability of recommendations in recommender systems [6]. By leveraging these techniques, effective recommendation algorithms can greatly enhance user satisfaction through personalized content that matches individual interests and preferences. Nonetheless, selection bias remains a significant challenge, as it can result in biased and inaccurate recommendations by disproportionately representing certain user groups or preferences based on skewed data [7]. Achieving fairness in these systems is vital to ensure that all users receive equitable recommendations, regardless of their demographics or past behaviors. The integration of large language models into recommender systems can further refine the understanding of user context and intent, resulting in more sophisticated and effective recommendations [8]. Combining these advanced methodologies helps to mitigate issues of bias and fairness, ultimately improving the overall performance and trustworthiness of recommender systems [9].

There are many applications of recommender systems. Various recommender system techniques are proposed for a variety of applications related to Government, Business, Online Shopping, Library, Learning, Tourism, Group activities, and Healthcare.

1) *Government*: The government may greatly improve its communication with its constituents and its ability to serve the public by adopting the internet recommender system. The citizen services discover and recommend to users more significant and interesting services. One-time items will receive ratings from the business perspective services [10]. By considering the citizen's profile, more relevant and engaging services to

the citizen are recommended. In business perspective, one-time items will receive ratings from the business perspective.

2) *Business*: Various recommender systems are developed for business promotions. Some of the systems pay attention to the recommendations initiated by individual customers that are Business-to-Customer (B2C) systems. Recommendations produced for business users on products and service is called Business-to-Business (B2B) systems.

3) *Online shopping*: One of the most significant tools in the realm of online purchasing is the recommender system [11]. Ratings for the purchased products by a user is the primary information that depicts the interest of the user [12]. Almost all commercial applications like Amazon, Netflix, and Flipkart provide the option for giving ratings for the products.

4) *Library*: To propose resources for research in the university's online libraries, Porcel *et al.* conducted research and created a recommender system [13], [14]. Applications for online libraries can employ systems of recommendations to help users search and select knowledge and data resources.

5) *Learning*: Learning recommender systems guide learners to select the subjects, courses, and learning information to perform learning activities [15]. Digital libraries contain a huge amount of e-documents that a user can choose from [16].

6) *Tourism*: Recommender Systems are also used to give recommendations for tourism places to tourists. It is mainly suggested on transportation, restaurants, and lodging for users to feel comfortable reaching their destinations. Users are directed to a wide range of online resources. These services contain different perspectives according to videos, music, and learning materials that are uploaded by users [17].

7) *Group activities*: As interactions through online have increased, the use of group activities has become more popular. Giving recommendations to a group of users having different opinions is a crucial task. The idea of group activities is to learn interactions between the users from the known group ratings [18]. In various cases, the decision has to be made by the users in both online or in without internet access. In such instances, the entire organization makes the decision to balance users' expectations in online as well as offline formats. The online group is to be formed by the system, but the offline group will be already formed [19].

8) *Healthcare*: Efficient and effective communication is very important in healthcare. A growing number of patients require the care of healthcare professionals from many specialties, especially those who have chronic illnesses or diseases [20].

There are many other applications in the fields of medicine, banking, telecom, media, social networks, e-commerce, internet of things (IoT) [21], [22] other than the ones already mentioned.

In Section II a brief analysis of the problem statement is discussed. In Section III we discussed about the taxonomy of recommender systems followed by in Section IV, a discussion on matrix factorization techniques is mentioned. In Section V the empirical evaluation of the datasets and the rating distribution plots are discussed. Section VI discusses about the results and discussion and in Section VII gives a brief

comparison of different matrix factorization methods. Finally, Section VIII ends up with final conclusion and future plan. The supplementary information for the results is placed in Section VIII-B.

II. PROBLEM STATEMENT

Consider a set of m users $U = \{U_1, U_2, \dots, U_m\}$ and a group of n items $I = \{I_1, I_2, \dots, I_n\}$ and a rating matrix R of size $m \times n$, R_{ij} denotes the rating provided by the user U_i to item I_j . Making user recommendations for unrated items presents a challenge.

The problem of recommender systems is depicted in Fig. 1.

	I_1	I_2	I_3	I_4	I_5	I_6
U_1	4	?	5	?	3	?
U_2	?	2	2	?	4	?
U_3	3	?	?	2	?	1
U_4	?	4	?	5	?	3
U_5	?	?	4	3	?	3
U_6	3	?	?	3	4	?
U_7	4	5	?	4	3	?

Fig. 1. Example for recommender systems.

This correlates to the issue of matrix completion, which is to fill the empty cells of R with rating information from the filled matrix entries. This problem is challenging because the real-world rating matrices are huge in size and sparse in nature. For instance, Amazon product recommendation is represented as a matrix containing around 197 million users and 12 million products. As a single user may not rate many products, the product vector of the user has more empty cells than ratings. Hence, it is extremely challenging to predict recommendations for the next user action based on these fewer interactions and it is called a data sparsity problem.

Handling sparse rating data in recommendation systems is challenging due to extreme sparsity, where missing data makes it hard to find patterns and leads to less accurate predictions. Overfitting occurs as models may capture noise from sparse data, reducing their ability to generalize. The cold start problem also arises when new users or items lack enough interaction history, limiting accurate recommendations. Solutions include regularization, hybrid models, and using implicit feedback or additional information to address these issues.

There are numerous methods of solving matrix completion issues. One of the traditional methods for matrix completion is matrix factorization (MF). This study focuses on solutions based on MF. Table I provides the notations utilised in this work.

Next section describes the existing methods of recommendation.

TABLE I. NOTATIONS

Notation	Usage
R	Rating Matrix
m	Number of Users
n	Number of Items
i	user
j	item/product
X	Latent features for the Users
Y	Latent features for the Items
k	Number of features extracted
U	Set of Users
I	Set of Items
δ	Regularization Constant
γ	Learning Rate
\hat{R}	Prediction rating
β	Distribution Parameter Set
α	Distribution Hyper parameter
P, Q	Orthogonal Matrices
s	Singular Matrix

III. LITERATURE

Information filtering in recommender systems involves choosing and displaying relevant information or items for users based on their preferences, behaviors, and interactions [23]. This entails processing large amounts of data to find and present content, products, or services that are most likely to appeal to the user [24]. By providing tailored suggestions that suit each user’s particular preferences and needs, the goal is to enhance the user experience. There are many classifications of recommender systems in practice. Fig. 2 shows a popular classification.

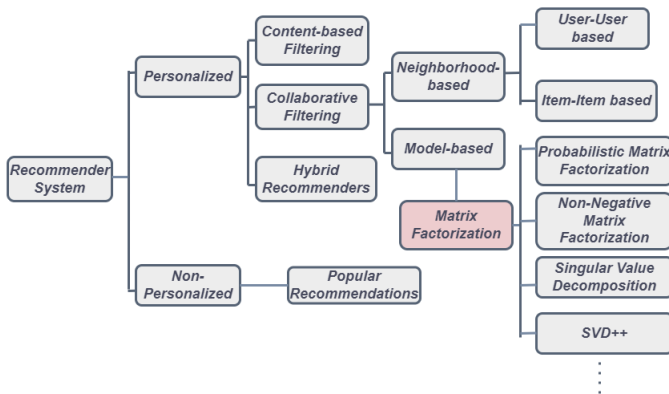


Fig. 2. Classification of recommender systems.

Non-personalized and personalised recommender systems are the two main classifications of recommender systems. Non-personalized recommenders show users only the most popular items, regardless of their purchases/interests. Based on their purchases and reviews, personalised recommenders analyse the tasks of users and make pertinent product recommendations. For instance, suggesting the most popular web series such as *MoneyHeist* being recommended to all Netflix subscribers irrespective of their genre choice can be regarded as a non-personalized recommendation. On the other hand, suggesting movies/TV shows belonging to a genre that the user watches/rates more often is a personalized recommendation.

Additional categories for personalised recommender sys-

tems include content-based filtering, collaborative filtering, and hybrid models. Content-Based (CB) Filtering utilises the item attributes in recommendation. The algorithm creates a list of products with characteristics comparable to those of products the customer has already bought or reviewed. A few items from the list with top similarity will be recommended to the user. For this purpose, metrics such as cosine, euclidean, pearson, or spearman are used [25]–[28].

An example showing content-based filtering is depicted in Fig. 3.

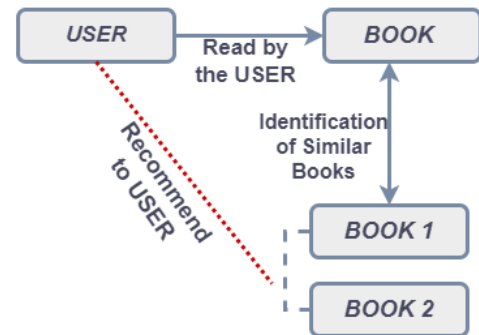


Fig. 3. Content-based filtering.

The recommender system finds additional books (let’s say BOOK1 and BOOK2) that are comparable to the one the USER has already read and recommends those to the USER.

The content-based recommendation requires domain knowledge to identify attributes that may be non-available due to privacy concerns. This is a major limitation of this class of recommendation systems. However, CB addresses the cold start problem effectively. Collaborative Filtering (CF) suggests items/products based on the user’s previous choices [29]–[34]. For a specific user i , CF finds additional users who share i ’s preferences and makes suggestions based on their choices. Their interactions with various products that user i purchased/rated can be used to determine if they have comparable tastes. Collaborative filtering key benefit is that it doesn’t require domain knowledge [35]. The process of CF for book recommendation is given in Fig. 4.

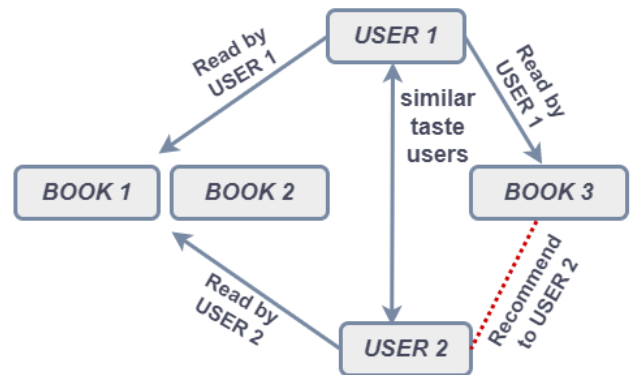


Fig. 4. Collaborative filtering.

In this example, there are two different users (say USER1 and USER2) who are having similar tastes. If there are two

different books (say BOOK1 and BOOK2) that are read by both users, the recommender system identifies the books that are read by only one user (say USER1) and recommends the remaining books (say BOOK3) that are not read to the other user (say USER2).

The two approaches of calculating user similarity are Neighborhood-based and Model-based methods.

Neighborhood-based Collaborative Filtering (NCF) methods are also called Memory-based models or Heuristic-based models. The NCF technique forecasts the similarity between users and items by analyzing user-item interactions through heuristics. It employs two approaches: user-user collaborative filtering and item-item collaborative filtering [36].

- User-User based Collaborative Filtering: When producing predictions, the user-based collaborative filtering finds the other users who are engaging in similar behaviors. For user's having similar interactions, the items are recommended. It predicts the interest of an item that depends on the rating information from similar users [37]. The steps to compute user-user similarities are given below:
 - Build a user vector A_i for each individual user i . A_i will be of size n , n being the number of items. $A_i[j]$ is 1 if user buys item j , otherwise it is zero.
 - Compute the similarity matrix M , of size $m \times m$ where m is the number of users such that $M[i_1, i_2] = \text{similarity}(i_1, i_2)$.
 - For every user i , identify a set of users $S \in U$, where S contains the users with top similarity score with i .
 - Suggest the items that are bought by the users in S , and that are not bought by i .

The example of user-user based collaborative filtering is depicted in Fig. 5.

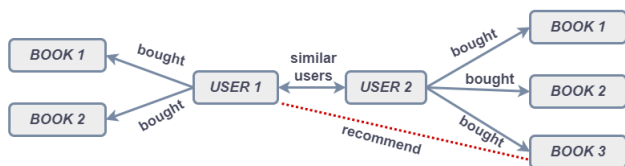


Fig. 5. User-user based collaborative filtering.

To recommend books to USER1, the recommender computes the similar users of USER1, which is USER2 in the first step. BOOK3 bought by USER2 is not bought by USER1, and is recommended to USER1.

Item-item based Collaborative Filtering (ICF): By detecting the associated subjects that users have previously rated, the item-item based collaborative filtering determines how similar the items are and provides predictions. It computes the similarity of how the target item is selected from the k-most similar items [38]. Additionally, the corresponding parallels are found. When comparable things are discovered, the prediction is made using the target user's rating as well as the average of the related items. The following are the steps to compute item-item similarities:

- Find the previously liked items of the target user from the historical data.
- Identify the most similar items for the previously liked items.
- Select the maximum likelihood items from similar item sets.
- Introduce the products to the target user.

The example of item-item based collaborative filtering is depicted in Fig. 6.

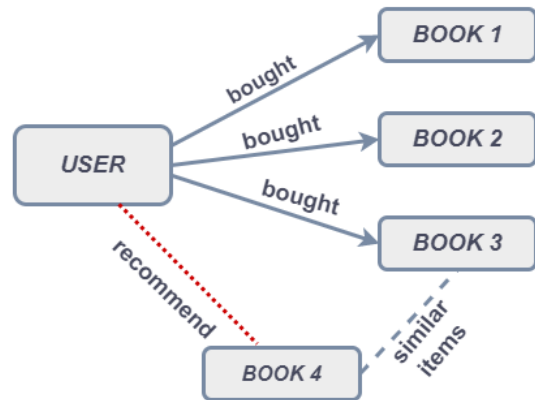


Fig. 6. Item-item based collaborative filtering.

To suggest books to a USER, the recommender system computes the maximum likelihood of similar books (say BOOK4) bought by the USER from the historical data and recommends them to the USER.

Model-based Collaborative Filtering trains a model using historical information on user-item ratings. Once the model is trained, ratings can be predicted using the model [39]. One of the well-liked model-based techniques is *Matrix Factorization (MF)*. The goal of this study is to compare and assess the performance of several MF approaches in different areas. The next section discusses MF and its variations in detail.

IV. MATRIX FACTORIZATION TECHNIQUES

In this section, various matrix factorization methods are discussed. In every matrix factorization method, ratings are predicted and the recommendations are given to the users. So, the evaluation metric for our analysis is limited to Root Mean Square Error (RMSE).

The basic procedure of MF is shown in Fig. 7.

Further, five variations of MF techniques namely matrix factorization (MF), probabilistic matrix factorization (PMF), non-negative matrix factorization (NMF), singular value decomposition (SVD), and SVD++ are elaborated. Each of the variant addresses the challenges of collaborative filtering like cold start problem, and data sparsity in different ways.

MF addresses data sparsity by decomposing the user-item interaction matrix into lower-dimensional user and item matrices, capturing latent factors that help to estimate missing values

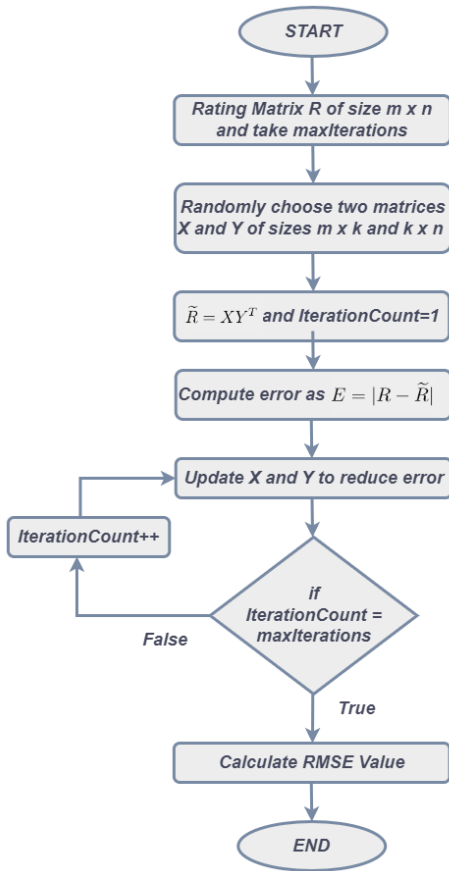


Fig. 7. Flow chart for the procedure of Matrix Factorization (MF).

and fill gaps created by unobserved ratings. However, MF faces challenges with the cold start problem, as it depends on historical interactions to learn these latent factors, making it difficult to generate accurate recommendations for new users or items without prior rating data. PMF extends MF with a probabilistic framework that regularizes factorization, helping to manage sparse data. While PMF also faces cold start challenges due to reliance on historical data, Bayesian approaches can mitigate this by incorporating priors on latent factors. NMF decomposes the matrix into non-negative latent factors, capturing additive relationships and handling sparse data more effectively. However, it still relies on sufficient observed data and struggles with cold start, though variations incorporating content-based data can help address this limitation. SVD factorizes the matrix into orthogonal components, capturing key features with reduced dimensions and approximating missing values through low-rank approximations. However, it requires observed data for decomposition, making it less effective for cold start scenarios, as it lacks a direct mechanism for handling users or items without prior interactions. SVD++ extends standard SVD by incorporating both explicit ratings and implicit feedback, such as clicks and views, which helps mitigate data sparsity by providing additional data points. While it improves cold start handling for users through implicit feedback, it still requires some interaction data and remains limited for completely new users or items.

A rating matrix R is of size $m \times n$ is input to any MF

method, where m, n are the number of users and items. R is factored into two latent feature matrices X and Y [40], [41].

Three steps are common in these MF methods:

- 1) Initialization of latent feature matrices X and Y : It is a common practice to initialize the matrices X and Y randomly. Different MF techniques use different data distributions to generate these random numbers.
- 2) Computation of predicted rating matrix: The predicted rating matrix \tilde{R} is computed by extracting k users and items latent features. The value of k can be fixed empirically. These latent feature matrices are multiplied to get the overall predicted matrix. The sample process is shown below.

$$\begin{matrix} \begin{bmatrix} r_{11} & \dots & r_{1n} \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ r_{m1} & \dots & r_{mn} \end{bmatrix} & = & \begin{bmatrix} x_{11} & \dots & x_{1k} \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ x_{m1} & \dots & x_{mk} \end{bmatrix} & \begin{bmatrix} y_{11} & \dots & y_{1n} \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ \cdot & \dots & \cdot \\ y_{k1} & \dots & y_{kn} \end{bmatrix} \\ m \times n & & m \times k & k \times n \end{matrix}$$

Where, k is the number of features extracted, m is the users, n is the items, X is a matrix representing latent features of the users, Y denotes the latent features of the items. The relation between R and \tilde{R} is shown in Eq. 1.

$$R \approx XY^T \quad (1)$$

$$\tilde{R} = XY^T$$

The error in the prediction is determined using the difference between corresponding cells in R and \tilde{R} after computing \tilde{R} . A few common error metrics include Mean Absolute Error (MAE), Regularised Square Error (RSE), and Root Mean Square Error (RMSE), as illustrated in Eq. 2, Eq. 3, and Eq. 4, respectively.

Eq. 2 calculates Root Mean Square Error (RMSE) by subtracting the original rating from the predicted rating values.

$$RMSE = \sqrt{\frac{1}{N} \sum (r_{ij} - \tilde{r}_{ij})^2} \quad (2)$$

where N is the number of predictions, \tilde{r}_{ij} is the predicted rating, and r_{ij} is the original rating.

According to Eq. 3, the Regularised Square Error (RSE) is produced by subtracting the original rating from the predicted rating values and adding regularisation factors.

$$RSE = (r_{ij} - \tilde{r}_{ij})^2 + \delta \quad (3)$$

Mean Absolute Error (MAE) is calculated as shown in Eq. 4, by subtracting the original rating's absolute value from the anticipated rating values.

$$MAE = \frac{1}{|N|} (|r_{ij} - \tilde{r}_{ij}|) \quad (4)$$

Minimising the discrepancy between the actual and predicted rating matrices is the key job here.

In general, an objective function shown in Eq. 5 is used for that task.

$$\min \frac{1}{2} \|R - XY\|^2 \quad (5)$$

The updating of the latent user and item matrices, which are covered below, is necessary for the goal function.

- 1) Update the latent feature matrices X and Y : Different update rules are used by MF methods to reduce the error computed in step 2.
- 2) Step 3 is repeated until either error doesn't remain the same in two successive steps or the error is less than a chosen threshold. But in most of the programming solutions, a fixed number of iterations is taken as a terminating point.

To summarise, different MF techniques vary in steps 1 (Initialization), and 3 (update rule). The following sections describe the variations in detail.

A. Matrix Factorization

1) *Initialization of latent feature matrices*: The initialization of X and Y are purely random values with 0 to 1 distribution in basic MF [42], [43].

2) *Update rule to reduce the error between actual and predicted rating matrices*:: By multiplying the rating vectors for the person and the object, as stated in Eq. 6, one can find the original rating.

$$r_{ij} \approx x_i y_j^T \quad (6)$$

Utilising the observed ratings while reducing the squared error is one method for computing the empty ratings in the matrix. The square error minimization is shown in Eq. 7.

$$\min \sum_{i,j} (r_{ij} - x_i y_j^T)^2 \quad (7)$$

The result will overfit the training data and to overcome the overfitting in squared error a regularization term is incorporated is shown in Eq. 8. Regularization is controlled by using a regularization constant δ known as Regularized Square Error (RSE).

$$\min \sum_{i,j} (r_{ij} - x_i y_j^T)^2 + \delta (\|x_i\|^2 + \|y_j\|^2) \quad (8)$$

where $\|\cdot\|$ is the frobenius norm. Alternating least squares or stochastic gradient descent can be used to estimate this value. According to Eq. 9, every rating within the training data has been predicted via stochastic gradient descent, and the prediction error is calculated.

$$e_{ij} = r_{ij} - x_i y_j^T \quad (9)$$

Then update the vectors y_j and x_i with a constant γ called the learning rate, and δ as the regularization constant. Updating the values of y_j and x_i is shown in Eq. 10.

$$\begin{aligned} y_j &\leftarrow y_j + \gamma(e_{ij}x_i - \delta y_j) \\ x_i &\leftarrow x_i + \gamma(e_{ij}y_j - \delta x_i) \end{aligned} \quad (10)$$

B. Probabilistic Matrix Factorization

1) *Initialization of latent feature matrices*: The initialization of X and Y are random values of 0 to 1 with normal distribution.

2) *Update rule to reduce the error between actual and predicted rating matrices*: By fixing the parameters, the log-posterior value on the predicted rating matrix \tilde{R} is observed from the original rating matrix R [44]. To maximize the log-posterior for the user's and item's latent features, some additional regularization hyperparameters are added and fixed to minimize the sum of squares as shown in Eq. 11.

$$\begin{aligned} E = & -\frac{1}{2} \left[\sum_{i=1}^m \sum_{j=1}^n (r_{ij} - x_i^T y_j)_{(i,j) \in \Omega_{R_{ij}}}^2 \right] \\ & - \frac{1}{2} \left[\delta_X \prod_{i=1}^m \|x_i\|_{Fro}^2 + \delta_Y \prod_{j=1}^n \|y_j\|_{Fro}^2 \right] \end{aligned} \quad (11)$$

where

$$\delta_X = \frac{\sigma_X^2}{\sigma^2}, \delta_Y = \frac{\sigma_Y^2}{\sigma^2}$$

The procedure for calculating the log-posterior distribution is as follows: An approach that offers a statistical framework using the Bayes theorem for the model rating matrix R called Probabilistic Matrix Factorization (PMF) which is proposed by Salkhutinov and Mnih [45]. PMF is a probabilistic linear model with gaussian distribution which is used for initial latent feature matrices X and Y . By fixing the parameters, the log-posterior value on the predicted rating matrix \tilde{R} is observed from the original rating matrix R [44].

$$p(\beta|Z, \alpha) = \frac{p(Z|\beta, \alpha)p(\beta|\alpha)}{p(Z|\alpha)} \propto p(Z|\beta, \alpha)p(\beta|\alpha) \quad (12)$$

In this case, Z represents dataset, β represents the distribution parameter set, and α represents the distribution hyper parameter. The posterior distribution, also known as a-posteriori, is denoted by $p(\beta|Z, \alpha)$. $p(Z|\beta, \alpha)$ is the likelihood and $p(\beta|\alpha)$ is the prior. More information about the data distribution can be obtained through the training process, and the model parameter β can be adjusted to fit the data. Let R_{ij} represent the rating of the user i on the item j and let $X \in R^{m \times k}$ and $Y \in R^{k \times n}$ are the users and items latent feature vectors respectively. Here we assume that the entries of R are normally distributed around the inner product of (X_i, Y_j)

with a common variance. We will now use our rating matrix for the predictions.

$$\beta = \{X, Y\}, \quad Z = R, \quad \alpha = \sigma^2$$

where σ^2 is the variance of the Gaussian distribution. We get this by substituting these values in Eq. 12.

$$p(X, Y|R, \sigma^2) = p(R|X, Y, \sigma^2)p(X, Y|\sigma_X^2, \sigma_Y^2) \quad (13)$$

In Eq. 13, X and Y values are independent of each other, and hence the equation can be rewritten as shown in Eq. 14.

$$p(X, Y|R, \sigma^2) = p(R|X, Y, \sigma^2)p(X, \sigma_X^2)p(Y, \sigma_Y^2) \quad (14)$$

Let I_{ij} be defined as the likelihood of R entries such that if the value is 1, the entry is observed, and if the value is 0 the entry is not observed. Adopt a gaussian-distributed probabilistic linear model and specify the conditional probability across the observed ratings as per Eq. 15.

$$p(R|X, Y, \sigma^2) = \prod_{i=1}^m \prod_{j=1}^n [N(r_{ij}|x_i^T y_j, \sigma^2)]^{I_{ij}} \quad (15)$$

The new assumption about the likelihood is that R 's entries are independent, every entry has a normal distribution, and entries all have the same variance σ^2 .

The prior distributions of X, Y are shown in Eq. 16 and Eq. 17.

$$p(X|\sigma_X^2) = \prod_{i=1}^m N(x_i|0, \sigma_X^2) \quad (16)$$

$$p(Y|\sigma_Y^2) = \prod_{j=1}^n N(y_j|0, \sigma_Y^2) \quad (17)$$

In these priors we assume that X and Y rows are correlated, every entry has a normal distribution, and entries all have the same variance σ^2 .

Replacing Eq. 15, Eq. 16 and, Eq. 17 in Eq. 14 we get,

$$p(R|X, Y, \sigma^2) = \prod_{i=1}^m \prod_{j=1}^n [N(r_{ij}|x_i^T y_j, \sigma^2)]^{I_{ij}} \prod_{i=1}^m N(x_i|0, \sigma_X^2) \prod_{j=1}^n N(y_j|0, \sigma_Y^2) \quad (18)$$

For training our model, we apply logarithms on both sides of Eq. 18 and then apply derivatives on both sides of the equation. Then the expression for log-posterior is like

$$\ln p(X, Y|R, \sigma^2) = -\frac{1}{2\sigma^2} \prod_{i=1}^m \prod_{j=1}^n I_{ij} (r_{ij} - x_i^T y_j)^2 - \frac{1}{2\sigma_X^2} \prod_{i=1}^m \|x_i\|_{Fro}^2 - \frac{1}{2\sigma_Y^2} \prod_{j=1}^n \|y_j\|_{Fro}^2$$

Here, the Fro suffix is called the Frobenius norm is given by

$$\|x\|_{Fro}^2 = x^T x$$

C. Non-Negative Matrix Factorization

There are many applications in which the data is analyzed to be non-negative, and many of the tools follow this property [46]. The idea of NMF, which forces the data to be non-negative, gave rise to the need for low-rank approximation for development. NMF is used as a tool for the analysis of high-dimensional with non-negative entries in the data. NMF should consist of only non-negative constraints as a part of the representation. There are different variants of NMF algorithms proposed [47] and we are using basic NMF. NMF was introduced with the name positive matrix factorization by Paatero and Tapper [48]. Researchers paid attention to NMF after the work given by Lee and Sung. They discussed more on usage and importance of NMF [49].

1) *Initialization of latent feature matrices:* The initialization of latent feature matrices X and Y are taken as non-negative random values.

2) *Update rule to reduce the error between actual and predicted rating matrices:* Finding an estimate of a non-negative matrix R , which is represented as the product of latent feature matrices X and Y , is the primary goal of NMF.

$$R \approx XY$$

where R is a $m \times n$ rating matrix. The approximation of R with product of matrices X ($m \times k$) and Y ($k \times n$) by considering $k \leq (m, n)$. The latent feature matrices X and Y , can be derived by using multiplicative update method that consist of some update rules. The rules are explained in [49]. The non-negativity property is maintained by both matrices X and Y .

$$\begin{aligned} X &= X \cdot \left((R \cdot / (X \times Y + (R == 0))) \times Y^T \right) \\ Y &= Y \cdot \left(X^T \times (R \cdot / (X \times Y + (R == 0))) \right) \end{aligned}$$

Similarly, the dot division of X and Y is $X \cdot / Y$, where $X \cdot \times Y$ is element-wise division calculated as the dot product of X and Y . The product of two matrices X and Y is $X \times Y$. The transposed version of the matrix X is X^T . In the denominator, the expression $R == 0$ is used to avoid division by zero.

The properties of NMF are only non-negative values are allowed into the resultant matrix, since non-negative values are only allowed, the matrix is allowed to only add but not subtract, and the result of the factorization is not unique.

D. Singular Value Decomposition

The singular value decomposition (SVD) method was first applied for recommender systems [50]. In MF using SVD, the rating matrix R decomposes into three latent feature matrices P , s , and Q , as shown in Eq. 19 where the rating matrix R is of size $m \times n$ and the latent factor matrices P is of size $m \times m$, s is of size $m \times n$, and Q is of size $n \times n$. Here, P and Q are orthogonal matrices, and s is a singular matrix. The latent feature matrices X and Y are computed as follows: $X=P.s$ and $Y= Q$ [51].

$$R \approx PsQ^T \quad (19)$$

Better performance is achieved by using the SVD approach on its applications that reduces the dimensionality of user and item matrix [52].

E. SVD++

The main purpose of SVD++ is to identify the missing ratings in the matrix by adding implicit feedback to the user's latent feature matrix [53]. This technique is observed to be more accurate in many cases, because of including implicit feedback to user latent feature matrix [54].

The implicit feedback matrix UV is calculated as follows: The matrix U is calculated as $U = [u_{ij}] \forall (i, j)$ is 1 if R_{ij} is present in the original rating matrix 0 otherwise. For every non-zero entry in the matrix i^{th} row is written as $\frac{1}{\sqrt{|I_i|}}$ and is an $m \times n$ matrix. V matrix is calculated as $V = [v_{ij}] \forall (i, j)$ which is same as an item feature matrix of order $n \times k$. Calculate the dot product of matrices UV of order $m \times n$ and $n \times k$, resulting in a matrix of the order of $m \times k$. The implicit feedback matrix UV is to be added to X (say $X = X + UV$) before performing the dot product of X and Y^T . V matrix is assigned to Y (say $Y = V$).

The variations of different MF methods used in this work are tabulated in Table II.

V. EMPIRICAL EVALUATION

The era of each dataset, where it was downloaded from, and the information that is contained in it are all fully described in this section. Exploratory data analysis reveals a full description of the dataset ratings. additionally explains the setup for conducting the analysis and arriving at the RMSE value.

A. Datasets

The primary goal of utilizing the datasets is to run the simulations. The datasets are downloaded from Kaggle, Konekt, and Github. Datasets namely Movie Lens-100K, Movie Lens-1M, Film Trust, Trip Advisor, Jester, and Market datasets are taken. Each dataset contains information about the users, items, ratings, and timestamps.

The period of Movie Lens-100K dataset is from September 19th, 1997 to April 22nd,1998 [55]. The period of Movie Lens-1M dataset is on December 2015 [56]. The period of the Film Trust dataset is on 2011 [57]. The period of Trip Advisor dataset is from March 3rd, 2001 to November 1st,

2009 [58]. The period of the Jester dataset is from November 2006 to November 2012 [59]. The period of Market dataset is from January 1st, to April 30th, 2021 [60]. Table III tabulates the dataset statistics for numerous datasets.

B. Exploratory Data Analysis

In Fig. 8, the rating distribution plots for several datasets, including Movie Lens -100K, Movie Lens -1M, Film, Trip Advisor, Jester, and Market, are displayed. The rating distribution for films in the MovieLens-100K dataset ranges from 1 to 5. In the dataset, the users are the individuals, and the items are the films. A low rating of one is provided by users 6110 (6.11%) beyond 100000 ratings, while users 34174 (27.14%) offer a rating of four for films. The rating distribution in the movie lens-1M dataset ranges from 0.5 to 5. In the dataset, the users represent individuals, while the items refer to movies. It has been noted that users 306221 (29.20%) have given films a rating of four stars, while users 8559 (0.81%) have given films a poor rating of 0.5 out of 1048576.

According to Fig. 8, the rating distribution for the film trust dataset is between 0.5 and 4. The films in the dataset are the items, and the users are the people. It has been noted that out of 35494 ratings, users 1060 (2.98%) give films a low rating of 0.5 and users 9170 (25.83%) offer films a rating of five. The 1 to 5 rating distribution is part of the Trip Advisor dataset. The hotels in the dataset are the items, and the users are the individuals. Users 10082 (5.73%) out of 175765 ratings offer a low rating of one for hotels, while users 77668 (44.18%) provide a rating of five for hotels.

The rating distribution for the jester dataset is between -10 and $+10$, as shown in Fig. 8. In the dataset, the users are the individuals, and the objects are the jokes. Out of the 1048575 ratings, it is noted that over 4000 people have given ratings of 10 or above. The rating distribution in the market dataset ranges from 1 to 5. In the dataset, the users are the individuals and the items are the things. It has been noted that out of 1048575 users' ratings, users 609417 (58.11%) give items a rating of five, while users 63082 (6.01%) give products a poor rating of two.

VI. RESULTS ANALYSIS

To forecast the missing values in the rating matrix, we apply five different MF methods such as MF, PMF, NMF, SVD, and SVD++ with different latent features on six data sets which are given in Section IV and Table III. For all methods, we computed RMSE value for different latent features $k = 10, 50, 100, 200, 300, 400$ with 10 steps. In this section, we have shown patterns of RMSE values with different latent features for various MF methods on six datasets. We have shown other errors (RSE, RMSE, and MAE) for latent features $k = 1, 2, \dots, 10$ with 100 steps in supplementary information as shown in Section VIII-B.

Fig. 9 describes RMSE value on six different datasets for the MF method with $k = 10, 50, 100, 200, 300, 400$ at different steps. In movie lens-1M, film trust, and jester datasets, if k -value is increasing there is a constant range of RMSE maintained. In the movie lens-100K dataset, it is observed that as k -value is increasing there is a constant range of RMSE for k values 10, 50, 100, 200 and the RMSE decreases for k values

TABLE II. DIFFERENCES BETWEEN DIFFERENT VARIATIONS OF MATRIX FACTORIZATION (MF) METHODS

Method	Initialization of Latent Features	Update of Latent Features
MF	Latent feature matrices X and Y are taken as random values between 0 and 1.	Update the parameters of X and Y by adding regularization constant and learning rate to minimize the error.
PMF	Latent feature matrices X and Y are taken as normal distribution random values between 0 and 1.	Update the parameters of X and Y by adding regularization hyperparameters to minimize the error.
NMF	Latent feature matrices X and Y are taken as non-negative random values.	Apply the rules of the multiplicative update method to update the parameters of X and Y .
SVD	Latent feature matrices X and Y are taken as floating point numeric dtype random values.	No update rule is used.
SVD++	Latent factor matrices X and Y are taken by adding an implicit feedback matrix to the user latent feature matrix X .	Adding an implicit feedback matrix to the user latent feature matrix X is itself an update that is performed.

TABLE III. INSIGHTS INTO SIX DIVERSE DATASETS: MOVIE LENS-100K, MOVIE LENS-1M, FILM TRUST, TRIP ADVISOR, JESTER, AND MARKET DATASETS

Dataset	Users	Items	Ratings	Rating Range	Average Rating	Sparsity
Movie Lens-100K	943	1682	100000	1-5	3.529	0.937
Movie Lens-1M	7848	65133	1048576	0.5-5	3.522	0.998
Film Trust	1508	2071	35494	0.5-4	3.002	0.988
Trip Advisor	145316	1759	175765	1-5	4.000	0.999
Jester	31958	140	1048575	-10 - +10	0.955	0.839
Market	941860	9849	1048575	1-5	4.062	0.999

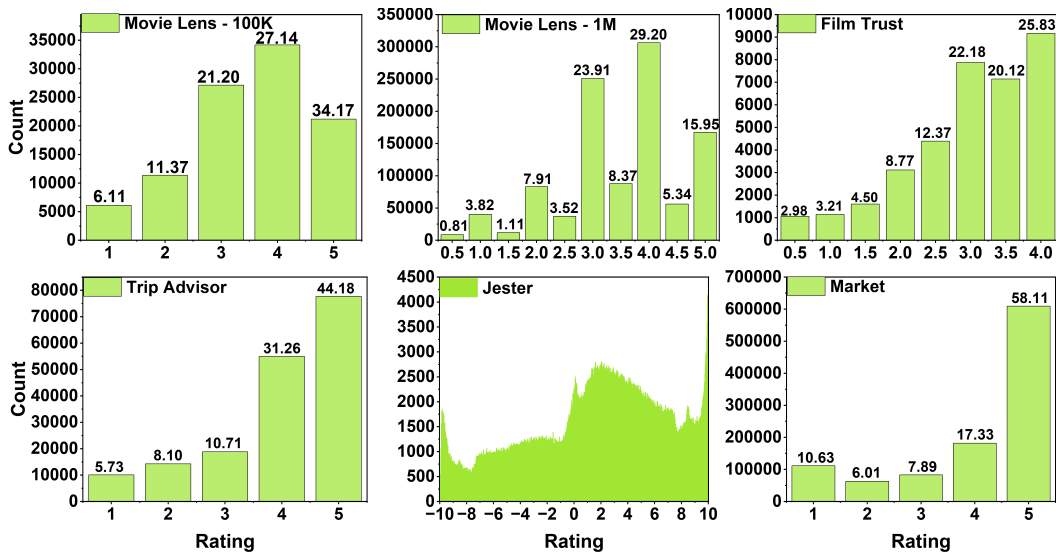


Fig. 8. Visual insights into Rating Distributions: Movie Lens-100K, Movie Lens-1M, Film Trust, Trip Advisor, Jester, and Market datasets.

300, 400. In the trip advisor dataset, the RMSE value decreases at finite steps for all k values and slightly increases. There is an increase in RMSE value if k is 10. In the market dataset, the RMSE value is decreased for all k values except for k value 200. Compared to all the datasets, the jester is giving less RMSE value. As there is a maximum of 140 items in the jester dataset, we can calculate RMSE value up to k less than or equal to $\min(m, n)$.

Fig. 10 describes RMSE value on six different datasets for PMF at different steps. There are similar fluctuations in RMSE value as the k value is altered. In the movie lens-100K, for k is 10 latent features, it is observed that there is a decrease in RMSE value from 1.75 to 1.2. At 50 latent features, the RMSE value increases from 1.35 to 1.45. At 100 and 300 latent features, the RMSE increases from 1.65 to 1.75. At 200 latent features, it is observed that the RMSE decreases more from 2.0 to 1.5. For = 400 latent features, there is an increase

in RMSE value from 1.55 to 1.9.

In the movie lens-1M, at 10 latent features, it is observed that there is a decrease in RMSE from 1.8 to 1.35. For k is 50 latent features, it is observed that the RMSE decreases from 1.55 to 1.35. For 100 latent features, the RMSE value reduces from 1.5 to 1.4. At 200 latent features, the RMSE value increases from 1.2 to 1.35. The RMSE value falls from 1.45 to 1.3 for 300 latent features. There is an increase in RMSE value from 1.35 to 1.55 at 400 latent features. In the film trust dataset, all are behaving in the same manner except for 50 latent features. There is a drastic change in RMSE value with different behavior from 1.20 to 0.99. For the remaining k values there are slight fluctuations in RMSE value. In the trip advisor dataset, different behavior is seen for $k = 10$ latent features. For all the remaining latent features, there is a decrease in RMSE value if the latent features are increased. In the jester dataset, there is an increase in RMSE value as

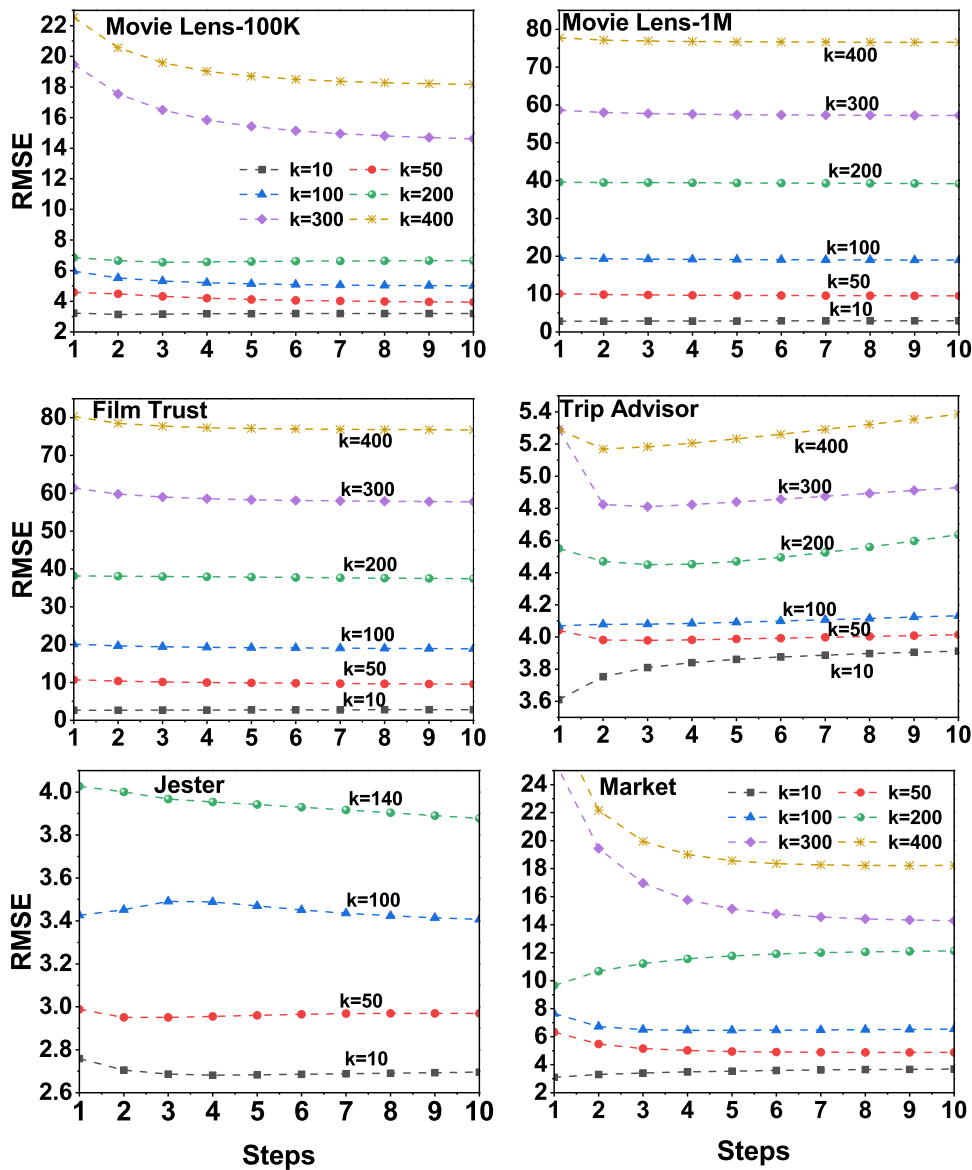


Fig. 9. Root Mean Square Error (RMSE) graphs for Matrix Factorization (MF) on six datasets, namely, Movie Lens-100K, Movie Lens-1M, Film Trust, Trip Advisor, Jester, and Market datasets for 10 steps and 400 latent features.

the latent features are increased. There is a decrease in RMSE value at 100. The constant RMSE is maintained at k value 140. In the market dataset, there is a decrease in RMSE value if the k -value increases.

For the NMF method, we have plotted RMSE values with different latent features on six datasets in Fig. 11. In all datasets, if the k -value is increasing RMSE decreases. Compared to all datasets, the NMF method gives less RMSE value for film trust and market datasets. More RMSE value for jester, movie lens-100K datasets. For any latent features, the RMSE value is decreasing with different steps.

Fig. 12 (a), (b), and (c) describes RMSE value of SVD method with various k values for pair of datasets movie lens-

100K, jester, movie lens-1M, market, and trip advisor, film trust respectively. Across all datasets, an increase in the k -value is observed to correspond with a decrease in the RMSE value. Compared to all datasets, the SVD method gives less RMSE value for film trust and market datasets. More RMSE value for movie lens-100K datasets. For the jester dataset a drastic change in the RMSE value of the SVD method when increasing the k -value because the number of items is very less.

Fig. 13 describes RMSE value on six different datasets for SVD++ method at different steps. In all datasets, if the k -value is increasing then RMSE also increases. In the jester dataset, it is observed that the RMSE is maintained constant for different steps. Compared to all datasets, the SVD++ method gives less

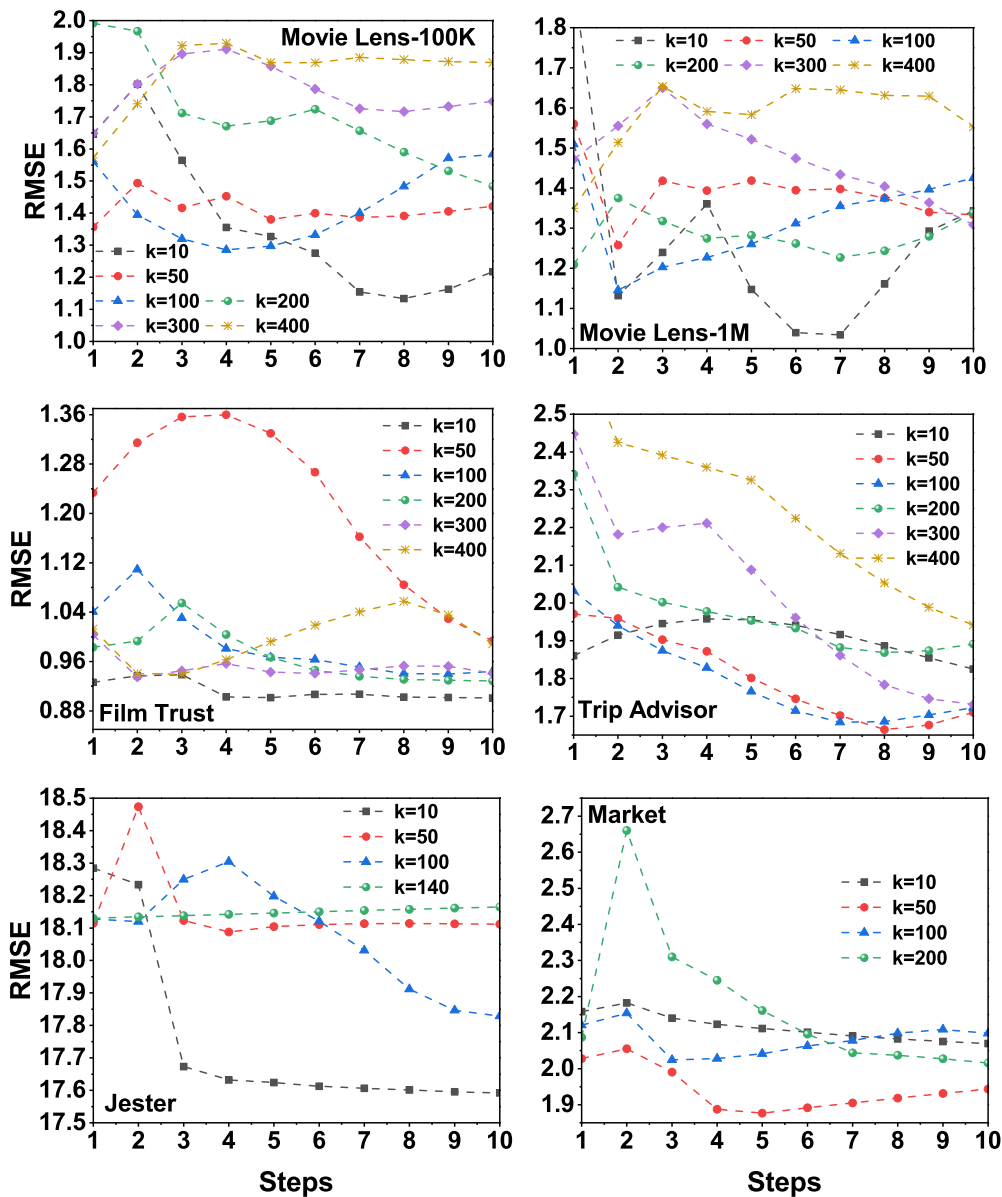


Fig. 10. Root Mean Square Error (RMSE) graphs for Probabilistic Matrix Factorization (PMF) on six datasets, namely, Movie Lens-100K, Movie Lens-1M, Film Trust, Trip Advisor, Jester, and Market datasets for 10 steps and 400 latent features.

RMSE value for the film trust dataset and more RMSE value for the jester dataset.

VII. COMPARISON OF MF METHODS

The comparison of different MF methods on different datasets namely movie lens-100K, film trust, movie lens-1M, trip advisor, jester, and market with $k = 5$ in Table IV, and $k = 10$ in Table V, and $k = 100$ in Table VI. Compared to all the MF methods NMF, and SVD methods gives less RMSE value with different k -values. It is observed that in MF, PMF, and SVD++ methods, by increasing the k -value RMSE also increases. Whereas in NMF, and SVD methods there is a decrease in RMSE value as k -value increases. The

k -value in each method is the number of latent features that are divided into the items. As compared to all MF methods, if we consider more groups for items in the dataset, MF, PMF, and SVD++ give more errors for suggesting an item to the user. Whereas, if we increase the groups in NMF and SVD methods, the error value that is obtained is less while predicting a recommendation to the user.

For film trust and movie lens-1M datasets, the prediction performance of MF is drastically decreasing with an increase in the number of latent features. The dominant observation is that the rating distribution for the two datasets is right skewed and sparse. However, NMF and SVD are not affected much by this skewed rating as well as the number of latent features.

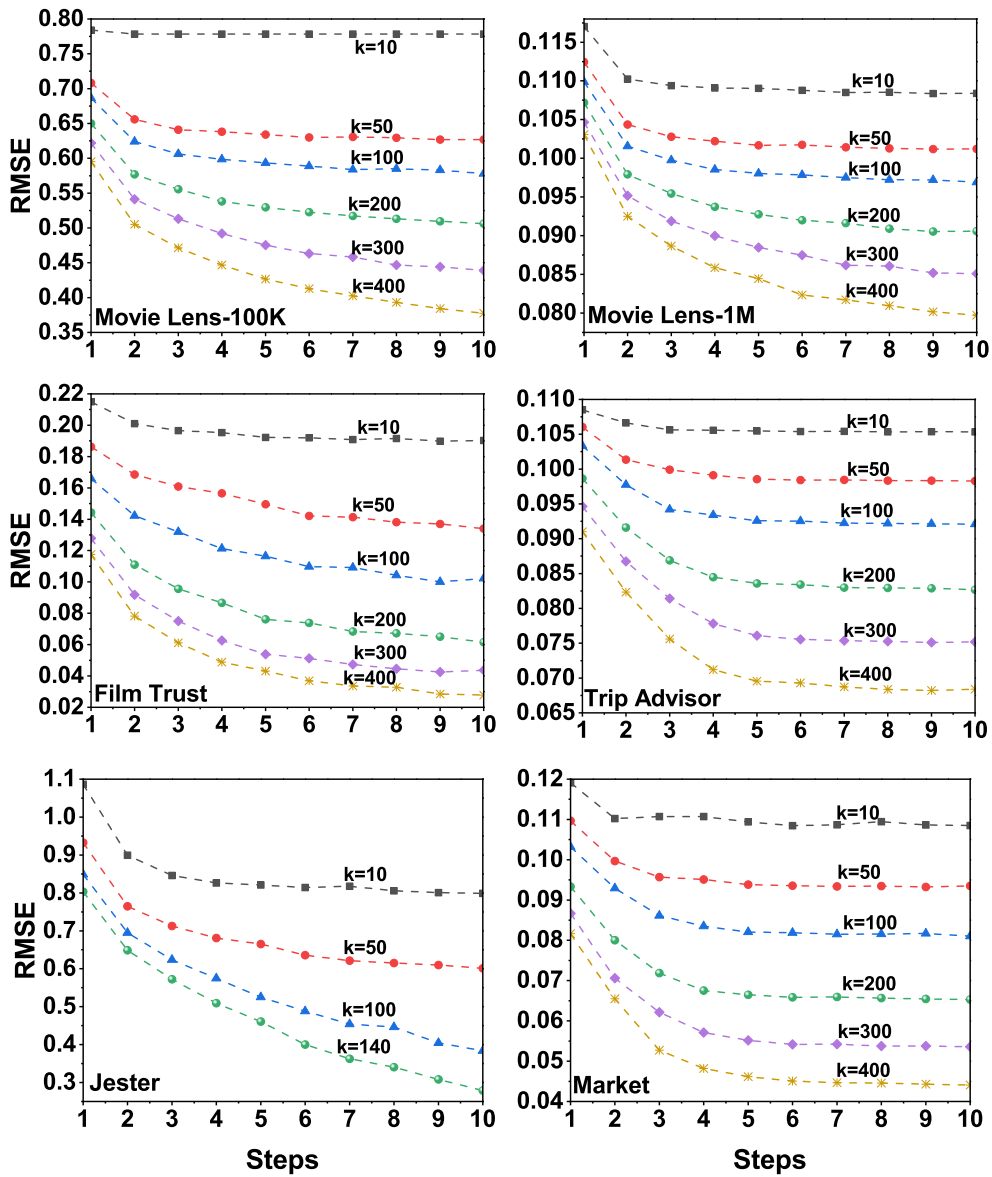


Fig. 11. Root Mean Square Error (RMSE) graphs for Non-Negative Matrix Factorization (NMF) on six datasets, namely, Movie Lens-100K, Movie Lens-1M, Film Trust, Trip Advisor, Jester, and Market datasets for 10 steps and 400 latent features.

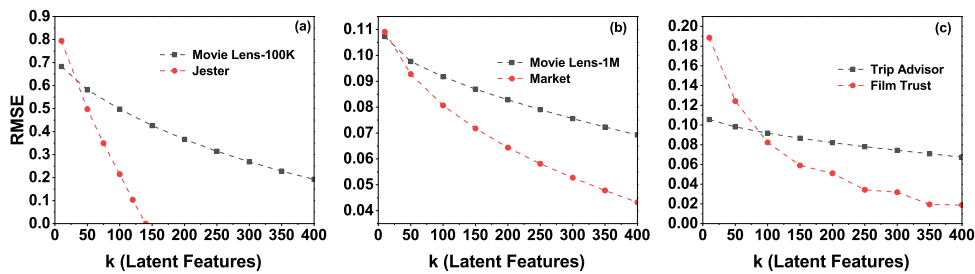


Fig. 12. Root Mean Square Error (RMSE) graphs for Singular Value Decomposition (SVD) on six datasets, namely, Movie Lens-100K, Movie Lens-1M, Film Trust, Trip Advisor, Jester, and Market datasets for 10 steps and 400 latent features.

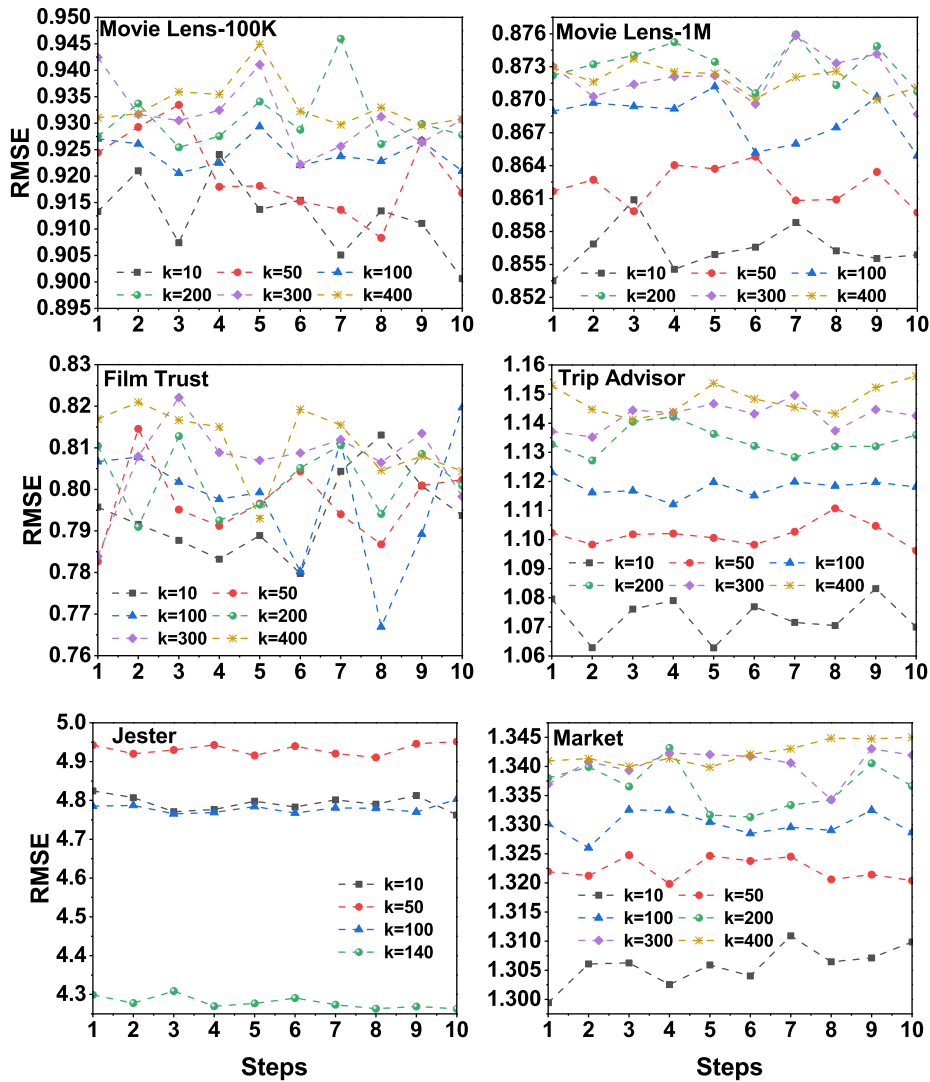


Fig. 13. Root Mean Square Error (RMSE) graphs for SVD++ on six datasets, namely, Movie Lens-100K, Movie Lens-1M, Film Trust, Trip Advisor, Jester, and Market datasets for 10 steps and 400 latent features.

TABLE IV. RMSE VALUES FOR DIFFERENT VARIATIONS OF MATRIX FACTORIZATION (MF) ON SIX DIFFERENT DATASETS NAMELY MOVIE LENS-100K, MOVIE LENS-1M, FILM TRUST, TRIP ADVISOR, JESTER, AND MARKET AT $k = 5$

Dataset (↓) / MF algorithm (→)	MF	PMF	NMF	SVD	SVD++
Movie Lens-100K	1.048	1.051	0.720	0.711	0.908
Movie Lens-1M	2.236	1.463	0.111	0.110	0.859
Film Trust	2.722	1.457	0.200	0.199	0.798
Trip Advisor	3.919	1.945	0.106	0.106	1.071
Jester	2.695	17.591	0.798	0.794	4.762
Market	3.597	2.041	0.112	0.112	1.294

TABLE V. RMSE VALUES FOR DIFFERENT VARIATIONS OF MATRIX FACTORIZATION (MF) ON SIX DIFFERENT DATASETS NAMELY MOVIE LENS-100K, MOVIE LENS-1M, FILM TRUST, TRIP ADVISOR, JESTER, AND MARKET AT $k = 10$

Dataset (↓) / MF algorithm (→)	MF	PMF	NMF	SVD	SVD++
Movie Lens-100K	3.205	1.343	0.690	0.682	0.900
Movie Lens-1M	2.953	1.271	0.108	0.107	0.855
Film Trust	2.804	0.900	0.190	0.188	0.793
Trip Advisor	3.912	1.824	0.105	0.105	1.069
Jester	3.499	17.521	0.867	0.861	4.707
Market	3.692	2.069	0.108	0.109	1.309

VIII. CONCLUSION AND FUTURE SCOPE

A. Conclusion

Information theory is essential for improving the effectiveness of recommendation systems. By measuring the un-

certainty and information content in user preferences and interactions, it offers a solid framework for creating more precise and efficient recommendation algorithms. This theoretical basis enhances the handling of sparse data, boosts prediction accuracy, and ensures more personalized user ex-

TABLE VI. RMSE VALUES FOR DIFFERENT VARIATIONS OF MATRIX FACTORIZATION (MF) ON SIX DIFFERENT DATASETS, NAMELY, MOVIE LENS-100K, MOVIE LENS-1M, FILM TRUST, TRIP ADVISOR, JESTER, AND MARKET AT $k = 100$

Dataset (\downarrow) / MF algorithm (\rightarrow)	MF	PMF	NMF	SVD	SVD++
Movie Lens-100K	5.014	1.583	0.577	0.497	0.920
Movie Lens-1M	18.994	1.425	0.096	0.091	0.864
Film Trust	18.886	0.944	0.102	0.082	0.879
Trip Advisor	4.131	1.722	0.092	0.091	1.118
Jester	3.407	17.828	0.383	0.219	4.803
Market	6.539	2.097	0.810	0.080	1.328

periences. As recommendation systems advance, the principles of information theory will continue to be vital in tackling challenges related to data complexity, user diversity, and changing preferences, resulting in more sophisticated and dependable recommendation solutions. In this study, various MF methods like MF, PMF, NMF, SVD, and SVD++ have been compared on different datasets namely movie lens-100K, film trust, movie lens-1M, trip advisor, jester, and market with different latent features on different steps. The performance of the MF methods is evaluated using RMSE. It is observed that MF is the least-performing MF method among all studied in this work. SVD is the outperforming method among all other MF algorithms. However, it has been observed that the number of latent features is affecting the prediction performance. The prediction power of MF, PMF, and SVD++ is reducing with an increase in the number of latent features. On the other hand, NMF and SVD are performing better with an increase in the number of latent features.

B. Future Scope

In the future, we would like to extend the concept of recommendation to different real-world contexts. For example, this study focuses solely on recommending a single item to an individual user. However, in practice, there are situations where recommendations are needed for a group of users. For example, a group of students might be advised on selecting an elective course based on their collective interests. This interest prompts the requirement for and creation of group recommender systems [61]. Cross-domain recommendation systems (CDR) can help mitigate this issue. CDRs use the ratings of the new item/user in one domain in another domain with transfer learning [62]–[64]. Utilizing these techniques will help recommendation systems work better by a variety of semantic information contained in knowledge graphs. A knowledge graph (KG) is a collection of relational facts, including information about the entities, entity categories, and collaborations among entities. KG embeds complex information about different relationships among real-world entities [65], [66].

REFERENCES

- [1] S. Tokala, M. K. Enduri, T. J. Lakshmi, A. Abdul, and J. Chen, "Empowering quality of recommendations by integrating matrix factorization approaches with louvain community detection," *IEEE Access*, 2024.
- [2] R. Chen, Q. Hua, B. Wang, M. Zheng, W. Guan, X. Ji, Q. Gao, and X. Kong, "A novel social recommendation method fusing user's social status and homophily based on matrix factorization techniques," *IEEE Access*, vol. 7, pp. 18 783–18 798, 2019.
- [3] S. Tokala, J. Nagaram, M. K. Enduri, and T. J. Lakshmi, "Enhanced movie recommender system using deep learning techniques," in *2024 3rd International Conference on Computational Modelling, Simulation and Optimization (ICCMO)*. IEEE, 2024, pp. 71–75.
- [4] S. Tokala, M. K. Enduri, and T. J. Lakshmi, "Unleashing the power of svd and louvain community detection for enhanced recommendations," in *2023 IEEE 15th International Conference on Computational Intelligence and Communication Networks (CICN)*. IEEE, 2023, pp. 807–811.
- [5] D. Roy and M. Dutta, "A systematic review and research perspective on recommender systems," *Journal of Big Data*, vol. 9, no. 1, p. 59, 2022.
- [6] C. Gao, Y. Zheng, W. Wang, F. Feng, X. He, and Y. Li, "Causal inference in recommender systems: A survey and future directions," *ACM Transactions on Information Systems*, vol. 42, no. 4, pp. 1–32, 2024.
- [7] J. Chen, H. Dong, X. Wang, F. Feng, M. Wang, and X. He, "Bias and debias in recommender system: A survey and future directions," *ACM Transactions on Information Systems*, vol. 41, no. 3, pp. 1–39, 2023.
- [8] Y. Hou, J. Zhang, Z. Lin, H. Lu, R. Xie, J. McAuley, and W. X. Zhao, "Large language models are zero-shot rankers for recommender systems," in *European Conference on Information Retrieval*. Springer, 2024, pp. 364–381.
- [9] Y. Wang, W. Ma, M. Zhang, Y. Liu, and S. Ma, "A survey on the fairness of recommender systems," *ACM Transactions on Information Systems*, vol. 41, no. 3, pp. 1–43, 2023.
- [10] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, "Recommender system application developments: a survey," *Decision Support Systems*, vol. 74, pp. 12–32, 2015.
- [11] J. K. Kim, Y. H. Cho, W. J. Kim, J. R. Kim, and J. H. Suh, "A personalized recommendation procedure for internet shopping support," *Electronic commerce research and applications*, vol. 1, no. 3–4, pp. 301–313, 2002.
- [12] K.-j. Kim and H. Ahn, "A recommender system using ga k-means clustering in an online shopping market," *Expert systems with applications*, vol. 34, no. 2, pp. 1200–1209, 2008.
- [13] C. Porcel and E. Herrera-Viedma, "Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries," *Knowledge-Based Systems*, vol. 23, no. 1, pp. 32–39, 2010.
- [14] C. Porcel, J. M. Moreno, and E. Herrera-Viedma, "A multi-disciplinary recommender system to advice research resources in university digital libraries," *Expert systems with applications*, vol. 36, no. 10, pp. 12 520–12 528, 2009.
- [15] R. Sikka, A. Dhankhar, and C. Rana, "A survey paper on e-learning recommender system," *International Journal of Computer Applications*, vol. 47, no. 9, pp. 27–30, 2012.
- [16] J. Lu, "A personalized e-learning material recommender system," in *International conference on information technology and applications*. Macquarie Scientific Publishing, 2004.
- [17] J. Borràs, A. Moreno, and A. Valls, "Intelligent tourism recommender systems: A survey," *Expert systems with applications*, vol. 41, no. 16, pp. 7370–7389, 2014.
- [18] Y.-L. Chen, L.-C. Cheng, and C.-N. Chuang, "A group recommendation system with consideration of interactions among group members," *Expert systems with applications*, vol. 34, no. 3, pp. 2082–2090, 2008.
- [19] I. Cantador, I. Fernández-Tobías, S. Berkovsky, and P. Cremonesi, "Cross-domain recommender systems," in *Recommender systems handbook*. Springer, 2015, pp. 919–959.
- [20] P. Vermeir, D. Vandijck, S. Degroote, R. Peleman, R. Verhaeghe, E. Mortier, G. Hallaert, S. Van Daele, W. Buylaert, and D. Vogelaers, "Communication in healthcare: a narrative review of the literature and practical recommendations," *International journal of clinical practice*, vol. 69, no. 11, pp. 1257–1267, 2015.
- [21] Y. Himeur, A. Sayed, A. Alsalemi, F. Bensaali, A. Amira, I. Varlamis, M. Eirinaki, C. Sardanios, and G. Dimitrakopoulos, "Blockchain-based recommender systems: Applications, challenges and future opportunities," *Computer Science Review*, vol. 43, p. 100439, 2022.
- [22] F. Karimova, "A survey of e-commerce recommender systems," *European Scientific Journal*, vol. 12, no. 34, pp. 75–89, 2016.

- [23] L. Wu, X. He, X. Wang, K. Zhang, and M. Wang, "A survey on accuracy-oriented neural recommendation: From collaborative filtering to information-rich recommendation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 5, pp. 4425–4445, 2022.
- [24] J. Liu, C. Shi, C. Yang, Z. Lu, and S. Y. Philip, "A survey on heterogeneous information network based recommender systems: Concepts, methods, applications and resources," *AI Open*, vol. 3, pp. 40–57, 2022.
- [25] U. Javed, K. Shaukat, I. A. Hameed, F. Iqbal, T. M. Alam, and S. Luo, "A review of content-based and context-based recommendation systems," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 3, pp. 274–306, 2021.
- [26] K. Kundegraber and S. Pletzl, "Basic approaches in recommendation systems," EasyChair, Tech. Rep., 2022.
- [27] S. Chen, S. Owusu, and L. Zhou, "Social network based recommendation systems: A short survey," in *2013 international conference on social computing*. IEEE, 2013, pp. 882–885.
- [28] S. Reddy, S. Nalluri, S. Kuniseti, S. Ashok, and B. Venkatesh, "Content-based movie recommendation system using genre correlation," in *Smart Intelligent Computing and Applications: Proceedings of the Second International Conference on SCI 2018, Volume 2*. Springer, 2019, pp. 391–397.
- [29] N. Y. Asabere, "Review of recommender systems for learners in mobile social/collaborative learning," *International Journal of Information*, vol. 2, no. 5, 2012.
- [30] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems*, vol. 22, no. 1, pp. 5–53, 2004.
- [31] M. D. Ekstrand, J. T. Riedl, J. A. Konstan *et al.*, "Collaborative filtering recommender systems," *Foundations and Trends® in Human-Computer Interaction*, vol. 4, no. 2, pp. 81–173, 2011.
- [32] M. A. Hameed, O. Al Jadaan, and S. Ramachandram, "Collaborative filtering based recommendation system: A survey," *International Journal on Computer Science and Engineering*, vol. 4, no. 5, p. 859, 2012.
- [33] S. A. Amin, J. Philips, and N. Tabrizi, "Current trends in collaborative filtering recommendation systems," in *World Congress on Services*. Springer, 2019, pp. 46–60.
- [34] M. Jalili, S. Ahmadian, M. Izadi, P. Moradi, and M. Salehi, "Evaluating collaborative filtering recommender algorithms: a survey," *IEEE access*, vol. 6, pp. 74 003–74 024, 2018.
- [35] S. Tokala, M. K. Enduri, T. J. Lakshmi, and H. Sharma, "Community-based matrix factorization (cbmf) approach for enhancing quality of recommendations," *Entropy*, vol. 25, no. 9, p. 1360, 2023.
- [36] S. Gong, "A collaborative filtering recommendation algorithm based on user clustering and item clustering," *J. Softw.*, vol. 5, no. 7, pp. 745–752, 2010.
- [37] J. Wang, A. P. De Vries, and M. J. Reinders, "Unifying user-based and item-based collaborative filtering approaches by similarity fusion," in *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, 2006, pp. 501–508.
- [38] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," in *Proceedings of the 10th international conference on World Wide Web*, 2001, pp. 285–295.
- [39] M.-P. T. Do, D. Nguyen, and L. Nguyen, "Model-based approach for collaborative filtering," in *6th International conference on information technology for education*, 2010, pp. 217–228.
- [40] J. Hintz, "Matrix factorization for collaborative filtering recommender systems," 2015.
- [41] D. kumar Bokde, S. Girase, and D. Mukhopadhyay, "Role of matrix factorization model in collaborative filtering algorithm: A survey," *Clinical Orthopaedics and Related Research*, abs/1503.07475, vol. 1, no. 12, 2015.
- [42] J. Ivarsson and M. Lindgren, "Movie recommendations using matrix factorization," 2016.
- [43] Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *Computer*, vol. 42, no. 8, pp. 30–37, 2009.
- [44] J. Liu, C. Wu, Y. Xiong, and W. Liu, "List-wise probabilistic matrix factorization for recommendation," *Information Sciences*, vol. 278, pp. 434–447, 2014.
- [45] A. Mnih and R. R. Salakhutdinov, "Probabilistic matrix factorization," *Advances in neural information processing systems*, vol. 20, 2007.
- [46] Z. Huang, A. Zhou, and G. Zhang, "Non-negative matrix factorization: a short survey on methods and applications," in *International Symposium on Intelligence Computation and Applications*. Springer, 2012, pp. 331–340.
- [47] Y.-X. Wang and Y.-J. Zhang, "Nonnegative matrix factorization: A comprehensive review," *IEEE Transactions on knowledge and data engineering*, vol. 25, no. 6, pp. 1336–1353, 2012.
- [48] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.
- [49] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.
- [50] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Application of dimensionality reduction in recommender system—a case study," Minnesota Univ Minneapolis Dept of Computer Science, Tech. Rep., 2000.
- [51] E. J. Lentilucci, "Using the singular value decomposition," *Rochester Institute of Technology, Rochester, New York, United States, Technical Report*, 2003.
- [52] R. Mehta and K. Rana, "A review on matrix factorization techniques in recommender systems," in *2017 2nd International Conference on Communication Systems, Computing and IT Applications*. IEEE, 2017, pp. 269–274.
- [53] J. Jiao, X. Zhang, F. Li, and Y. Wang, "A novel learning rate function and its application on the svd++ recommendation algorithm," *IEEE Access*, vol. 8, pp. 14 112–14 122, 2019.
- [54] R. Kumar, B. Verma, and S. S. Rastogi, "Social popularity based svd++ recommender system," *International Journal of Computer Applications*, vol. 87, no. 14, 2014.
- [55] "Kaggle," <https://www.kaggle.com/datasets/prajitdata/movielens-100k-dataset> (accessed on 31 July 2023).
- [56] "Kaggle," <https://www.kaggle.com/datasets/odedgolden/movielens-1m-dataset> (accessed on 31 July 2023).
- [57] "Konec," <http://konec.cc/networks/librec-filmtrust-ratings/> (accessed on 31 July 2023).
- [58] "Konec," <http://konec.cc/networks/wang-tripadvisor/> (accessed on 31 July 2023).
- [59] "Konec," <http://konec.cc/networks/jester2/> (accessed on 31 July 2023).
- [60] "Github," <https://github.com/MengtingWan/marketBias> (accessed on 31 July 2023).
- [61] S. Dara, C. R. Chowdary, and C. Kumar, "A survey on group recommender systems," *Journal of Intelligent Information Systems*, vol. 54, no. 2, pp. 271–295, 2020.
- [62] M. M. Khan, R. Ibrahim, and I. Ghani, "Cross domain recommender systems: a systematic literature review," *Association for Computing Machinery Computing Surveys*, vol. 50, no. 3, pp. 1–34, 2017.
- [63] M. Enrich, M. Braunhofer, and F. Ricci, "Cold-start management with cross-domain collaborative filtering and tags," in *International Conference on Electronic Commerce and Web Technologies*. Springer, 2013, pp. 101–112.
- [64] I. Fernández-Tobías, I. Cantador, M. Kaminskas, and F. Ricci, "Cross-domain recommender systems: A survey of the state of the art," in *Spanish conference on information retrieval*, vol. 24. ACM Valencia, Spain, 2012.
- [65] S. Bouraga, I. Jureta, S. Faulkner, and C. Herssens, "Knowledge-based recommendation systems: A survey," *International Journal of Intelligent Information Technologies*, vol. 10, no. 2, pp. 1–19, 2014.
- [66] R. Burke, "Knowledge-based recommender systems," *Encyclopedia of library and information systems*, vol. 69, no. Supplement 32, pp. 175–186, 2000.

SUPPLEMENTARY INFORMATION

In the supplementary information section, RSE graphs for MF at $k = 10, 50, 100, 200, 300, 400$ at 10 steps and 100 steps at 10 latent features are shown and RMSE graphs for MF, PMF, NMF, and SVD methods are provided at 100 steps and 10 latent features and MAE graphs for the SVD++ method at $k = 10, 50, 100, 200, 300, 400$ at 10 steps are shown.

Fig. 14 describes the RSE value on six different datasets for the MF method on 10 steps and 400 latent features. In all datasets, it is observed that there is an increase in RSE value as the k -value increases. Compared to all the datasets less RMSE value is given by film trust and more RMSE value is given by the jester dataset.

Fig. 15 describes the RSE value on six different datasets for the MF method on 100 steps and 10 latent features. In all datasets, except for jester, it is observed that if the k -value is increasing then RSE decreases. Compared to all datasets, the MF method gives less RSE value for the market dataset and more RSE value for the jester.

Fig. 16 describes the RMSE value on six different datasets for the MF method at different steps. In all datasets, if the k -value is increasing then RMSE also increases. Compared to all datasets, the MF method gives less RMSE value for film trust, movie lens-1M datasets, and more RMSE value for movie lens-100K, trip advisor, film trust, and market.

Fig. 17 describes RMSE value on six different datasets for the PMF method on 100 steps and 10 latent features. In the movie lens-100K dataset, at $k = 1$ there is a constant behavior maintained. For all the remaining k values there are many alterations in RMSE value. In movie lens-1M, for $k = 2$, there is a major deviation in the RMSE value as compared to

all remaining k values. In the film trust dataset, for all the k values between 1 to 9, there are fluctuations as they are altered. For $k = 10$ the RMSE value is high as compared to the remaining k values. In the trip advisor dataset, all the k values exhibit different behavior. In the jester dataset, as compared to the remaining datasets, the RMSE value is too high as there is a decrease in the RMSE value with different steps. In the market dataset, all the k values have different behavior with an increase in steps.

Fig. 18 describes RMSE value on six different datasets for the NMF method on 100 steps and 10 latent features. In all datasets, if the k -value is increasing then RMSE also decreases. Compared to all the datasets, the NMF method gives less RMSE value for trip advisor, movie lens-1M, and market datasets and some more RMSE value for movie lens-100K, and film trust datasets. More RMSE value is given by the jester dataset due to less items.

Fig. 19 (a), (b), and (c) describes RMSE value of SVD method with various k values for pair of datasets movie lens-100K, jester, movie lens-1M, trip advisor, and film trust, market, respectively. In all datasets, it is observed that as the k value increases there is a decrease in RMSE value. Compared to all the datasets, the SVD method gives less RMSE value for trip advisor, movie lens-1M, and market datasets. More RMSE value for movie lens-100K and jester datasets.

Fig. 20 describes the MAE value on six different datasets for the SVD++ at different steps. In movie lens-100K, movie lens-1M, trip advisor, and market datasets, it is observed that there is an increase in RMSE value as the step increases. In the jester dataset, a similar constant RMSE value is maintained at different steps.

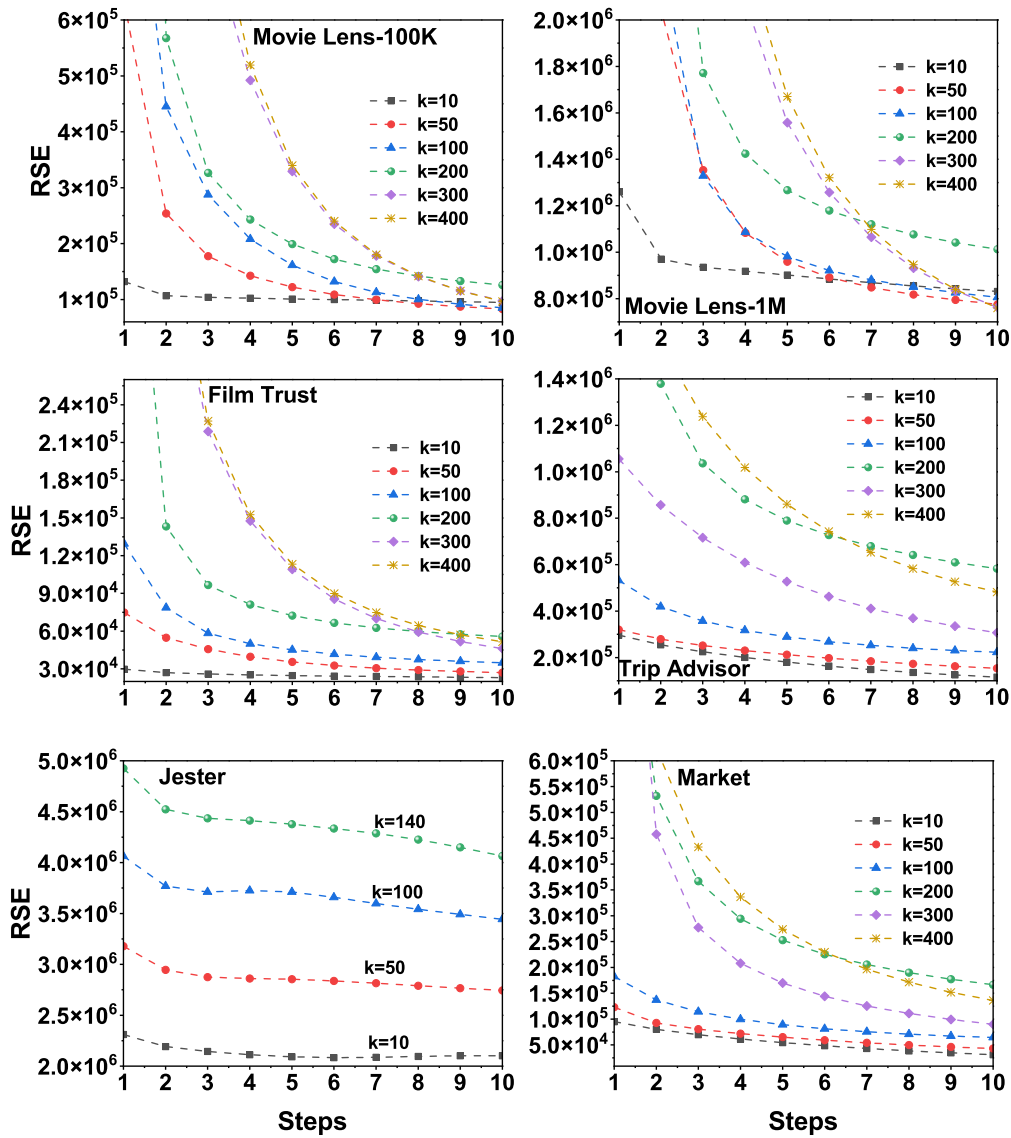


Fig. 14. Regularized Square Error (RSE) graphs for basic Matrix Factorization (MF) on six datasets, namely, Movie Lens-100K, Movie Lens-1M, Film Trust, Trip Advisor, Jester, and Market datasets for 10 steps and 400 latent features.

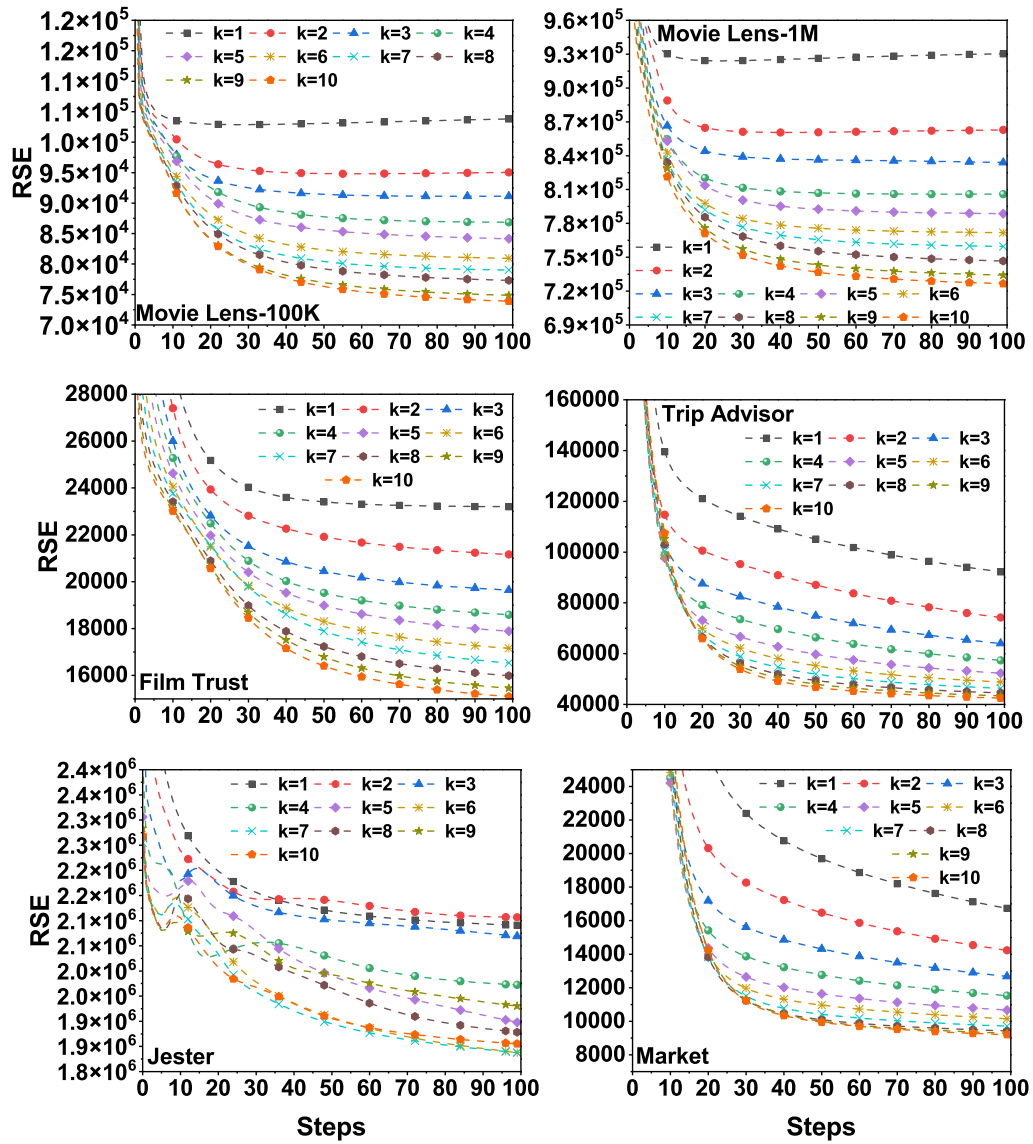


Fig. 15. Regularized Square Error (RSE) graphs for basic Matrix Factorization (MF) on six datasets, namely, Movie Lens-100K, Movie Lens-1M, Film Trust, Trip Advisor, Jester, and Market datasets for 100 steps and 10 latent features.

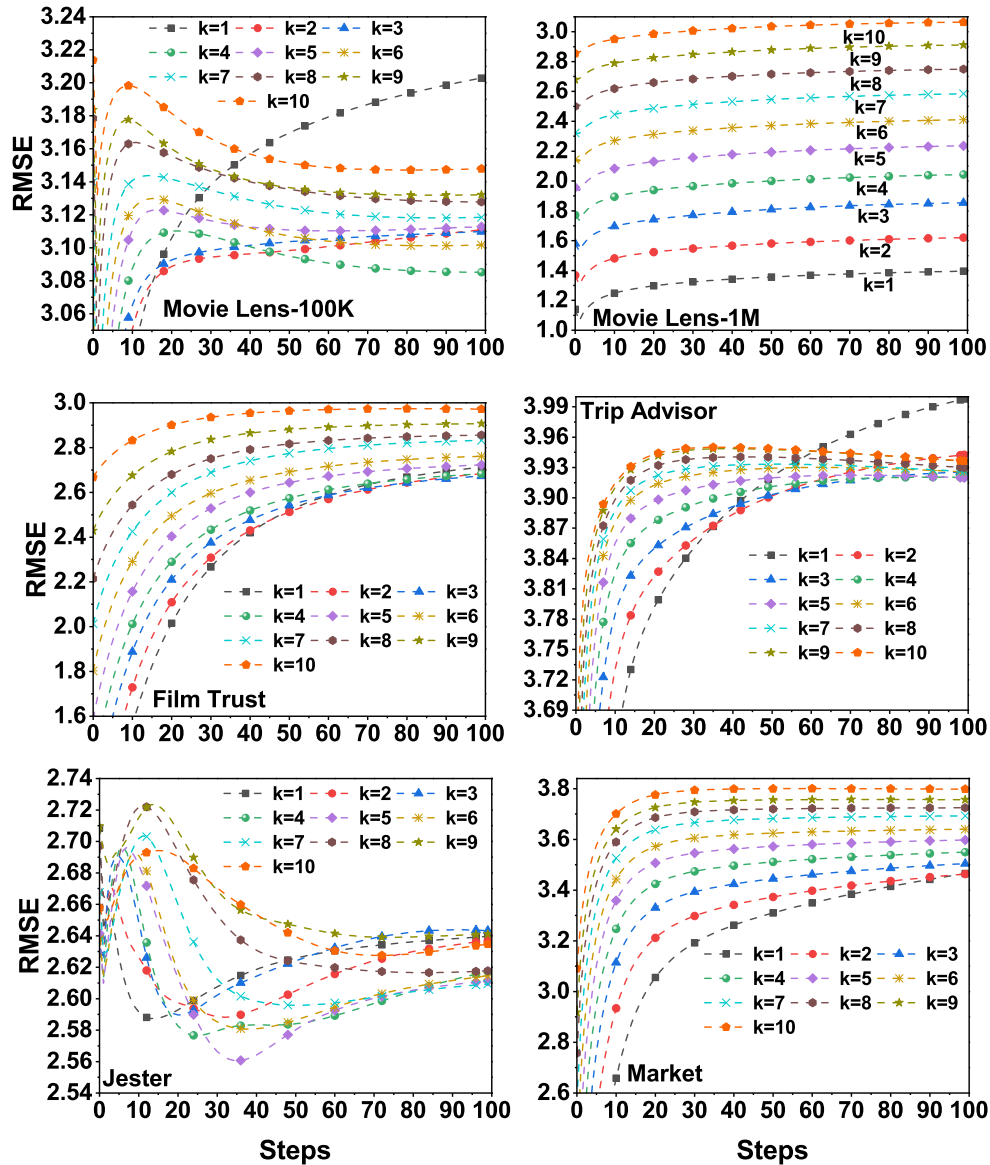


Fig. 16. Root Mean Square Error (RMSE) graphs for basic Matrix Factorization (MF) on six datasets, namely, Movie Lens-100K, Movie Lens-1M, Film Trust, Trip Advisor, Jester, and Market datasets for 100 steps and 10 latent features.

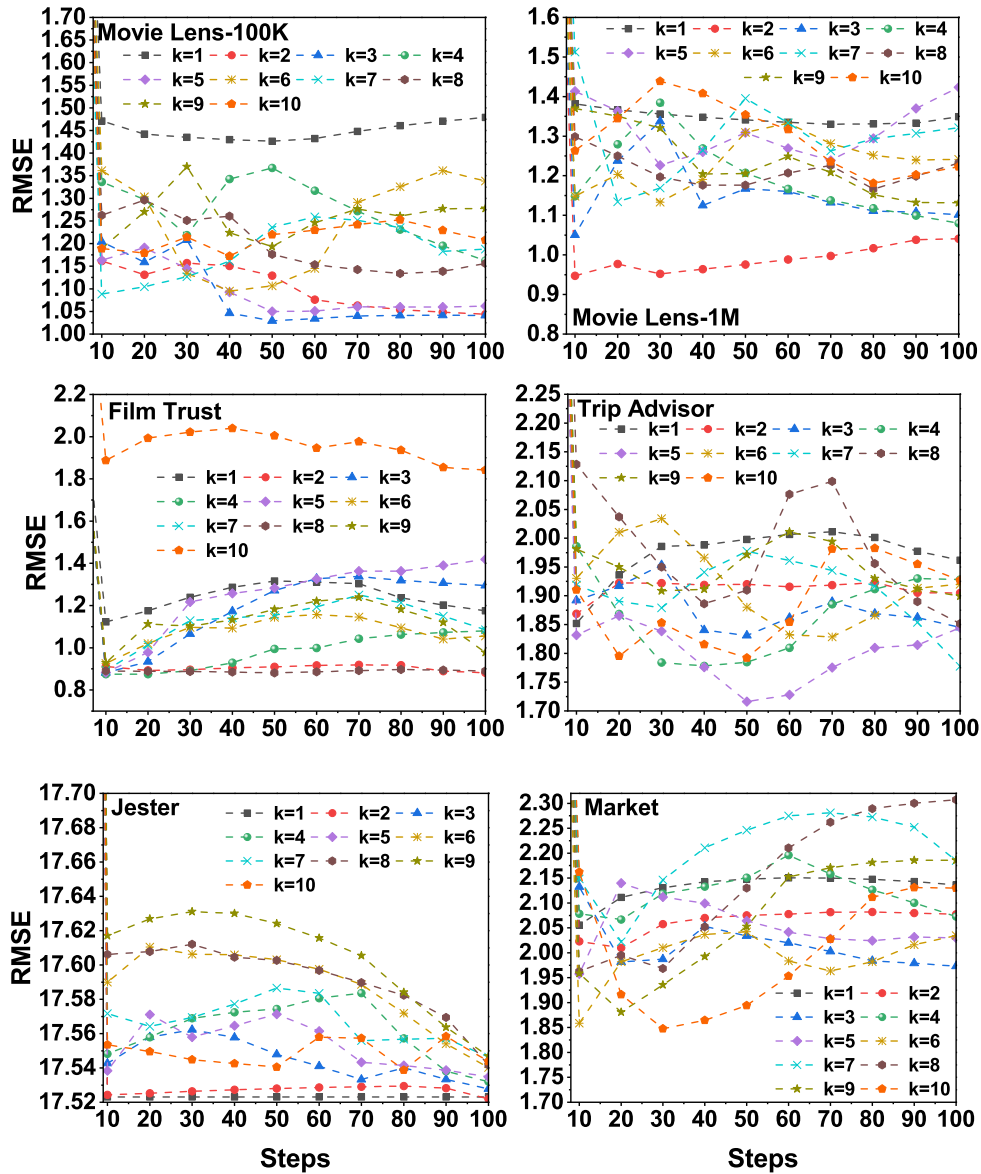


Fig. 17. Root Mean Square Error (RMSE) graphs for probabilistic Matrix Factorization (PMF) on six datasets, namely, Movie Lens-100K, Movie Lens-1M, Film Trust, Trip Advisor, Jester, and Market datasets for 100 steps and 10 latent features.

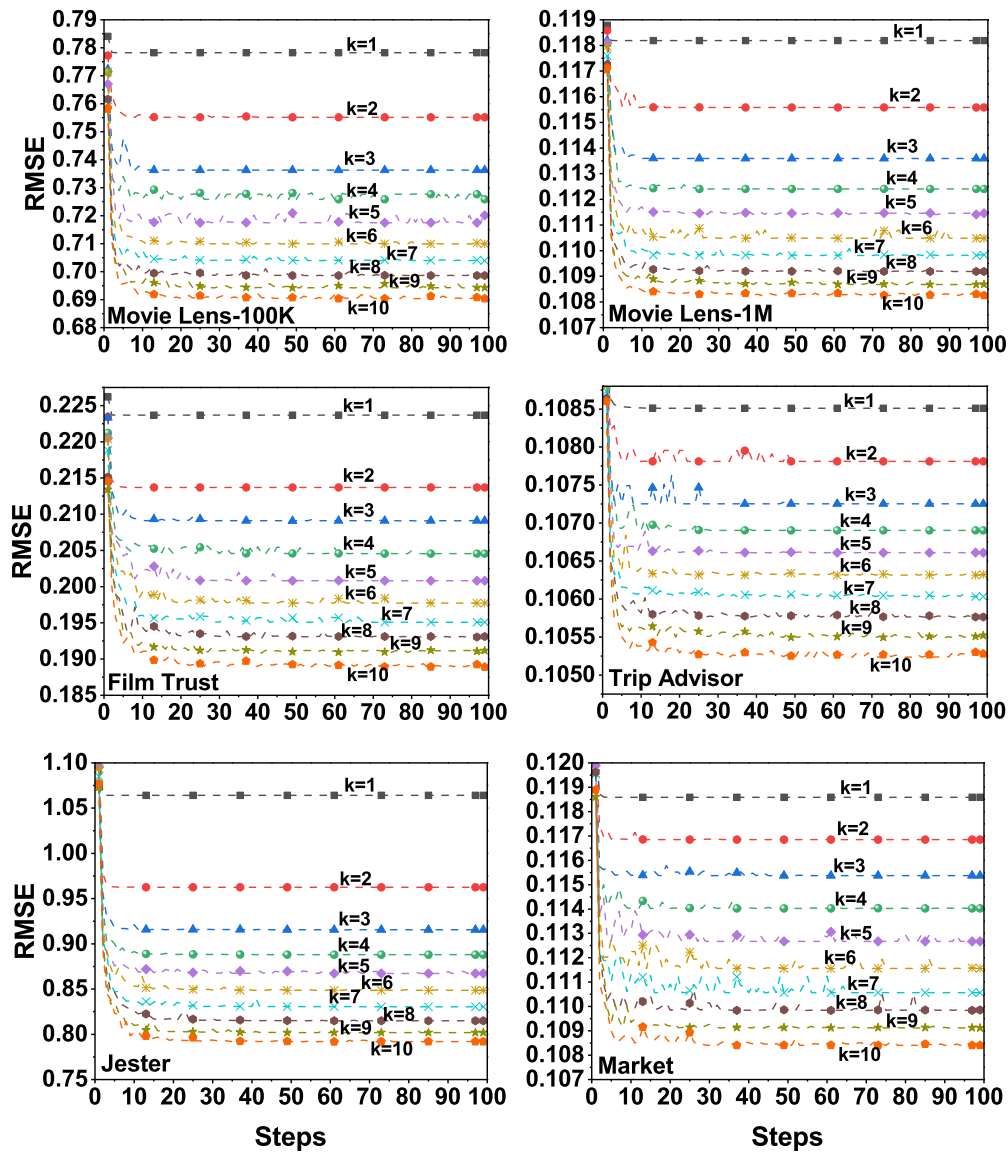


Fig. 18. Root Mean Square Error (RMSE) graphs for Non-Negative Matrix Factorization (NMF) on six datasets, namely, Movie Lens-100K, Movie Lens-1M, Film Trust, Trip Advisor, Jester, and Market datasets for 100 steps and 10 latent features.

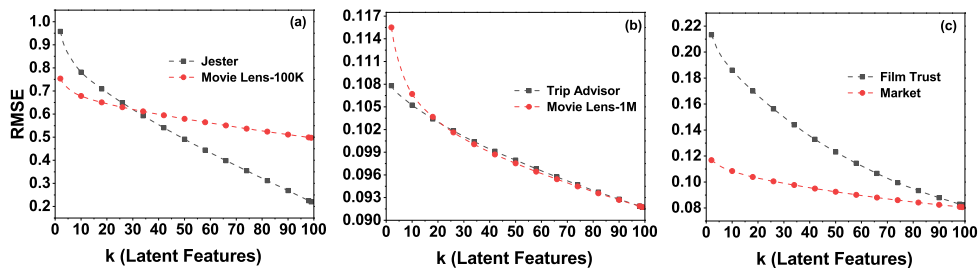


Fig. 19. Root Mean Square Error (RMSE) graphs for Singular Value Decomposition (SVD) on six datasets, namely, Movie Lens-100K, Movie Lens-1M, Film Trust, Trip Advisor, Jester, and Market datasets for 100 steps and 100 latent features.

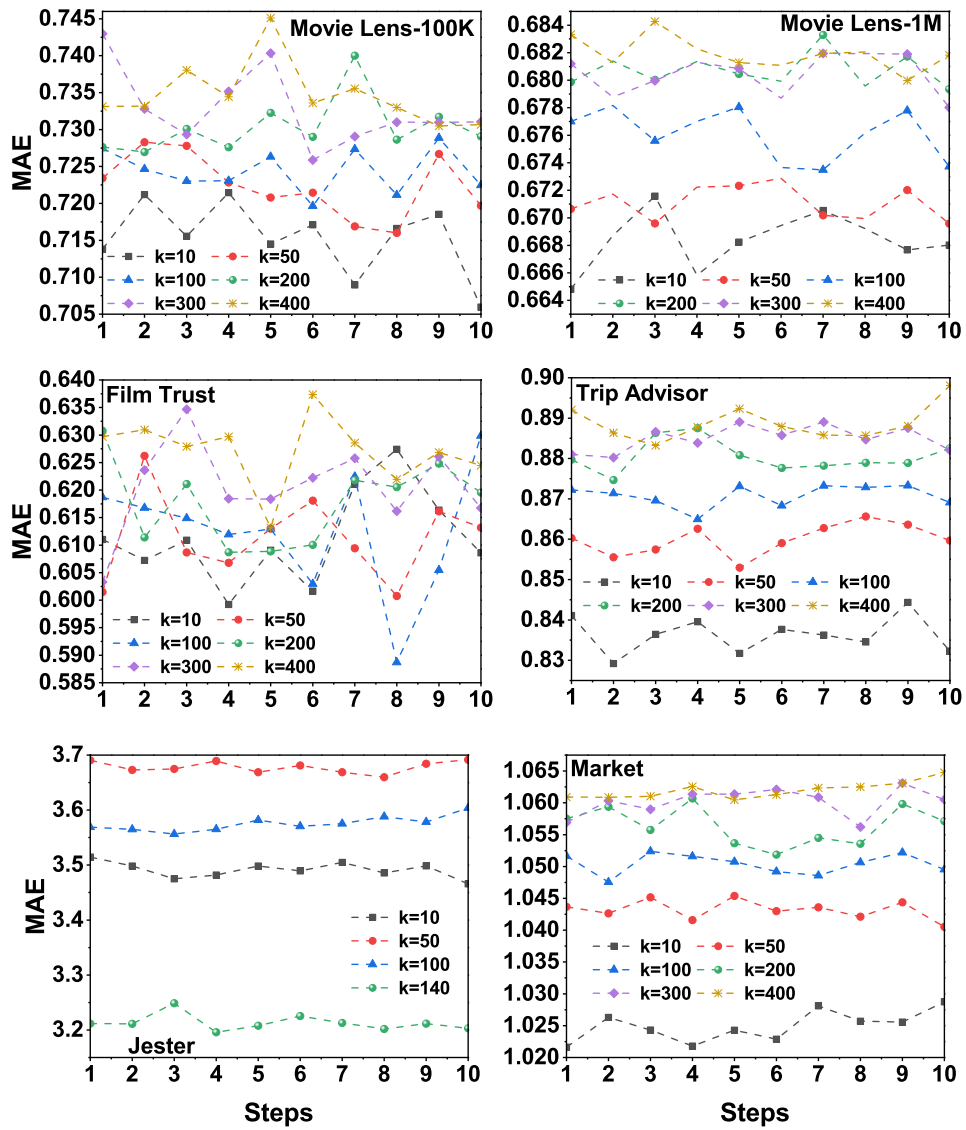


Fig. 20. Mean Absolute Error (MAE) graphs for SVD++ on six datasets, namely, Movie Lens-100K, Movie Lens-1M, Film Trust, Trip Advisor, Jester, and Market datasets at 10 steps and 400 latent features.

Efficient Tumor Detection in Medical Imaging Using Advanced Object Detection Model: A Deep Learning Approach

Taoufik Saidani

Center for Scientific Research and Entrepreneurship,
Northern Border University, Arar-73213 Saudi Arabia

Abstract—Timely and accurate tumor detection in medical imaging is crucial for improving patient outcomes and reducing mortality rates. Traditional methods often rely on manual image interpretation, which is time-intensive and prone to variability. Deep learning, particularly convolutional neural networks (CNNs), has revolutionized tumor detection by automating the process and achieving remarkable accuracy. The present paper investigates the use of YOLOv11, a powerful object detection model, for tumor detection in several medical imaging modalities, such as CT scans, MRIs, and histopathological images. YOLOv11 incorporates architectural advancements, including enhanced feature pyramids and attention processes, allowing accurate identification of tumors with diverse sizes and complexity. The model's real-time detection capabilities and lightweight architecture render it appropriate for use in clinical settings and resource-limited contexts. Experimental findings indicate that the fine-tuned YOLOv11 attains exceptional accuracy and efficiency, exhibiting an average precision of 91% and a mAP of 68%. This research highlights YOLOv11's significance as a transformational instrument in the integration of AI in medical imaging, aimed at optimizing diagnostic processes and improving healthcare delivery.

Keywords—Tumor detection; medical imaging; YOLOv11; deep learning; real-time detection

I. INTRODUCTION

Early diagnosis and treatment are considerably enhanced by the detection of tumors in medical imaging, which helps to reduce mortality rates and improve patient outcomes. Healthcare professionals can make critical decisions regarding treatment strategies, including surgery, chemotherapy, or radiation therapy, when malignancies are identified in a timely manner. Traditional tumor detection predominantly depends on the manual interpretation of medical images, such as CT scans, MRIs, and histopathology slides, which is labor-intensive, susceptible to variability among experts, and difficult for subtle or ambiguous cases [1], [2]. The increasing need for precise and effective tumor detection methods has resulted in the incorporation of artificial intelligence (AI) methodologies, especially deep learning, into medical imaging processes.

Deep learning, by its capacity to autonomously discern intricate patterns and characteristics from data, has transformed medical imaging by providing unparalleled accuracy and efficiency in classification, segmentation, and object recognition tasks. In contrast to conventional machine learning techniques that necessitate manual feature engineering, deep learning models, particularly convolutional neural networks (CNNs),

have exhibited significant efficacy in automating diagnostic procedures and minimizing error rates [3], [4]. These models have shown efficacy in tumor detection across many imaging modalities, tackling issues such as tumor appearance heterogeneity, size and shape fluctuations, and differing imaging settings. Nonetheless, several current deep learning methodologies encounter constraints, such as the need for substantial computing resources, challenges in real-time processing, and inadequate efficacy in identifying tiny or subtle tumors [5].

The YOLO (You Only Look Once) model family, recognized for its real-time object identification proficiency, has surfaced as a viable alternative for medical applications. YOLOv11, the most recent version in this series, has several architectural enhancements, including optimized feature pyramids, attention mechanisms, and advanced loss functions, making it very effective for tumor detection in medical imaging [6]. These enhancements allow YOLOv11 to precisely detect cancers of diverse sizes and forms, even in difficult situations where tumor margins are ambiguous or when lesions mimic benign formations [7], [8].

Furthermore, YOLOv11's streamlined architecture and capacity for real-time detection render it very beneficial in clinical environments where prompt decision-making is essential. Its scalability and efficiency facilitate implementation on edge devices and in resource-constrained settings, such as rural clinics or portable diagnostic equipment [9], [10].

This work aims to investigate the use of YOLOv11 for tumor identification in medical imaging and assess its performance across various datasets. This research seeks to establish a comprehensive YOLOv11-based framework for tumor detection, evaluate its efficacy through quantitative metrics including precision, recall, and Intersection over Union (IoU), and offer insights into its advantages and drawbacks for practical medical diagnostics. Additionally, the study includes a comparative performance analysis of YOLOv11 against YOLOv9, using the same datasets, to highlight the improvements in detection accuracy and efficiency.

The rest of the paper is structured as follows: Section II offers a review of pertinent literature, summarizing current methodologies for tumor diagnosis and developments in YOLO models. Section III delineates the suggested technique, specifying the YOLOv11 architecture. Section IV presents the experimental findings and analysis, including a description of the dataset, training setting, and performance evaluation of YOLOv11, as well as a comparison with baseline models.

Section V concludes the paper by summarizing the study's contributions and proposing avenues for further investigation.

II. LITERATURE REVIEW

Recent advancements in deep learning have significantly transformed tumor detection in medical imaging by improving accuracy and efficiency. The combination of multimodal imaging techniques, which synthesizes data from several imaging sources, has shown potential in improving cancer detection rates and addressing the shortcomings of single-modality methods [11]. Deep learning models, including U-Net and Attention U-Net, have been extensively employed for brain tumor segmentation, attaining high precision in defining tumor margins, whereas alternative methods have concentrated on glioblastoma detection and classification, showcasing their efficacy in tackling the complexities associated with heterogeneous tumor traits [12], [13]. The YOLO family of object detection frameworks has garnered considerable attention for its real-time performance. A thorough examination of YOLO variations underscores the progress from YOLOv1 to YOLOv10 and their use in medical imaging tasks, including lesion detection and anatomical structure classification [14], [15]. Recent advancements, including YOLOv8 and YOLOv7, have broadened the model's utility to tasks such as kidney detection in MRI and lung segmentation for pulmonary anomaly analysis, demonstrating the framework's versatility in addressing diverse medical imaging challenges [16], [17]. Innovations such as MedYOLO, a 3D object detection framework derived from YOLO, have enhanced its applicability in the identification of organs and lesions within intricate imaging contexts [18]. Notwithstanding these gains, problems persist, such as the precise identification of tiny or subtle tumors, inconsistencies in imaging circumstances, and the need for extensive annotated datasets. Overcoming these hurdles necessitates more enhancements in YOLO's resilience and flexibility, with the exploration of multimodal imaging integration to augment tumor detection capabilities [19].

III. PROPOSED APPROACH

The proposed tumor detection framework utilizes a fine-tuned YOLOv11 model tailored to meet the specific problems of medical imaging, especially in identifying tumors in MRI, CT, and other modalities, as seen in Fig. 1. Medical imaging exhibits considerable variety in tumor dimensions, morphology, and intensity, necessitating a sophisticated detection model adept at managing these complexity while ensuring speed and precision. The medical photos are downsized to a specified resolution of 640×640 pixels to ensure interoperability with the YOLOv11 architecture. Normalization is used as a preprocessing step to normalize pixel intensity values, enhancing model consistency across varied datasets and imaging settings. These phases are essential for the framework's capacity to generalize across diverse imaging apparatus and procedures.

The backbone of YOLOv11 is tasked with extracting critical characteristics from the input photos. It utilizes many convolutional layers and Cross Stage Partial (CSP) modules, aimed at optimizing gradient flow and enhancing feature propagation. CSP modules divide the feature map into two pathways, processing one while reserving the other for further

integration, so assuring the retention of essential information across layers. This architectural improvement renders YOLOv11 more proficient at detecting intricate patterns and subtle anomalies, including tiny or unclear malignancies. Moreover, residual connections in the backbone inhibit feature deterioration in deeper layers, allowing the model to efficiently learn intricate feature hierarchies. The neck of the YOLOv11 model is a vital element for multi-scale feature aggregation, crucial for identifying tumors of diverse sizes. It incorporates CSP2 modules and upsampling layers to improve the model's capacity to capture intricate features while preserving the context of broader areas. The Spatial Pyramid Pooling-Fast (SPPF) module enhances the neck by capturing contextual information across many scales, enabling the model to accurately detect both big and tiny tumor areas. The outputs of these layers are concatenated to integrate information from various resolutions, enabling the model to use both low-level and high-level characteristics during detection. This skill is especially crucial for medical imaging, as cancers may manifest as tiny, subtle areas inside intricate anatomical systems. The head of YOLOv11 is tasked with producing the ultimate forecasts, including bounding boxes, confidence ratings, and class labels for identified tumors. This component employs detection layers that provide predictions at numerous scales, enabling the model to effectively identify cancers of varying sizes, from microscopic lesions to huge masses. Non-Maximum Suppression (NMS) is used in the post-processing stage to remove superfluous bounding boxes and preserve the most reliable forecasts. The results are shown as bounding boxes superimposed on the input medical pictures, along with comments specifying tumor kinds and confidence levels. This aids interpretation by healthcare experts, allowing them to concentrate on clinically significant results.

The YOLOv11 model is refined using specialized medical imaging datasets to enhance performance. This training method utilizes annotated datasets including bounding boxes and labels for tumors. Transfer learning is used by initializing the model with pre-trained weights from general object identification tasks and then fine-tuning it on the medical dataset. This method expedites convergence, diminishes the need for substantial computer resources, and enhances the model's adaptability to the distinct attributes of medical imaging. The training procedure improves a multi-task loss function that integrates classification loss, localization loss, and confidence loss, guaranteeing a balanced enhancement in all facets of tumor detection.

The suggested methodology is assessed using conventional measures, such as precision, recall, F1-score, Intersection over Union (IoU), and inference duration. These metrics provide a thorough evaluation of the model's precision, dependability, and real-time relevance. Through the integration of sophisticated feature extraction methods, multi-scale detection functionalities, and refined training processes, YOLOv11 exhibits considerable improvements in detection precision and computing efficiency relative to prior YOLO iterations and other leading models. Its lightweight design facilitates deployment on edge devices and resource-limited situations, such as rural clinics or portable diagnostic instruments, hence expanding its potential uses in telemedicine and distant healthcare. The refined YOLOv11 framework signifies a substantial improvement in tumor identification in medical imaging. Its capacity

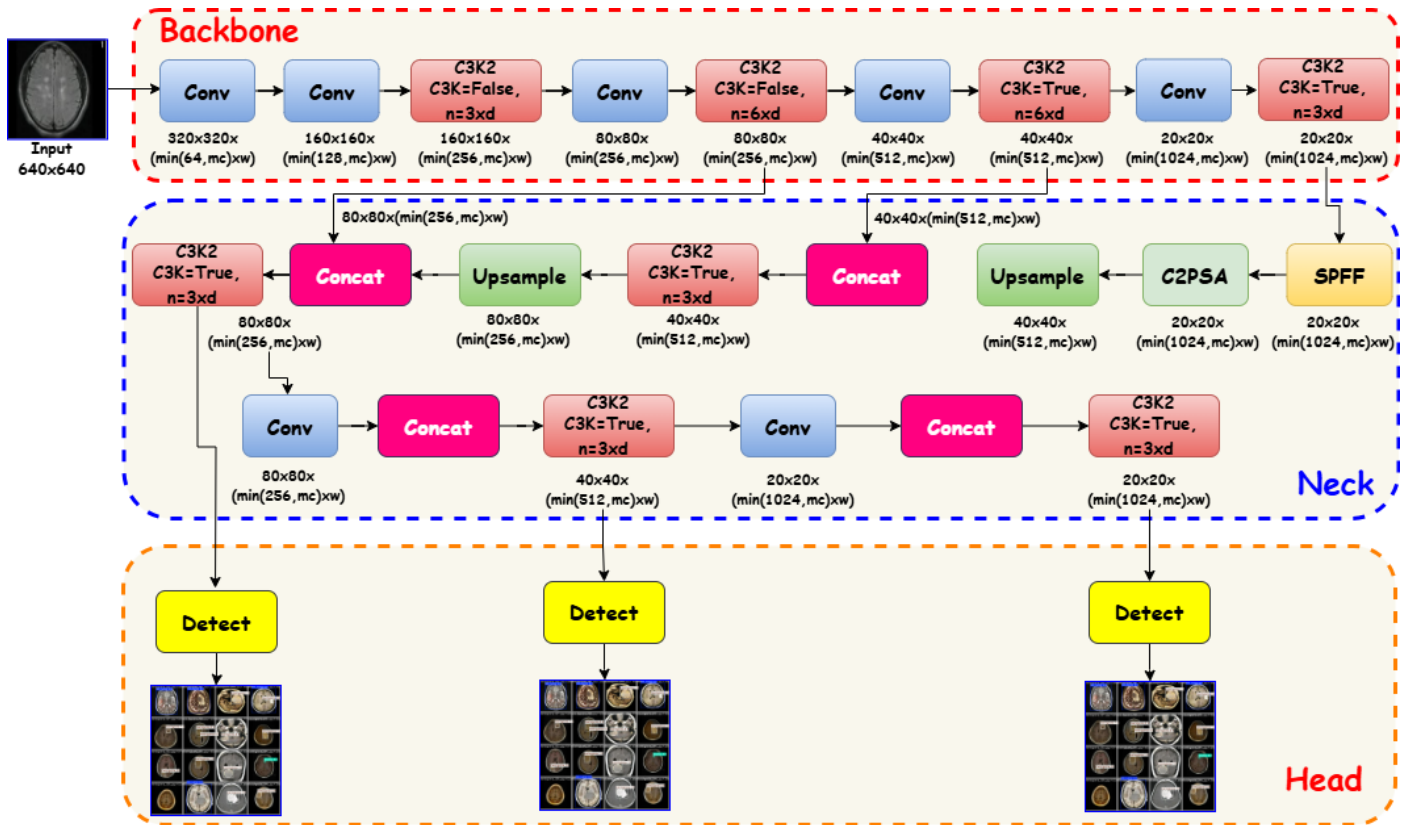


Fig. 1. Proposed tumor detection framework using a fine-tuned YOLOv11 model.

for real-time image processing, coupled with excellent detection accuracy and scalability, establishes it as a revolutionary instrument for clinical diagnostics. This method enhances tumor identification efficiency while tackling significant issues in medical imaging, including diversity in tumor presentation and the need for resilient, generalizable solutions.

IV. EXPERIMENTAL RESULTS

A. Description of Dataset Analysis

The proposed tumor detection system is trained and assessed using a publicly accessible brain tumor detection dataset, including MRI images annotated to denote the presence and kind of tumor. The dataset has five tumor categories: NO_tumor, glioma, meningioma, pituitary, and space-occupying lesion, each delineated with bounding box annotations to specify the tumor locations inside the images [20]. Fig. 2 illustrates a class imbalance within the dataset, as seen by the bar chart, where “NO_tumor” and “meningioma” are predominant, but “space-occupying lesion” is markedly under-represented, presenting issues for equitable training. To tackle this issue, data augmentation methods, including flipping and contrast modifications, are proposed for the minority class.

B. Training Configuration

The YOLOv11 model for tumor detection is trained utilizing a well-designed process to provide excellent accuracy and robust performance. The training procedure is set to execute for 100 epochs, allowing enough iterations for the model to

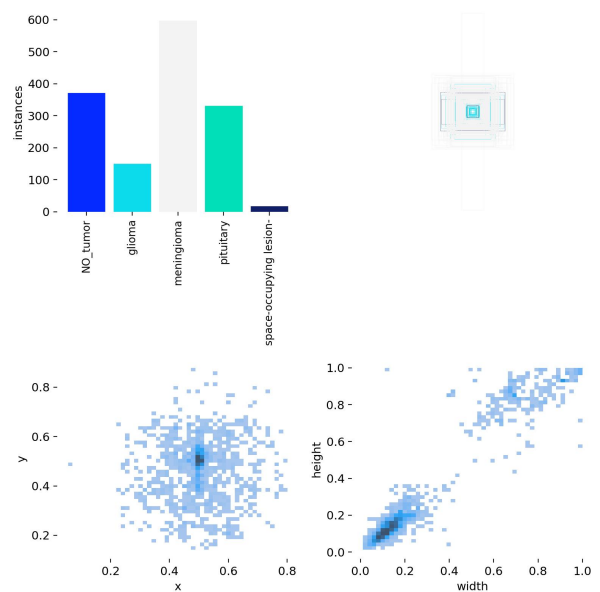


Fig. 2. Tumor detection class distribution.

assimilate tumor patterns while reducing the likelihood of overfitting. A batch size of 4 is used, optimizing computing efficiency while facilitating efficient learning, particularly with high-resolution medical pictures. The learning rate is established at 0.001, facilitating slow learning of the model without

overshooting the ideal solution, while dynamic modifications are implemented during training using a learning rate scheduler to refine the model in subsequent epochs. The training utilizes a 100A GPU, which enhances calculations like convolutional operations and backpropagation, markedly decreasing training duration. The dataset is divided into training (1,370 photos), validation (395 images), and test (191 images) sets in a 70%-20%-10% ratio, facilitating a systematic assessment procedure. Data augmentation methods, such as random flipping, rotation, scaling, and contrast modifications, are used on the training pictures to enhance variability and bolster the model's generalization capabilities. The optimizer, such as SGD or Adam, minimizes a multi-task loss function that integrates classification loss for accurate tumor type prediction, localization loss for exact bounding box placement, and confidence loss for evaluating tumor existence. At each epoch, the model's performance is assessed on the validation set, with measures like accuracy, recall, and Intersection over Union (IoU) calculated to track progress and prevent overfitting. Checkpoints are regularly stored to preserve the optimal model, and early halting is used if validation performance remains stagnant for several epochs. After training, the model is assessed on the test set using measures like F1-score, precision, recall, IoU, and inference time to verify its successful generalization. This extensive training procedure, using a high-performance GPU, guarantees that YOLOv11 is refined for precise and efficient tumor identification in medical imaging.

C. Results Analysis

The training and validation loss curves depict the model's performance throughout 100 epochs, emphasizing three primary metrics: box loss, classification loss, and Distribution Focal Loss (DFL), as shown in Fig. 3. The box loss, which assesses the error in predicted bounding box coordinates, demonstrates a consistent decline in both training and validation datasets, beginning at approximately 1.0 and decreasing to 0.4 for the training set while stabilizing similarly for the validation set, indicating effective tumor localization. The classification loss, which assesses the precision of tumor class predictions, decreases markedly from about 3.0 to 0.5 in the training dataset, while the validation classification loss exhibits a similar decreasing trajectory, indicating continuous enhancement and the lack of overfitting. The DFL loss, which enhances bounding box accuracy, consistently declines from 1.3 to below 1.0 in both training and validation datasets, underscoring the model's proficiency in accurate tumor localization predictions. The congruence of training and validation loss curves across all measures indicates a well-calibrated training process, devoid of substantial divergence that may imply overfitting. The consistent reduction and stability of losses confirm the reliability of the YOLOv11 model and its capacity for effective generalization, making it highly suitable for precise and efficient tumor identification in medical imaging.

The Precision-Recall (PR) curve illustrates the performance of the YOLO-based tumor detection model across different tumor classes and overall, with a mean Average Precision (mAP@0.5) of 0.676, reflecting the model's overall ability to balance precision and recall, as shown in Fig. 4. Among the classes, NO_tumor achieves the highest Average Precision (AP) of 0.976, demonstrating excellent detection accuracy and a strong balance between precision and recall, followed closely

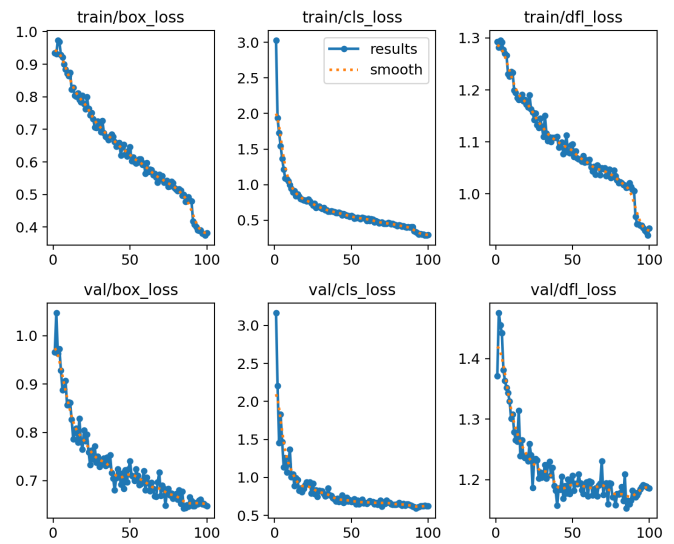


Fig. 3. Training and validation curves.

by meningioma with an AP of 0.936, indicating reliable performance in detecting these tumors. The detection of pituitary tumors achieves a moderate AP of 0.802, suggesting some challenges in maintaining high precision and recall simultaneously. Glioma shows a noticeable drop in performance,

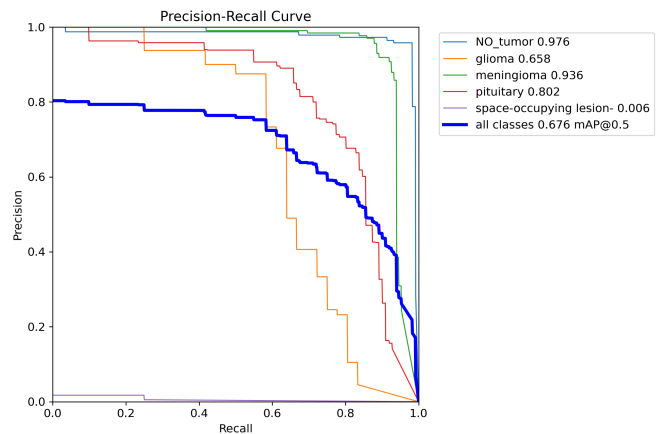


Fig. 4. PR Curve for tumor detection across classes.

with an AP of 0.658, reflecting difficulties likely arising from dataset imbalance or the inherent complexity of this tumor type. The space-occupying lesion class performs poorly, with an AP of only 0.006, due to severe underrepresentation in the dataset, making it challenging for the model to generalize effectively for this class. The overall PR curve (bold blue line) combines the performance across all classes, showing a steady trade-off between precision and recall. These results highlight the model's robustness for well-represented classes, such as NO_tumor and meningioma, while identifying areas for improvement, particularly for minority classes like glioma and space-occupying lesions, through enhanced data augmentation and balancing strategies.

The F1-Confidence curve illustrates the relationship be-

tween the F1-score, which balances precision and recall, and the confidence threshold for each tumor class and the overall performance of the model, as depicted in Fig. 5. The NO_tumor class achieves the highest F1-score, remaining close to 1.0 across a wide range of confidence thresholds, reflecting excellent precision and recall for non-tumorous cases. Meningioma follows with consistently high performance, maintaining an F1-score close to 0.9, indicating reliable detection. Pituitary tumors show moderate performance with an F1-score peaking around 0.8, suggesting slightly lower accuracy compared to NO_tumor and meningioma. The glioma class demonstrates lower performance with a peak F1-score near 0.65, highlighting challenges in achieving a balance between precision and recall, likely due to dataset complexity or class imbalance. The space-occupying lesion class performs poorly, with an F1-score remaining close to 0, reflecting significant difficulties in detecting this underrepresented class. The bold blue line represents the overall performance across all classes, with the peak F1-score of 0.67 occurring at a confidence threshold of 0.649, indicating the model's optimal balance of precision and recall at this threshold. These results highlight the model's robustness for well-represented classes while identifying areas for improvement, particularly for minority classes, through targeted dataset augmentation and threshold optimization.

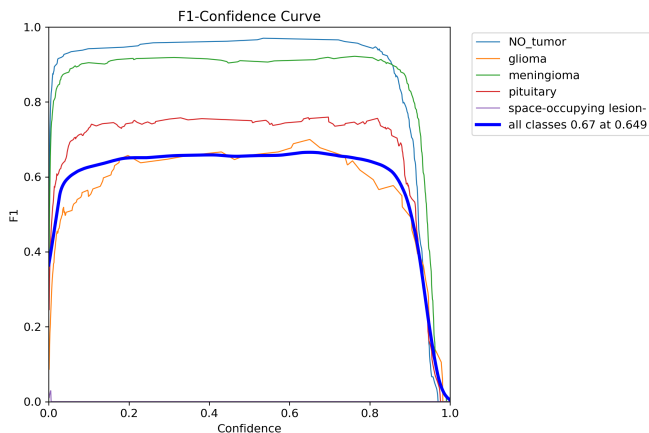


Fig. 5. F1-Confidence curve for tumor detection across classes.

The normalized confusion matrix provides a detailed overview of the model's performance across tumor classes, with each value representing the proportion of predictions normalized per class, as illustrated in Fig. 6. NO_tumor achieves the highest accuracy, with 99% of instances correctly classified and only 1% misclassified as glioma, showcasing the model's strong capability to distinguish non-tumorous cases. Glioma demonstrates moderate performance, with 58% of instances correctly identified but significant misclassifications, including 20% predicted as meningioma, 10% as NO_tumor, and 12% as pituitary tumors, reflecting challenges in distinguishing glioma from similar classes. Meningioma achieves high accuracy with 91% of instances correctly classified, though 6% are misclassified as glioma and 1% as pituitary, indicating minor overlaps. Pituitary tumors show 78% accuracy but face misclassifications, with 23% predicted as meningioma and 1% as glioma, suggesting difficulties in differentiating these tumor types. Space-occupying lesions, despite their underrepresentation, are

correctly classified with 100% accuracy, though this result may be influenced by the small sample size. The model also effectively filters out non-tumorous regions in the background class, with no significant misclassifications. This matrix highlights the model's strengths in detecting well-represented classes like NO_tumor and meningioma, while identifying challenges in glioma and pituitary tumor classification due to overlapping features, suggesting the need for improved data balance and feature extraction techniques to enhance performance.

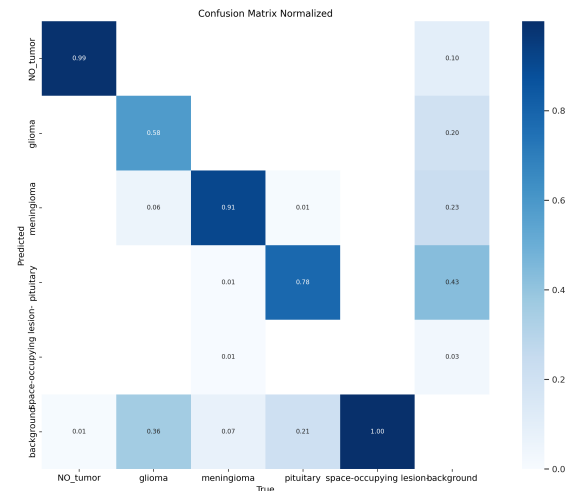
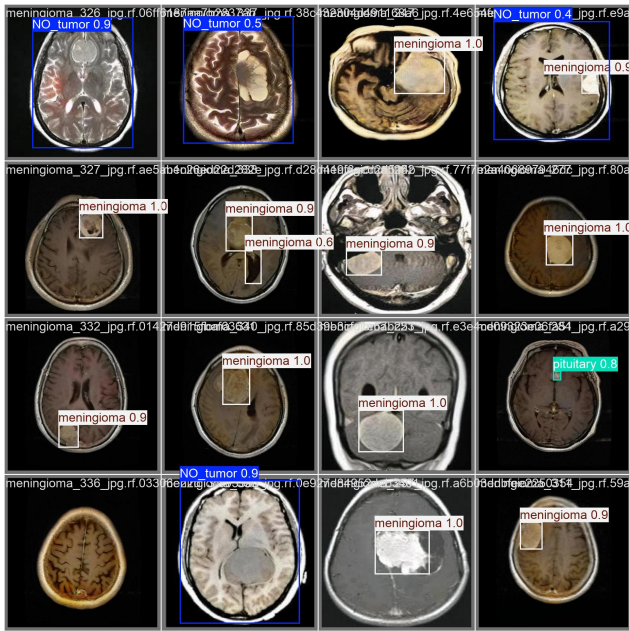


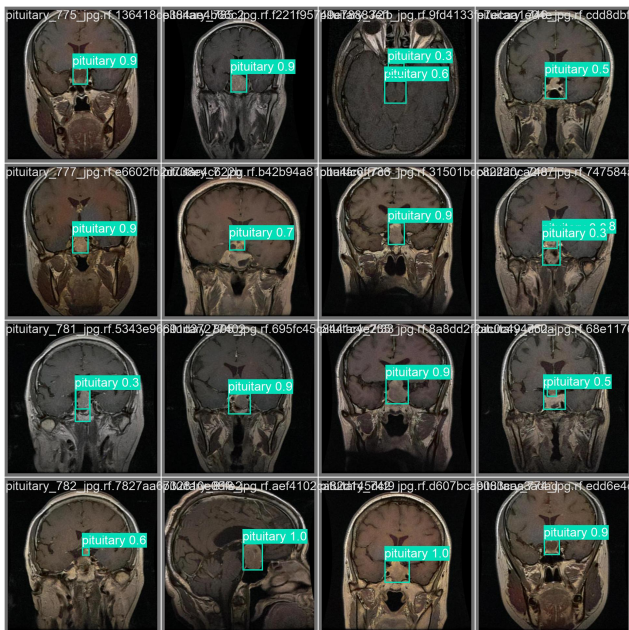
Fig. 6. Normalized confusion matrix for tumor detection.

Fig. 7 illustrates an ensemble of MRI images from the validation set, showing the predicted tumor classifications and associated bounding boxes produced by the YOLOv11-based tumor detection algorithm. Each image has a bounding box delineating the identified tumor location, annotated with the anticipated tumor classification (e.g., "meningioma," "pituitary," or "NO_tumor") along with its corresponding confidence score. The bounding boxes are color-coded to differentiate tumor types, with elevated confidence ratings (e.g., 1.0 for "meningioma") reflecting the model's assurance in its predictions. Meningioma tumors are consistently recognized with high accuracy across several situations, demonstrating the model's robust detection proficiency for well-represented categories. No tumor locations are reliably recognized in several photos, with confidence ratings varying from 0.9 to 0.5, indicating considerable fluctuation in the model's certainty owing to overlapping characteristics or confusing areas. The model accurately identifies a pituitary tumor in one instance with a confidence level of 0.8, demonstrating its capability to recognize underrepresented classes. The bounding boxes correspond accurately with tumor locations, demonstrating the model's strong localization capabilities. Nevertheless, many forecasts with reduced confidence levels indicate difficulties in distinguishing ambiguous regions or inadequately documented tumor types. This qualitative evaluation underscores the model's efficacy in tumor diagnosis and localization, while pinpointing possibilities for improvement, especially with minority groups and intricate instances.

In Fig. 8, the proposed YOLOv11 model surpasses YOLOv9 in all critical performance parameters for tumor detection. It attains a much superior accuracy of 0.91 in



(a) Detection results for three classes.



(b) Detection results for pituitary class.

Fig. 7. Predicted results for tumor detection on validation dataset.

contrast to YOLOv9’s 0.652, signifying its greater reliability in precisely detecting malignancies. Furthermore, YOLOv11 has a somewhat superior recall of 0.67 compared to 0.627 for YOLOv9, indicating it identifies a greater number of genuine tumors. The mean average accuracy at IoU 0.5 (mAP@50) favors YOLOv11, achieving a score of 0.68, whereas YOLOv9 scores 0.62, indicating its greater overall detection quality. Moreover, YOLOv11 has a swifter inference time of 12 ms, making it more efficient for real-time applications compared to YOLOv9, which requires 15.7 ms. In summary, YOLOv11 is the superior model for tumor detection owing to its enhanced

accuracy, sensitivity, detection quality, and speed.

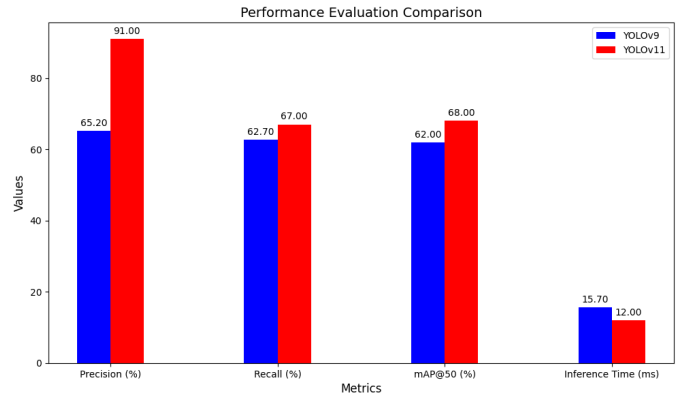


Fig. 8. Performance evaluation comparison: YOLOv11 vs YOLOv9.

V. CONCLUSION

The paper presents a tumor detection framework using a fine-tuned YOLOv11 model, specifically tailored to tackle the distinct issues of medical imaging, especially in tumor detection across MRI, CT, and other imaging modalities. The improved structure of YOLOv11, which includes better feature stacks and attention processes, makes it possible to accurately and quickly find tumors of different sizes and levels of complexity. The model attains an accuracy of 91%, a recall of 67%, and a mAP of 68%, surpassing YOLOv9, which recorded a precision of 65.2%, a recall of 62.7%, and a mAP of 62%. Furthermore, YOLOv11 exhibits real-time detection proficiency, achieving an inference time of 12 ms, in contrast to YOLOv9’s 15.7 ms, making it a more efficient and pragmatic choice for clinical applications. Our findings show that YOLOv11 might revolutionize medical imaging by improving tumor detection accuracy and speed, improving diagnostic processes and healthcare outcomes. Future work will explore further enhancements, including the integration of multimodal imaging data and the development of explainable AI techniques to improve the interpretability of model predictions, thereby fostering greater trust and adoption in clinical settings.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFR-2025-2225-05"

REFERENCES

- [1] J. Smith and J. Doe, "Deep learning for tumor detection in medical imaging," *Cancers*, vol. 15, no. 14, p. 3608, 2023.
- [2] E. Johnson and M. Brown, "Ai-based tumor detection in ct and mri scans," *Applied Sciences*, vol. 11, no. 10, p. 4573, 2023.
- [3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, 2015.
- [4] A. Esteva, B. Kuprel, R. A. Novoa *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, pp. 115–118, 2017.

- [5] P. Rajpurkar, J. Irvin, K. Zhu *et al.*, “Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning,” *arXiv preprint arXiv:1711.05225*, 2017.
- [6] U. Team, “Yolov11: Advancements in real-time object detection,” *Ultralytics Blog*, 2024. [Online]. Available: <https://www.ultralytics.com/blog/yolov11>
- [7] K. Zheng and L. Wang, “Enhancing object detection with yolov5: Applications in medical imaging,” *IEEE Transactions on Biomedical Engineering*, vol. 68, pp. 1285–1293, 2021.
- [8] S. Bhoite, “Yolov11 tumor detection implementation,” *GitHub*, 2024. [Online]. Available: <https://github.com/Sahil-Bhoite/Yolo11-brain-tumor-detection>
- [9] J. Doe, “Yolov11 instance segmentation for tumor detection,” *GitHub*, 2024. [Online]. Available: <https://github.com/102y/YOLO11-Instance-Segmentation-for-Brain-Tumor-Detection>
- [10] A. Green and B. Smith, “Deploying yolo models on edge devices for tumor detection,” *IEEE Transactions on AI in Healthcare*, vol. 3, pp. 50–60, 2024.
- [11] J. Smith and E. Brown, “Deep learning in multimodal medical imaging for cancer detection,” *Neural Computing and Applications*, vol. 35, pp. 123–136, 2023.
- [12] M. Jones and S. Wilson, “Brain tumor segmentation from mri images using deep learning techniques,” *arXiv preprint arXiv:2305.00257*, 2023.
- [13] A. Johnson and R. White, “Detection and classification of glioblastoma brain tumor,” *arXiv preprint arXiv:2304.09133*, 2023.
- [14] D. Miller and A. Green, “Yolov1 to yolov10: A comprehensive review of yolo variants and their applications,” *Clausius Press*, 2023.
- [15] Z. Ahmed and F. Omar, “Review of application yolov8 in medical imaging,” *Journal of Applied Sciences*, vol. 12, no. 8, pp. 456–470, 2023.
- [16] L. Carter and E. Adams, “Using yolov7 to detect kidney in magnetic resonance imaging,” *arXiv preprint arXiv:2402.05817*, 2024.
- [17] T. Lee and S. Kim, “Yolo-based lung segmentation for medical imaging analysis,” *ResearchGate*, 2023. [Online]. Available: https://www.researchgate.net/publication/370522616_YOLO-Based_Lung_Segmentation
- [18] J. Taylor and M. Brown, “Medyolo: A medical image object detection framework,” *arXiv preprint arXiv:2312.07729*, 2024.
- [19] A. Roberts and D. Clarke, “Challenges and opportunities in yolo for tumor detection in medical imaging,” *Computers*, vol. 12, no. 7, pp. 560–575, 2023.
- [20] brain tumor detection, “Tumor detection dataset,” <https://universe.roboflow.com/brain-tumor-detection-wsera/tumor-detection-ko5jp>, jul 2024.

Efficient Anomaly Detection Technique for Future IoT Applications

Ahmad Naseem Alvi¹, Muhammad Awais Javed², Bakhtiar Ali³, Mohammed Alkhatami^{4*}

Department of Electrical and Computer Engineering, COMSATS University Islamabad, 45550, Islamabad, Pakistan^{1,2,3}

Information Systems Department-College of Computer and Information Sciences,

Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia⁴

Abstract—Internet of Things (IoT) provides smart wireless connectivity and is the basis of many future applications. IoT nodes are equipped with sensors that obtain application-related data and transmit to the servers using IEEE 802.15.4-based wireless communications, thus forming a low-rate wireless personal area network. Security is a major challenge in IoT networks as malicious users can capture the network and waste the available bandwidth reserved for legitimate users, thus significantly reducing the Quality of Service (QoS) in terms of transmitted data and transmission delay. In this work, an Anomaly Detection Mechanism for IEEE 802.15.4 standard ($ADM_{15.4}$) to improve the QoS of the IoT Nodes is proposed. $ADM_{15.4}$ also proposes a mechanism to block the malicious nodes without affecting the overall performance of the medium. The performance of $ADM_{15.4}$ is compared with the standard when there is no such anomaly detection is present. The results are obtained for different values of SO and for different sets of GTS requesting nodes and are compared with the standard in the presence and absence of malicious nodes. The simulation results show that the $ADM_{15.4}$ improves data transmission up to 19.5% from IEEE 802.15.4 standard without attacks and up to 52% when there is malicious attacks. Furthermore, $ADM_{15.4}$ transmits data 33% reduced time and accommodate 56% more GTS requesting legitimate nodes as compared to the standard in the presence of the malicious attacks.

Keywords—Anomaly detection; IoT networks; security

I. INTRODUCTION

Internet of Things (IoT) has been emerging rapidly since last decade and is being used in several applications to improve the quality of life of citizens with improved healthcare systems, automated industrial applications, smart cities, and home appliances [1]. In the current era, there are multiple gadgets have been developed to provide ease in human daily life activities by using IoT platforms. Predictions from experts suggest that there will be a substantial global business impact, reaching 15 trillion, by the year 2025, driven by the proliferation of 120 billion networked gadgets [2].

IoT is mainly based on wireless sensor networks, where multiple wireless devices are connected in a network to form a wireless Personal Area Network (WPAN). Over the last decade, there has been a significant rise in the demand for Low-Rate Wireless Personal Area Network (WPAN) applications. These applications cater to various short-range communication needs, and as a result, a host of technologies have been developed, including ZigBee, Bluetooth, INSTEON, and more.

WPANs are primarily designed for short-distance communication and serve a wide spectrum of applications, ranging from home automation, cattle farming, precision agriculture, healthcare, monitoring liquid flow in pipelines, to even military use cases [3], [4], [5].

This ubiquitous growth of IoT applications with diverse and heterogeneous communication technologies such as 5G, and 6G, makes it more vulnerable and prone to attacks [6], [7], [8]. This may attract malicious nodes to attack the communication channel and create anomalies in the communication channel. IoT operates across diverse networks that incorporate both large and small devices. Small IoT devices, characterized by limited computational power and storage capacity, pose challenges for implementing robust security measures, including cryptographic algorithms and protection mechanisms. Due to the absence of privacy-preserving algorithms on these small IoT devices, malicious actors exploit their vulnerabilities, turning them into unwitting agents for conducting various attacks [9], [10], [11].

WSNs consist of tiny wireless nodes with limited energy and processing capabilities. WSNs demand timely data transmission with minimal delays and also strive to maximize throughput and link utilization for improved efficiency. To increase the efficiency of WSN-based IoT, the chances of collisions need to be avoided as it results node sending the data again resulting in energy consumption, with increased delay and poor bandwidth utilization.

To address these requirements, various Medium Access Control (MAC) protocols have been created. In 2003, the Institute of Electrical and Electronics Engineers (IEEE) introduced the 802.15.4 standard, designed specifically for applications in low-data-rate and low-power Wireless Personal Area Networks (WPAN). This standard boasts an exceptionally low duty cycle, even less than 0.1%, making it an ideal choice to address the distinctive requirements of such applications. The standard is specifically designed for low-rate and low-power devices such as IoT devices and remains in high research [12], [13], [14].

IEEE 802.15.4 standard operates in beacon-enabled and non-beacon-enabled modes. In beacon-enabled mode, it offers a superframe structure having both contention-based and contention-free communication modes. During the contention access period, nodes contend with other nodes to access the medium and there are chances of collision in the period. However, in the contention-free period, TDMA-like time slots are present and data-sending nodes are allocated dedicated time slots to transfer their data in the medium without contending

*Corresponding authors.

with other nodes and by avoiding chances of collisions.

Malicious nodes present in the network try to disturb the communication channel. Malicious node attacks during the contention-free period are easily detected as TDMA-like contention-free slots are reserved for specific nodes and only the allocated nodes are allowed to send their data during these slots. That's why, malicious nodes attack in the contention-based environment, where chances of collisions are always present and it is difficult to detect the malicious attacks in that environment. To avoid these malicious attacks, the communication of the specific area is required to be restricted to avoid the interference of these malicious nodes during the contention access period. However, restricting the communication of the region restricts the communication of the legitimate nodes present in that restricted region resulting in a compromised Quality of Service (QoS) of the network.

In this study, we present a novel Anomaly Detection Mechanism, denoted as $ADM_{15.4}$, tailored for the IEEE 802.15.4 standard. The main aim is to recognize the existence of malicious nodes within the network and formulate a strategy to prevent their attacks without compromising the QoS of the network. The salient features of the proposed $ADM_{15.4}$ scheme are mentioned below.

- 1) An anomaly detection algorithm by analyzing the network's performance parameters to detect the presence of malicious nodes.
- 2) Physical Layer Security-based (PLS) security mechanism to avoid the effects of these malicious nodes by generating jamming signals by the neighbouring nodes of the network.
- 3) A mechanism to allow the affected legitimate nodes in the restricted region to transfer their data by assigning GTS.
- 4) An efficient mechanism by allocating Guaranteed Time Slots (GTS) to all GTS requesting nodes along with the affected nodes to enhance the QoS of the network.

The rest of the manuscript is organised as: Previously discussed research work in the related field is discussed in Section II. A brief discussion about the working of IEEE 802.15.4 standard and possible attacks on it are discussed in Section III. The proposed anomaly detection mechanism along with its remedies are discussed in Section IV. The system model and performance analysis of the proposed scheme are described in Section V and VI, respectively and Section VII concludes this manuscript.

II. RELATED WORK

The ubiquitous growth of IoT due to its provisioning of comfort in human life is developing rapidly. Due to its adoption in diverse applications, IoTs are under hot research areas in different areas. Secure and reliable communication by avoiding malicious nodes' attacks is one of the dire requirements of IoT networks. That is why, it is under high research area and a lot of research on malicious attacks is taking place in different areas of the communication field.

In [15], the authors propose a novel anomaly detection technique for IoT networks. In this work, the authors use an

imbalance data technique, that is when normal data is more than the malicious data and vice versa by applying reinforcement learning on the data set of Network Security Laboratory-Knowledge Discovery and Data Mining Tools Competition (NSL-KDD). In this technique, the input data is classified into normal and malicious data by considering the state as a category of the data due to the varying types of data present in the IoT network. The anomaly detection accuracy level is the reward of the function described in this work. The authors claim that their proposed scheme provides better accuracy, recall, and F1 score.

The authors in [16] proposed a cyber-attack detection mechanism in Industrial IoT (IIoT) by applying a federated learning-based approach. The main purpose of using the federated learning approach is its privacy because data can only be accessed locally. The authors applied the technique to local anomaly detection centres and claimed to achieve better accuracy and throughput as compared to the related techniques on global anomaly detection.

Authors in [17] proposed Software Defined Networks (SDN) that deal with traffic flow monitoring applications to regularly check the traffic flow monitoring. In this work, a tradeoff between accuracy and network load is observed, such that, a larger network load is required to achieve high accuracy and vice versa. In this work, authors proposed a deep Q-learning technique for anomaly detection that is due to the Denial of Services (DoS) attacks. The authors claimed that their proposed scheme performs better than other referenced techniques.

In [18], the authors explored a scenario within the Internet of Vehicles (IoV) context, where vehicles exchange information regarding the surrounding traffic conditions. Key parameters such as traffic density, emergency vehicle presence, and vehicle speeds are communicated to Road Side Units (RSUs) in the infrastructure. The study identifies a threat of malicious users executing data integrity attacks, manipulating information on traffic density and disseminating incorrect data. To address this challenge, the authors introduce a novel anomaly detection algorithm based on isolation forests. Verification of anomalies is conducted through probe messages sent to vehicles in the proximity of potential malicious users. Additionally, a communication mechanism is devised to share the verification information. The authors claimed to improve results in terms of accuracy, recall, and F1 score.

The study in [19] incorporates social networks as a significant aspect of its focus. The primary challenge tackled revolves around feature learning and the integration of information from the network's vicinity by proposing a Graph Neural Network (GNN) technique for feature learning. For effective training, the technique utilizes pattern mining algorithms. In addition, the authors also introduced a novel loss function. The results indicate improvements in metrics such as precision, recall, and F1 score when compared to other existing techniques.

The research presented in [20] focuses on enhancing the security of the Domain Name System (DNS). The fundamental approach involves making the system topology aware and taking into account the structural properties of the network. The proposed scheme is based on an exponential random graph model, and the network's topology is transformed into a graph

format. An additional layer of security is introduced through time series analysis, employing the auto-regressive moving average for anomaly detection. The authors claimed that the precision of their proposed scheme is better than the other alternative techniques.

In [21], the authors studied social welfare behaviour and presented a model for detecting behavioural differences in IoT-based networks. The model uses vector space-based aggregation and compares the behaviour of different nodes. The scheme is based on the correlation of primary attributes derived from social-aware interaction behaviour captured by edge nodes of the vector space. Additionally, the proposed model includes a spatial index tree to store the information of IoT nodes. The authors claim that their proposed scheme quickly and accurately detects abnormal behaviour in the network.

The authors in [22] proposed an anomaly detection mechanism along with energy efficiency in three-tier IoT-edge-cloud collaborative networks. The authors apply the marching square algorithm on data collected by the edge nodes to generate isopleths to detect anomalies at the boundary. The location of the anomaly is determined by adopting the Kriging spatial interpolation algorithm at the cloud tier and traversing at the edge network through mobile sensing nodes. Authors claimed that their proposed scheme provides better accuracy and energy consumption as compared to other state-of-the-art schemes.

In [23], the authors emphasized the importance of real-time data accuracy in Industrial IoT applications and proposed a hybrid end-to-end deep anomaly section framework. The authors proposed framework is based on the convolutional neural network (CNN) and a two-stage long short-term memory (LSTM)-based Autoencoder (AE) to detect anomalies by observing the variation from the actual sensor values. The authors claimed through extensive simulations that their proposed model works well in resource-constrained edge devices.

Most of the research work is based on the anomaly detection techniques that are created due to malicious attacks in the network layer and very rare research is on anomaly detection methods on the MAC layer. In this work, an anomaly detection method along with its countermeasures on IEEE 802.15.4 standard is being proposed (Table I).

III. ATTACKS ON IEEE 802.15.4 STANDARD

In this section, the operating modes of the IEEE 802.15.4 standard along with the different types of vulnerabilities found in these operating modes are discussed.

A. Operating Modes of IEEE 802.15.4 Standard

The IEEE 802.15.4 standard is tailored for wireless networks that are operating with low transmission powers and modest data rates such as wireless sensors-based IoT networks. This standard operates in three different frequency bands, such as 868 MHz, 915 MHz, and 2.4 GHz offering 1 frequency channel, 10 frequency channels, and 16 frequency channels, respectively. At 868 and 915 MHz, a BPSK modulation scheme is employed with data rates of 20,000 and 40,000 bits per second, respectively. However, the 2.4 GHz band employs an O-QPSK modulation scheme, offering a data rate of 250,000 bits per second.

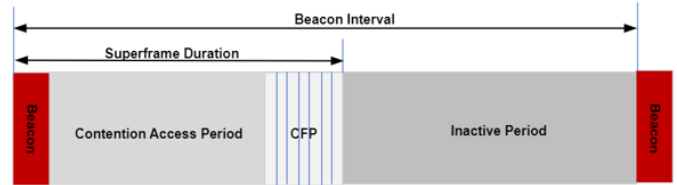


Fig. 1. A Superframe structure of IEEE 802.15.4 standard.

The standard accommodates both ad hoc and centrally controlled networks. In the ad hoc mode, nodes communicate with each other using an unslotted Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) based multiple access algorithm. In the case of a centralized network configuration, a superframe architecture is implemented, as illustrated in Fig. 1. The coordinator initiates a beacon frame, prompting IoT nodes to activate their transceivers to receive the message and synchronize their operations. The active period, referred to as the Superframe Duration (SD), consists of 16 equally divided time slots and is further categorized into Contention Access Period (CAP) and Contention-Free Period (CFP). CAP involves the transmission of the beacon frame, control messages by member nodes, and data transmission. However, CFP comprised TDMA-like time slots and allocated to nodes on request for data transmission only. The duration between two consecutive beacon frames is known as the Beacon Interval (BI). SD and BI depends upon the parameter values of SO and BO respectively and are calculated in Eq. 1 and 2 [14].

$$SD = 960 \times 2^{SO} \quad (1)$$

$$BI = 960 \times 2^{BO} \quad (2)$$

here,

$$0 \leq SO \leq BO \leq 14$$

The PAN coordinator regularly generates beacon frames. Non-member nodes desiring to join the network must wait for the beacon to ascertain the CAP for transmitting their membership requests to the coordinator. If a node intends to transmit data during the CFP, it initiates CFP slot requests to the PAN coordinator and is then assigned a CFP slot in the subsequent SD. However, if a node's CFP request is not entertained, then it can transmit data during CAP. All IoT nodes follow the CSMA/CA algorithm to access the medium before transmitting their frames.

The CSMA/CA primarily comprises three parameters, such as the Number of Backoffs (NB), Backoff Exponent (BE) and Contention Window (CW). NB is about the number of tries to access the medium for transmitting a frame. Its initial value is 0 and ranges up to the value as defined in parameter *MaxCSMABackoffs*. The default value of *MaxCSMABackoffs* is 4, which allows a node to make four attempts to access the medium availability before transmitting the frame. If it cannot access the medium then it declares the failed transmission with medium access busy

TABLE I. COMPARATIVE SUMMARY OF REFERENCED RESEARCH

Ref. No.	Addressed Area	Proposed Scheme	Results
[15]	Anomaly detection technique for IoT networks	Reinforcement learning on imbalanced data set with normal and malicious data classification	Better accuracy, recall, and F1 score
[16]	Cyber-attack detection mechanism in Industrial IoT	Federated learning-based approach to local anomaly detection centers	Better accuracy and throughput as compared to the related techniques
[17]	Traffic flow monitoring applications	Deep Q-learning technique for anomaly detection for DoS attack	Performs better during DOS attacks than other referenced techniques
[18]	Traffic density along with emergency traffic conditions	Anomaly detection algorithm based on isolation forests by sending probe messages	Improvement in terms of accuracy, recall, and F1 score
[19]	Incorporates social networks in feature learning	GNN technique for feature learning	Improve precision, recall, and F1 score
[20]	Security concerns of the Domain Name System	Exponential random graph model and time series analysis	Improved precision in security of the Domain Name System
[21]	Focused on social welfare behaviour in IoT-based networks	Vector space-based aggregation with spatial index tree	Quick and accurate detection of abnormal behaviour in the network
[22]	Energy efficient Anomaly detection mechanism in three-tier IoT-edge-cloud networks	Marching square algorithm on data collected by the edge nodes to generate isopleths	Better accuracy and energy consumption as compared to other schemes
[23]	Emphasized on real-time data accuracy in Industrial IoT applications	Deep anomaly section framework based on CNN and a two-stage LSTM	Improves efficiency of the resource-constrained edge devices

notification. BE determines the number of backoff periods, a node has to wait before accessing the channel and is calculated as $2^{BE} - 1$. The initial default value of BE is 3 and the number of random backoff periods, a node has to wait initially is in the range of 0 – 7. If it cannot find the medium idle, then the algorithm increments the BE value and the waiting range before accessing the medium increases to 0 – 15. Parameter CW allows a node to check the medium availability twice before transmitting the frame.

If the transmitted frame cannot reach its destination due to collision with another frame in the medium then it is re-transmitted. If several re-transmissions reach the parameter limit defined in $macMaxFrameRetries$ parameter, then the transmission is considered unsuccessful.

B. Attacks on IEEE 802.15.4 Standard

Malicious nodes interfere with the medium to disturb the communication of legitimate nodes. This work focuses on the malicious attacks during CAP of IEEE 802.15.4 standard. The following three types of malicious node attacks are quite common in the MAC protocols to disturb the communication standards of the protocol:

- 1) Exhaustion Attack
- 2) Collision Attack
- 3) Unfairness Attack

1) *Exhaustion attack*: During the CAP, nodes utilize CSMA/CA before transmitting a frame into the medium. They assess the medium's availability by conducting a Clear Channel Assessment (CCA). Malicious nodes keep the medium occupied by transmitting a long stream of messages. This results legitimate node finding the medium busy even after multiple tries as mentioned in $MaxCSMABackoffs$ parameters and the required message initiated by the upper layer is exhausted.

When a node transmits its packet and cannot receive its acknowledgment, then it has to resend the packet again and again till its maximum limit and then finally declares that the packet can not be transmitted.

2) *Collision attack*: Collision occurs when two or more nodes transmit their packets in the medium at the same time and cause the collision. Nodes wait for the acknowledgment for a certain time as mentioned in the parameter $macAckWaitDuration$ of the standard. If transmitting nodes do not receive the acknowledgment within the specific time, then it re-transmits the frame and if the number of retries reaches the limit mentioned in $macMaxFrameRetries$, then the transmission is declared unsuccessful. Malicious nodes disturb the communication after intentionally transmitting a short message while detecting the medium busy causing collisions of the frames transmitted by legitimate nodes.

3) *Unfairness attack*: The standard offers fairness by allowing all nodes equal chances to assess the medium after the decrement of the backoff period. A node after completing its backoff period can access the medium in transmitting its frame. Similarly, the standard allocates GTS to nodes, on a First Come First Serve (FCFS) basis. In case, the PAN coordinator receives GTS requests more than its available limit of 7, then it assigns GTS to those nodes, whose requests arrive first. Malicious nodes do not wait for their assigned backoff periods and initiate their requests at once which reduces the fair chances of other nodes to access the medium. Similarly, it occupies the GTS by initiating early GTS requests to the PAN coordinator and GTS requests of legitimate nodes of the networks are not entertained.

These malicious node attacks create an anomaly in the IoT network applications and QoS is compromised. In this work, an Anomaly Detection Mechanism for IEEE 802.15.4 standard ($ADM_{15.4}$) in an IoT network is proposed. $ADM_{15.4}$ detects malicious attacks in the network and then proposes a comprehensive mechanism to improve the QoS of the network by avoiding malicious attacks.

IV. PROPOSED SCHEME

In this work, malicious nodes' presence is identified by proposing an anomaly detection mechanism during CAP of IEEE 802.15.4 standard. The proposed $ADM_{15.4}$ detects anomalies in the network by introducing an anomaly detection

method and then proposes a solution to neutralize its effect to improve the QoS of the IoT network. The main features of our proposed scheme and described below and its flow is mentioned in Fig. 2.

- Anomaly detection mechanism to detect the anomaly in the medium through a soft function.
- Once an anomaly is detected in a medium, the communication of the region is restricted to prevent a malicious attack by transmitting a jamming signal.
- Data transmission of the affected legitimate nodes available in the restricted region along with an efficient GTS allocating method to improve the QoS.

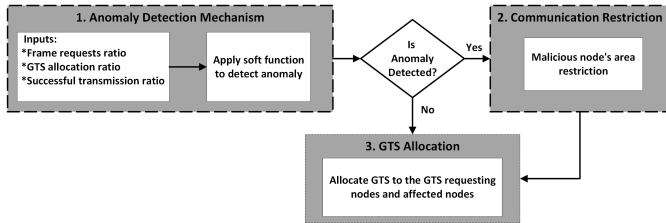


Fig. 2. A Flow of different sections of the proposed scheme.

A. Anomaly Detection Mechanism

Physical and MAC layers of most of the IoT-based networks follow IEEE 802.15.4 standard. MAC layer attacks of malicious nodes compromise the efficiency of the network. In the IEEE 802.15.4 standard, most of the attacks are during its CAP and disturb its performance. The proposed method, based on [24], is used to detect anomalies in the network using various parameters at the end of each SD. The method involves several steps, as shown in Fig. 3.

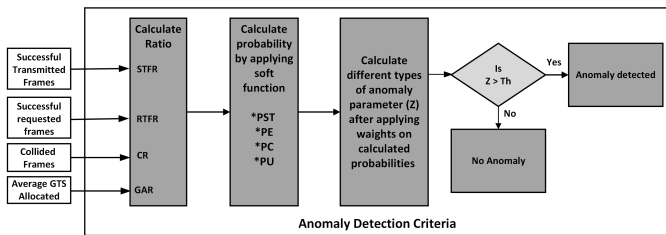


Fig. 3. Flow of the proposed anomaly detection mechanism.

1) *Transmission and collision ratio*: PAN coordinator at the end of each SD computes the following ratios of the different parameters that are observed during the SD.

a) *Successful Transmitted Frames (STF) ratio*: STF is calculated during each SD of the standard by calculating the number of successfully transmitted packets against the total number of requests.

b) *Requested Frame (RF) ratio*: RF is calculated as the number of frames successfully transmitted in the medium to the total number of the frames, nodes intend to transmit during an SD.

c) *Collision Ratio (CR)*: CR is computed by dividing the total number of collisions detected by the total number of frames transmitted in the medium during SD. This metric provides insight into the efficiency of the network by quantifying the proportion of frames that experienced collisions during the specified period.

d) *GTS Allocation Ratio (GAR)*: GAR is the maximum value among all the GTS requested nodes that is calculated as the average number of GTS allocated to a node against a total number of GTS requests received.

2) *Implementation of soft function*: After calculating all the ratios during the SD, a soft function (ψ) is formulated to determine the probabilities of various events based on input values. Specifically, it calculates the probability of successful transmission (PST), the probability of exhaustion attacks (PE), the probability of collision attacks (PC), and the probability of unfairness attacks (PU) using the input values of STF , RF , CR , and GAR , respectively. The mathematical expression is as follows:

$$\psi(X) = \frac{1}{1 + e^{-E \times (V - F)}} \quad (3)$$

Here, $\psi(X)$ is in the range between 0 and 1 and its outcome is the PST , PE , PC , and PU , while replacing V with inputs of STF , RF , CR , and GAR respectively in the soft function. The value of V can be determined through the desired value (Y_D) and real values (Y_R) as mentioned in Eq. 4.

$$J(V) = (Y_D - Y_R)^2 \quad (4)$$

However, E represents slope and F represents the centre of the curve. The shape of the curve is contingent upon these two values, and their dynamics evolve, recalculated after each SD as:

$$E_{K+1} = E_K + (\phi \times \frac{\partial J}{\partial E}) \quad (5)$$

Here ϕ ranges between 0 and 1 and $\frac{\partial J}{\partial E}$ are calculated as:

$$\frac{\partial J}{\partial E} = 2(Y_D - Y_R) \times \frac{E_K}{[1 + e^{-E_K \times (V - F_K)}]^2} \quad (6)$$

Similarly F_{K+1} is calculated as:

$$F_{K+1} = F_K + (\phi \times \frac{\partial J}{\partial F}) \quad (7)$$

Here $\frac{\partial J}{\partial F}$ is calculated as:

$$\frac{\partial J}{\partial F} = 2(Y_D - Y_R) \times \frac{-E_K \times e^{-E_K \times (V - F_K)}}{[1 + e^{-E_K \times (V - F_K)}]^2} \quad (8)$$

Algorithm 1: Anomaly Detection Algorithm

Input: Successful Transmission Ratio STR , Requested Frame Ratio RF , Collision Ratio CR , GTS Allocation Ratio GAR ,

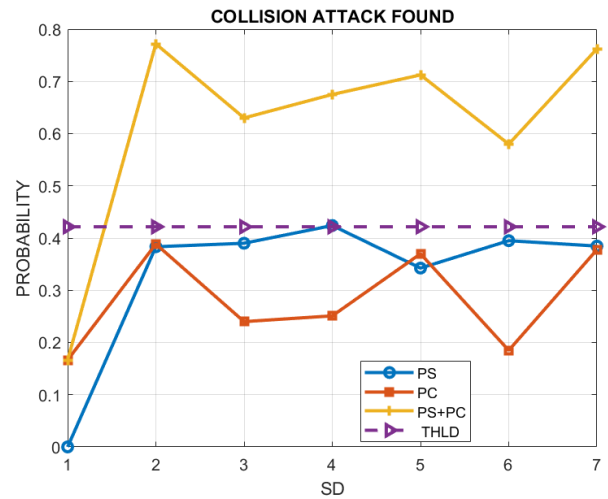
- 1 Compute $PST = 1/1 + \exp(-E \times (STR - F))$
- 2 Compute $PE = 1/1 + \exp(-E \times (RF - F))$
- 3 Compute $PC = 1/1 + \exp(-E \times (CR - F))$
- 4 Compute $PU = 1/1 + \exp(-E \times (GAR - F))$
- 5 Compute $Z_1 = (PST \times \psi) + (PE \times \theta)$
- 6 Compute $Z_2 = (PST \times \psi) + (PC \times \theta)$
- 7 Compute $Z_3 = (PST \times \psi) + (PU \times \theta)$
- 8 **if** $Z_1 > Th$
- 9 Exhaustion Attack
- 10 **else**
- 11 No Exhaustion Attack
- 12 **if** $Z_2 > Th$
- 13 Collision Attack
- 14 **else**
- 15 No Collision Attack
- 16 **if** $Z_3 > Th$
- 17 Unfairness Attack
- 18 **else**
- 19 No Unfairness Attack

3) *Anomaly detection with results:* After determining the PST , PE , PC , and PU , the PAN coordinator assigns weights that are within the range of 0 and 1 to each of the calculated probabilities. Each of the exhaustion, collision, and unfairness probability in combination with the weighted successful transmission probability compute the anomaly value. The calculated anomaly value is compared with the threshold value calculated in [25] to find the anomaly. The proposed anomaly detection algorithm is shown in Algorithm 1.

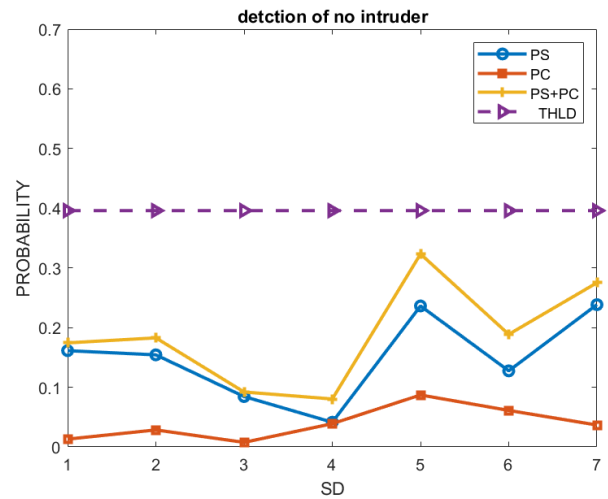
Results in Fig. 4 show the anomaly detection by the proposed algorithm to find the collision. Sub-figure 4b shows when there are no collision attacks found in the network as they are below the threshold level. However, sub-figure 4a shows the collision detection as the collision found in the network is more than the threshold limit as calculated in [25].

Results in Fig. 5 represent the presence of exhaustion attacks in the network and it is comprised of two sub-plots. Subplot 5b exhibits when there are no exhaustion attacks found as the exhaustion value represented by Z_1 in the algorithm is less than the threshold value. However, exhaustion attacks are found as shown in subplot 5a, when the exhaustion value is greater than the threshold value.

Results in Fig. 6 represent the unfairness attacks. Unfairness attacks are calculated from the GTS allocation in the standard as described in Section III-B and are determined from exhaustion value Z_3 from the algorithm. The figure comprises two subplots. Subplot 6a shows when the exhaustion value is greater than the threshold value due to the exhaustion attacks, however subplot 6b represents when the exhaustion value is less than the threshold limit resulting in no unfairness attacks found in the medium.



(a) Collision attacks.



(b) Without collision attack.

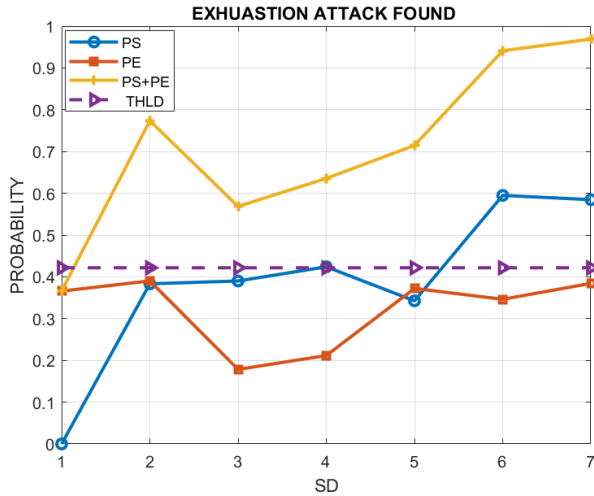
Fig. 4. With and without collision attack.

B. Prevention of Malicious Attacks

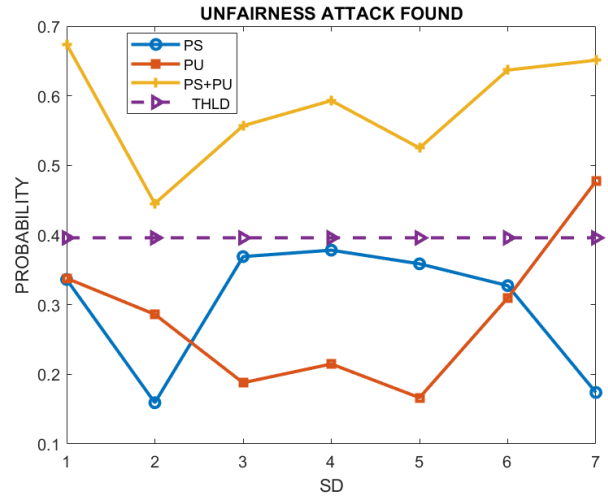
After successfully detecting the presence of a malicious node, its attacks are required to be neutralized by blocking its communication. In a terrestrial IoT network architecture, the PAN coordinator is supposed to know the location of each IoT node placed in the network with the help of its short address provided by the PAN coordinator of IEEE 802.15.4 standard. To stop the communication of the malicious nodes in the network, the PAN coordinator in its beacon frame requests one of the neighbouring malicious nodes, which has the highest residual energy level, to transmit jamming signals during CAP. Generating a jamming signal restricts the communication of all the nodes present in that area resulting in compromised QoS in that area as mentioned in Fig. 7.

C. Communication of the Affected Nodes

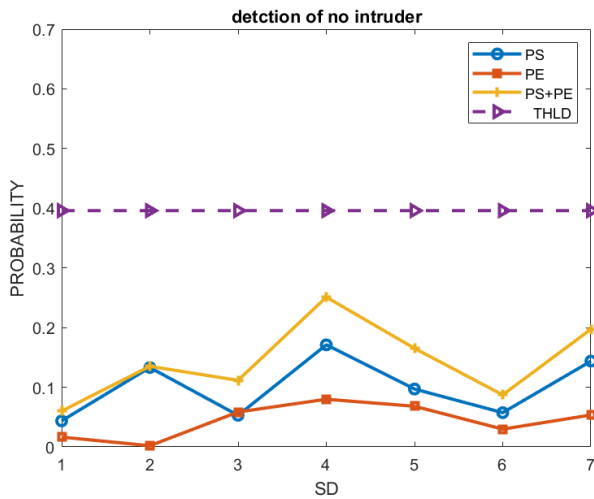
The communication of the legitimate nodes present in the restricted areas is provided by allocating GTS in the upcoming



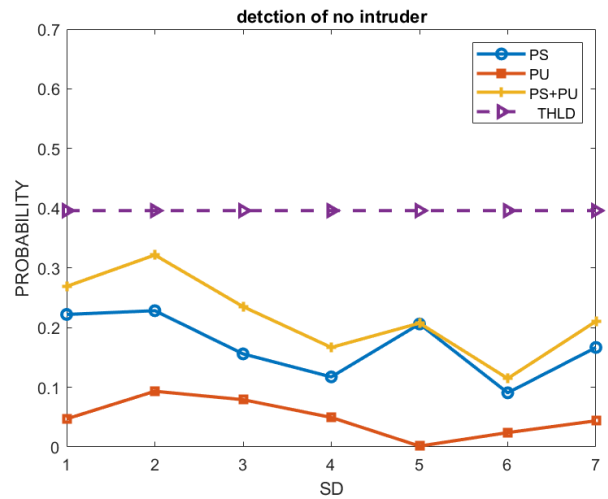
(a) Exhaustion attacks.



(a) Unfairness attacks.



(b) Without exhaustion attack.



(b) Without unfairness attack.

Fig. 5. With and without exhaustion attack.

Fig. 6. With and without unfairness attack.

SD. Due to restricted CAP, these affected nodes are unable to transmit their GTS requests to the PAN coordinator, In such a scenario, the GTS are assigned to these affected nodes by analyzing the nodes' previous transmission pattern. Suppose PAN coordinator receives j number of requests during past k sessions, then its expected GTS requests (GTS_i) is calculated as:

$$GTS_i = (K_{last} - K_{cur}) + \left\lceil \frac{K}{J} \right\rceil \quad (9)$$

Here, K_{last} is represented as the last SD when node i initiated the request and K_{cur} is the upcoming SD.

Each SD of IEEE 802.15.4 standard consists of 7 TDMA-like CFP slots. PAN coordinator after determining the expected GTS allocation to the affected nodes, assigns the remaining slots against the GTS requests received from the unaffected legitimate nodes. Suppose the PAN coordinator has m CFP

slots available and the number of GTS required to be allocated to affected nodes is n , then the PAN coordinator can only accommodate $m - n$ GTS requested slots in the next SD. If the number of GTS requested by the unaffected legitimate nodes is less than $m - n$ slots, then it can entertain all the GTS requests. However, in case, the number of requested GTS exceeds the available slots, a scrutiny of GTS takes place based on their priority levels. This is accomplished by employing the 0/1 knapsack algorithm.

To determine the priority of a node, the default GTS requesting command frame format has been modified by utilizing its two reserved bits. Each node requesting GTS informs its PAN coordinator about the number of GTS required, along with its priority level. This information is conveyed in the last two reserved bits of the GTS characteristic field, which is part of the GTS request command frame format specified in the IEEE 802.15.4 standard, as illustrated in the accompanying Fig. 8. Priority levels of each IoT node are categorized into

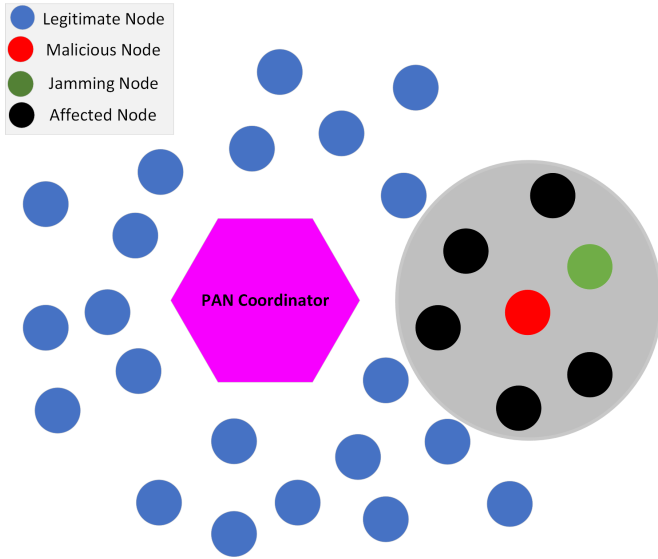


Fig. 7. IoT Network with malicious and affected nodes.

four different levels ranging from 00 to 11 representing the lowest to the highest priority levels, respectively.

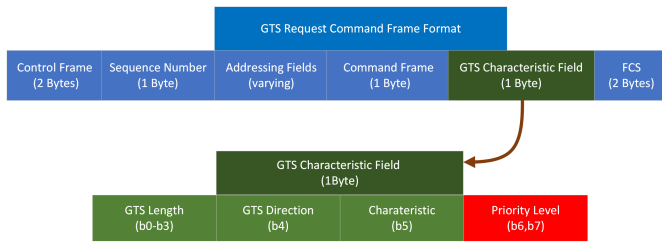


Fig. 8. Modified GTS request command frame format of IEEE 802.15.4 standard.

The PAN coordinator scrutinizes the GTS requests by applying the 0/1 knapsack algorithm. The available CFP slots in the upcoming SD are considered as sack capacity. Each GTS requesting node is mapped with an item and its requested slots are mapped as the weight of the item. The value of the item (V_i) is mapped with the value of the GTS requesting node i and depends upon its priority (P_i) and the time (T_i) that it has to wait after initiating its GTS request as:

$$Val_i = P_i \times T_i \quad (10)$$

Priority of the requested GTS as calculated from the proposed GTS requesting command frame format as shown in Fig. 7. However, the waiting time is calculated as:

$$T_i = N_i + \frac{(960 \times 2^{BO}) - X_i}{960 \times 2^{BO}} \quad (11)$$

Here, N_i represents the consecutive number of requests, the PAN coordinator receives from node i . If there is no GTS request in the previous beacon interval (BI), then its value is 0, however, if the same node is requesting GTS for the last

two BI and its request is not entertained, then the value of N is 2. X_i is the duration in symbols and it is calculated as the time between the start of the beacon frame and the time when the PAN coordinator receives the GTS request.

The proposed $ADM_{15.4}$ assists the PAN coordinator in allocating the available GTS to the GTS requesting nodes as:

- 1) Assign GTS to all GTS requesting nodes if requesting slots are less than the available GTS in the upcoming SD.
- 2) Scrutinize the GTS requesting node in allocating the GTS if the number of requesting slots is more than the available GTS.

A complete algorithm for GTS allocating nodes in upcoming SD for PAN coordinator is shown in Algorithm 2.

V. SYSTEM MODEL

Wireless sensor-based IoT nodes are being used in diverse wireless applications. Most of the wireless sensor networks use IEEE 802.15.4 standard in their MAC and Physical layers. Malicious wireless nodes being a part of the network, try to disturb the communication of the legitimate nodes and create an anomaly. In this work, the superframe structure of IEEE 802.15.4 standard operating at a 2.4GHz frequency channel is used for communication between all wireless connected nodes creating a Wireless Personal Area Network (WPAN). WPAN comprises a PAN coordinator and its member nodes. A system model of this work comprises of WPAN coordinator with legitimate member nodes and few malicious nodes as shown in Fig. 9. A WPAN coordinator acting as Cluster Head (CH) is selected based on the higher residual energy level among all nodes. All other nodes in the WPAN act as member nodes. All member nodes are in direct connection with the CH and do not use any relaying node to reach CH. Malicious nodes are part of the network and disturb the medium access of all legitimate nodes by transmitting information during CAP of the standard. This causes legitimate nodes to find the medium busy as well as increases the chances of collision in the medium with increased unfairness of the legitimate nodes.

Nodes can transmit their data during CAP as well as during CFP. A data frame transmitted during a CAP is successfully delivered, if it receives its acknowledgment within a stipulated time as mentioned in different parameters in IEEE 802.15.4 standard. Total time (ζ) required in transmitting a requesting frame during CAP is calculated as sum of backoff count (BC), data transmitting duration (TD), Propagation delay (PD), turn around (TA) time, Acknowledgment frame time (AF), and Inter-frame space (IFS) as mentioned in Eq. 12 and is shown in Fig. 10.

$$\zeta = BC + TD + (2 \times PD) + TA + AF + IFS \quad (12)$$

If X number of legitimate nodes are successful in transmitting its frames during CAP of a SD , then accumulated time (σ_{CAP}) calculated in successful transmission of data requested frames during CAP in q number of SD is calculated as:

Algorithm 2: GTS Allocation Mechanism

```

1   $w \leftarrow$  Current slot number
2   $W \leftarrow$  Max. number of available GTS
3   $a \leftarrow$  Node ID
4   $v \leftarrow$  Maximum Number of GTS requesting nodes
5   $k \leftarrow$  Maximum Number of GTS requested
6   $A[a, w] \leftarrow$  Cell value of  $a^{th}$  node and  $w^{th}$  slot
7   $w_a \leftarrow$  Slots requested by  $a^{th}$  node
8  if  $K < W$  then
9  | Allocate GTS to all requesting nodes
10 end
11 else
12 | Scrutinize nodes by applying 0/1 knapsack
13 | Populating the 0/1 knapsack table:
14 | for  $w = 0$  to  $W$  do
15 | |  $A[0, w] = 0$ 
16 | end
17 | for  $a = 1$  to  $v$  do
18 | |  $A[a, 0] = 0$ 
19 | end
20 | for  $a = 1$  to  $v$  do
21 | | for  $w = 0$  to  $W$  do
22 | | | if  $w_a \leq w$  then
23 | | | | if  $w_a + A[a - 1, w - w_a] > A[a - 1, w]$ 
24 | | | | | then
25 | | | | | |  $A[a, w] = w_a + A[a - 1, w - w_a]$ 
26 | | | | | end
27 | | | | | else
28 | | | | | |  $A[a, w] = A[a - 1, w]$ 
29 | | | | | end
30 | | | | | else
31 | | | | | |  $A[a, w] = A[a - 1, w]$ 
32 | | | | | end
33 | | | end
34 | | end
35 | | Nodes selection Criteria:
36 | | while  $a > 1$  and  $w > 1$  do
37 | | | if  $A[a, w] > A[a - 1, w]$  then
38 | | | |  $a^{th}$  node is selected
39 | | | |  $a = a - 1$ 
40 | | | |  $w = w - w_a$ 
41 | | | end
42 | | | else
43 | | | |  $a = a - 1$ 
44 | | | end
45 | | end
46 end

```

$$\sigma_{CAP} = \sum_{a=1}^q \sum_{b=1}^X SD_a \times \zeta_b \quad (13)$$

The time required in transmitting data during CFP is calculated as the time when a node, initiates its request during CAP since it transfers its data in CFP slots. Number of GTS required (CFP_i) to send m amount of data by a node i in transferring its data is calculated as:

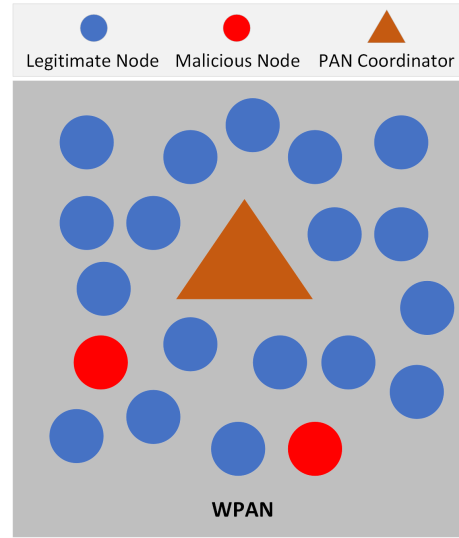


Fig. 9. System model of proposed scheme.

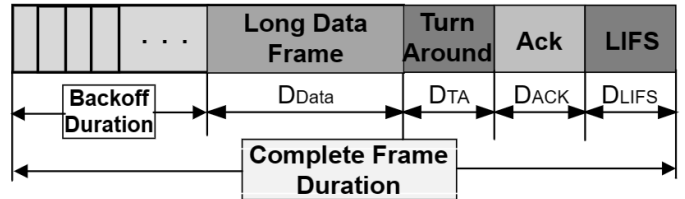


Fig. 10. A Complete frame length including acknowledgment.

$$CFP_i = \left\lceil \frac{m}{30 \times 2^{SO}} \right\rceil$$

Suppose node i sends a request of k number of GTS to the WPAN coordinator during CAP. If the WPAN coordinator successfully allocates its required GTS just before the j slots of the CFP period in the next SD, then the complete delay (CD_i) in transferring its data in its allocated GTS is calculated as:

$$CD_i = BI + SD - j + \left(\frac{m}{30 \times 2^{SO}} \right) \quad (14)$$

If p nodes are allocated GTS in each SD , then accumulated delay (σ_{GTS}) of all the WPAN in transferring data during CFP for q number of SD is calculated as:

$$\sigma_{GTS} = \sum_{a=1}^q \sum_{b=1}^p SD_a \times CD_b \quad (15)$$

VI. RESULTS AND ANALYSIS

In this section, the performance of the proposed scheme is thoroughly examined across various dimensions. The analysis delves into different aspects, evaluating the efficacy of the proposed scheme within the system model outlined in the preceding Section V. A simulation environment is established by deploying a fixed number of legitimate nodes within a

WPAN. This network configuration includes one WPAN coordinator alongside legitimate nodes, and notably, one additional node designated as a malicious node. All nodes follow the IEEE 802.15.4 standard and communicate with each other on the same frequency channel of the 2.4 GHz frequency band. Malicious nodes are present in specific areas and their position is supposed to be identified by the PAN coordinator. A “10 X 10” meters area around the malicious node is blocked by transmitting Jamming signals during CAP by one of the legitimate nodes in that area. A random number of nodes during each SD generate their GTS requests to transmit their data during CFP. The results are analyzed for a fixed duty cycle of 50% along with varying duty cycles for different values of *SO* and different numbers of nodes. A list of simulation parameters is presented in the Table II.

TABLE II. SIMULATION PARAMETERS

Parameters	Values
Total Number of Member Nodes	19, 8, 12
Network Size	100 × 100
Data Rate	250Kbps
Number of Legitimate Nodes	19
Number of Malicious Node	1
Cluster Head	1
Superframe Order	0,2
Superframe Duration (msec)	15.4,61.4
Beacon Interval (msec)	30.7, 122.9
Slot Duration (msec)	1.92, 7.68
Duty Cycle	50%
Offered Load (Bytes)	50 to 125

The simulation results are observed in different prospects with and without attacks and the performance of our proposed $ADM_{15.4}$ scheme is evaluated. The performance is compared with the standard by data transmission, average transmission time, and number of GTS allocated nodes accommodated by the PAN coordinator.

A. Transmitted Data

The data transmitted is calculated for only those legitimate nodes that are allowed to transfer their data during CFP. The proposed scheme applies 0/1 knapsack algorithm in allocating GTS to the legitimate nodes. However, the IEEE 802.15.4 standard applies FCFS in allocating GTS to the requesting nodes.

Results shown in Fig. 11 represent the effect on data transmission with and without attacks. The results show that the data transmission for the same number of data-requesting nodes increases at the same rate when there is no malicious attack. On the other hand, the data transmission is affected due to malicious attacks in the second SD. However, in the proposed scheme, the data transmission is affected in the second SD, however after the blocking of the region at the start of the 3rd SD, the rest of the nodes keep on transmitting their data. It can be observed from the results that between 1 and 2 SD values, there is no malicious attack and all nodes are transmitting data with same rate. However, in the 2nd SD, malicious nodes attack the medium and attacks are detected by the proposed scheme at the end of the 2nd SD and a prevention mechanism is applied in the 3rd SD by transmitting jamming signals in the surrounding of the malicious node. This affects the communication of nodes in the specific area, however,

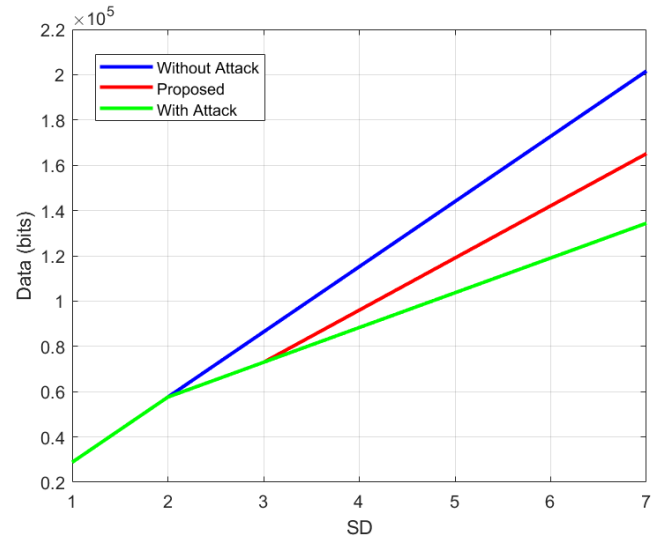


Fig. 11. Data transmission of nodes with and without attacks.

nodes present in the rest of the area remain unaffected and keep on transmitting their data.

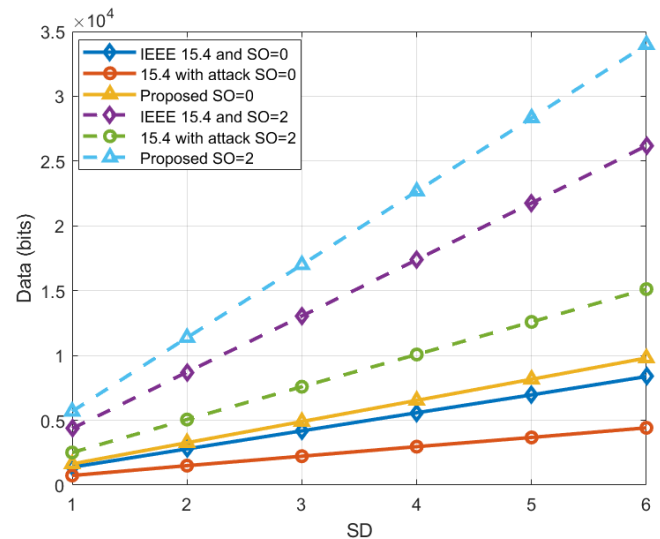


Fig. 12. Data transmission of GTS requesting nodes.

When attacks were found then $ADM_{15.4}$ allocates GTS to the nodes affected in that area as described in Section IV-B. Results shown in Fig. 12 represent the total amount of data transmitted by all nodes during CFP duration in the network when $SO=0$ and $SO=2$. The performance of the proposed scheme is evaluated by comparing its results with both the IEEE 802.15.4 standard under attack scenarios and the IEEE 802.15.4 standard in the absence of attacks. The results show that, for both values of SO , the data transmission in the proposed scheme is 30% more than the standard without attacks and 122% more than the standard with attack. This is due to the efficient allocation of GTS among GTS requesting nodes by applying the 0/1 knapsack algorithm because it allows the PAN coordinator to optimally allocate GTS among

the requesting nodes.

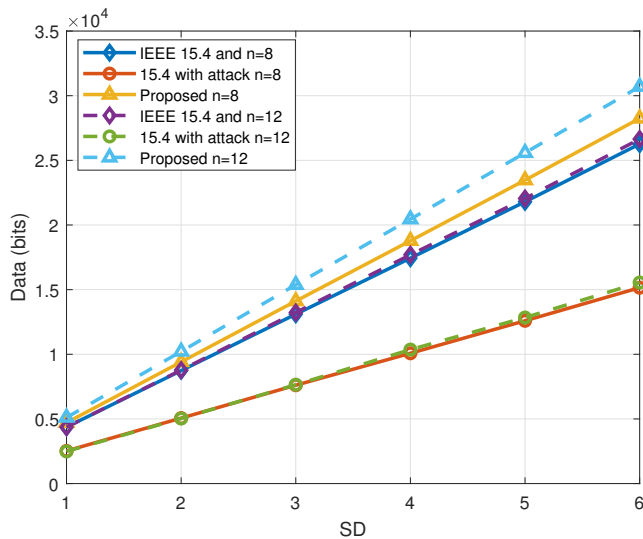


Fig. 13. Data transmission of GTS requesting nodes for different number of nodes.

The performance of the proposed scheme is validated by calculating the transmitted data during CFP when there was a random number of GTS requests from 8 and 12 legitimate nodes with an SO value of 2 as shown in Fig. 13. The results showed that the proposed scheme allowed for more data transmission for both 8 and 12 requesting nodes compared to the standard, with and without attacks. This demonstrates an optimal allocation of GTS among requesting nodes to enable more data transmission in an SD. Moreover, the results highlighted that the data transmission of the standard was severely affected during attacks because the PAN coordinator was unable to differentiate between legitimate and malicious node requests. This led to the PAN coordinator assigning GTS to the malicious nodes at the start of the CAP by applying FCFS.

B. Transmission Delay

The delay in transmitting data is calculated for those nodes that have initiated the GTS requests. The time to transmit data for all those nodes, which are successfully allocated GTS are calculated by following the Eq. 14. However, the transmission time of all those nodes which are not allocated GTS are supposed to be assigned GTS in the next SD automatically by passing through another BI.

Results in Fig. 14 show the accumulated time calculated for all GTS requesting nodes in transmitting their data. It is evident from the results, that due to malicious attacks, the overall data transmission time of GTS requesting nodes increases due to less number of legitimate nodes being assigned GTS in a superframe duration and the rest are allowed to send their data in the next superframe duration with an increase in BI time interval. However, data transmission time in proposed $ADM_{15.4}$ is the least among all and even less than the IEEE 802.15.4 standard because it accommodates a maximum number of GTS requesting nodes in transmitting their data during CFP in the current SD and less number of GTS

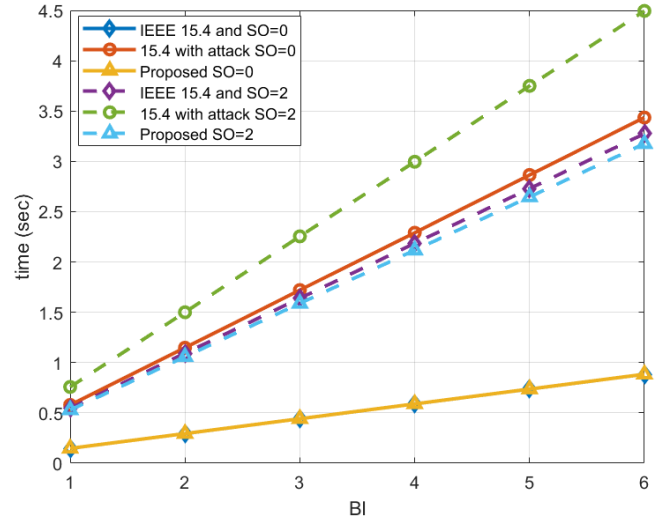


Fig. 14. Delay in transmitting data during contention free period.

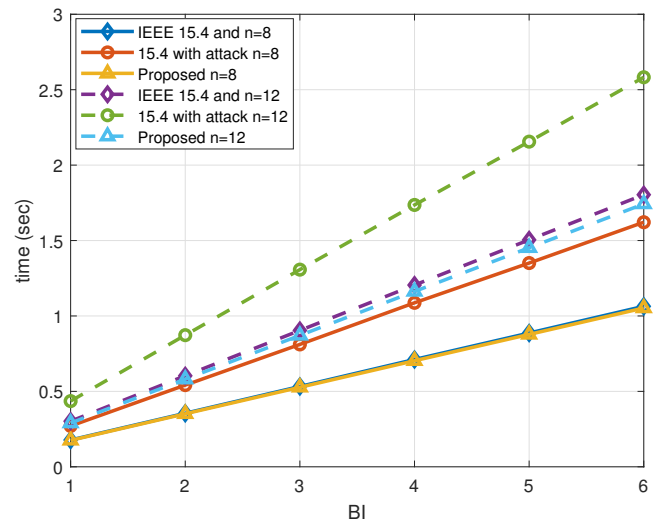


Fig. 15. Delay in transmitting data during contention free period for different number of nodes.

requesting nodes transmit their data in the next SD resulting in a reduced network delay.

Results shown in Fig. 15 represent the network delay of all GTS requesting nodes when the number of GTS requesting nodes are 8 and 12 with a 50% duty cycle. The results clearly show that the accumulated delay of all GTS requesting nodes in transmitting their data in $ADM_{15.4}$ is 0.5% to 3% less than the standard when there is no attack for both 8 and 12 GTS requesting nodes respectively. However, it is 58% and 49% less in the presence of malicious attacks for number of nodes are 8 and 12, respectively because it allocates GTS to the malicious nodes and most of the legitimate nodes are unattended and are not allocated GTS.

Results in Fig. 16 show a comprehensive picture by calculating the difference in delay between the proposed scheme

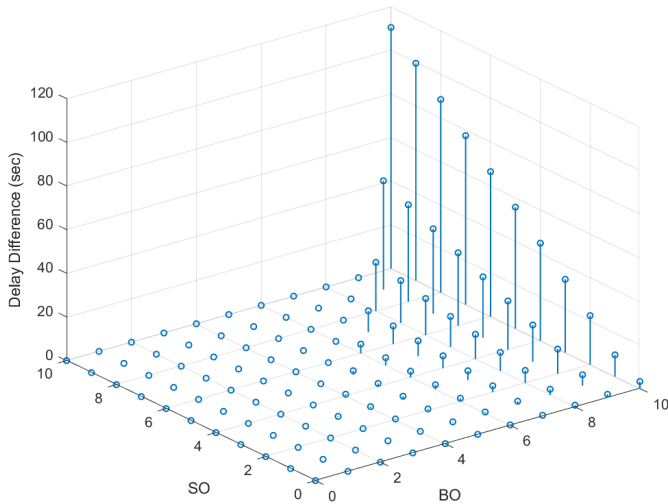


Fig. 16. Accumulated delay difference for all possible values of SO when BO=10.

and IEEE 802.15.4 standard with attacks. The results are obtained by accumulating the total difference in delay faced by 19 legitimate nodes against all the possible values of SO when BO ranges from 0 to 10. The results show that with the increase in BO, the delay difference increases because higher BO allows an increased number of SO options and the accumulated sum of all the differences against all the possible values also increases. Furthermore, increased BO increases the BI, and unsuccessful GTS requesting nodes have to wait for another BI resulting in more delay.

C. GTS Allocating Nodes

GTS allocating nodes are calculated as the total number of GTS requests of legitimate nodes entertained by the PAN coordinator during an SD. The results are obtained for two different values of SO when the number of GTS requesting nodes is 20, and when the number SO is fixed and the number of GTS requesting nodes is 8 and 12 as shown in Fig. 17 and 18, respectively.

Fig. 17 shows the total number of GTS requesting nodes, that have been successfully allocated GTS in a SD by WPAN. It is evident from the results that $ADM_{15.4}$ entertains the maximum number of GTS requesting nodes in a SD and number of GTS entertained for $SO = 2$ are 24% and 110% more than IEEE 802.15.4 standard without attacks and with attacks, respectively. However, when $SO = 0$, then the proposed scheme allocates the same number of GTS requesting nodes as nodes entertained by IEEE 802.15.4 standard without attacks. This is due to the reason that the optimal number of GTS requesting nodes is also entertained by the PAN coordinator in IEEE 802.15.4 standard. However, due to unfairness attacks, some CFP slots are allocated to malicious nodes, resulting in less number of CFP slots left that are allocated to legitimate nodes.

Results in Fig. 18 show that the number of nodes entertained throughout the different superframe durations in the

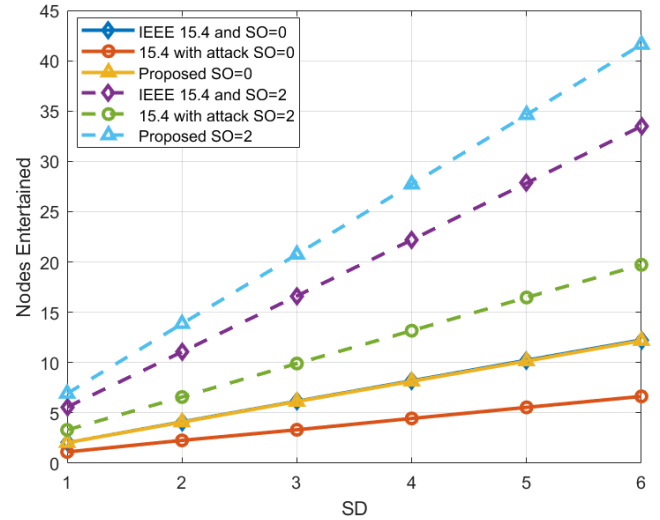


Fig. 17. Number of GTS requesting nodes entertained.

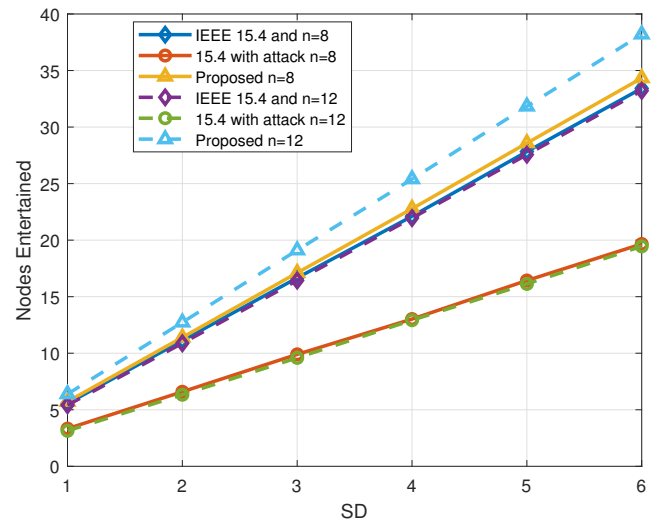


Fig. 18. Number of GTS requesting nodes entertained for different number of nodes.

proposed scheme is the highest for both numbers of GTS requesting nodes. It is evident from the results that the accumulated number of GTS requests entertained by the PAN coordinator is maximum when GTS requesting nodes are 12 in the proposed scheme. This is due to the optimal allocation of GTS to the GTS requesting nodes by applying the 0/1 knapsack algorithm as compared to FCFS used in the IEEE 802.15.4 standard. The results further show that the least number of GTS requests of the legitimate nodes are entertained in the presence of the malicious attacks because the standard does not differentiate the malicious attacks and some of the GTS are allocated to malicious nodes resulting in less number of GTS left for allocation to legitimate nodes.

VII. CONCLUSION

This work addresses the compromised QoS due to anomaly created by malicious nodes in the communication medium of IEEE 802.15.4 standard. In this work, an Anomaly Detection Mechanism for IEEE 802.15.4 standard $ADM_{15.4}$ is proposed. The proposed scheme detects the different types of anomaly caused by malicious node attacks during the contention access period of the superframe structure of the standard. Furthermore, $ADM_{15.4}$ proposes a PLC-based mechanism to stop the interference caused by a malicious node by transmitting jamming signals to its nearby node. This causes an interruption in a specific region and nodes in that region are unable to communicate during the contention access period. To overcome their communication interruption, these nodes are allocated GTS to transmit their information to WPAN applying a 0/1 knapsack algorithm in such a way that maximum GTS requesting nodes are entertained. The simulation results show that the proposed scheme improves the data transmission of legitimate nodes by 122% and 30% as compared to the standard with and without attacks respectively. The transmission delay of legitimate GTS requesting nodes is also reduced by 58% and 3% as compared the standard with and without attacks and accommodates up to 24% and 110% more GTS requesting nodes to transmit their data during CFP period in the current superframe duration. The improved data transmission and reduced transmission delay makes the proposed scheme suitable for future IoT applications. In the future, we will explore methods to detect anomalies due to data integrity attacks and faulty IoT sensors.

ACKNOWLEDGMENT

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (IMSIU) (Grant number IMSIU-RP23082).

REFERENCES

- [1] K. Fizza, P. P. Jayaraman, A. Banerjee, N. Auluck, and R. Ranjan, "Iot-qwatch: A novel framework to support the development of quality-aware autonomic iot applications," *IEEE Internet of Things Journal*, vol. 10, no. 20, pp. 17 666–17 679, 2023.
- [2] A. A. Abdulameer and R. K. Oubida, "The impact of iot on real-world future decisions," in *AIP Conference Proceedings*, vol. 2591, no. 1. AIP Publishing, 2023.
- [3] A. Gupta, T. Gulati, and A. K. Bindal, "Wsn based iot applications: A review," in *2022 10th International Conference on Emerging Trends in Engineering and Technology - Signal and Information Processing (ICETET-SIP-22)*, 2022, pp. 1–6.
- [4] B. Yao, R. Tang, and S. Ma, "Consideration in wsn applying for the health monitoring of transport aircraft," in *2023 9th International Symposium on System Security, Safety, and Reliability (ISSSR)*, 2023, pp. 44–48.
- [5] X. Li, B. Hou, R. Zhang, and Y. Liu, "A review of rgb image-based internet of things in smart agriculture," *IEEE Sensors Journal*, vol. 23, no. 20, pp. 24 107–24 122, 2023.
- [6] B. H. S. Alamri, M. M. Monowar, and S. Alshehri, "Privacy-preserving trust-aware group-based framework in mobile crowdsensing," *IEEE Access*, vol. 10, pp. 134 770–134 784, 2022.
- [7] S. You, K. Radivojevic, J. Nabrzyski, and P. Brenner, "Trust in the context of blockchain applications," in *2022 Fourth International Conference on Blockchain Computing and Applications (BCCA)*, 2022, pp. 111–118.
- [8] D. Popovic, H. K. Gedawy, and K. A. Harras, "Fedteams: Towards trust-based and resource-aware federated learning," in *2022 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, 2022, pp. 121–128.
- [9] D.-Y. Kim, N. Alodadi, Z. Chen, K. P. Joshi, A. Crainiceanu, and D. Needham, "Mats: A multi-aspect and adaptive trust-based situation-aware access control framework for federated data-as-a-service systems," in *2022 IEEE International Conference on Services Computing (SCC)*, 2022, pp. 54–64.
- [10] J. Guo, A. Liu, K. Ota, M. Dong, X. Deng, and N. N. Xiong, "Ictn: An intelligent trust collaboration network system in iot," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 1, pp. 203–218, 2022.
- [11] X. Shen, W. Lv, J. Qiu, A. Kaur, F. Xiao, and F. Xia, "Trust-aware detection of malicious users in dating social networks," *IEEE Transactions on Computational Social Systems*, pp. 1–12, 2022.
- [12] A. N. Alvi, S. Khan, M. A. Javed, K. Konstantin, A. O. Almagrabi, A. K. Bashir, and R. Nawaz, "Ogmad: Optimal gts-allocation mechanism for adaptive data requirements in ieee 802.15.4 based internet of things," *IEEE Access*, vol. 7, pp. 170 629–170 639, 2019.
- [13] S. Khan, A. N. Alvi, M. A. Javed, Y. D. Al-Otaibi, and A. K. Bashir, "An efficient medium access control protocol for rf energy harvesting based iot devices," *Computer Communications*, vol. 171, pp. 28–38, 2021.
- [14] S. Khan, A. N. Alvi, M. A. Javed, and S. H. Bouk, "An enhanced superframe structure of ieee 802.15.4 standard for adaptive data requirement," *Computer Communications*, vol. 169, pp. 59–70, 2021.
- [15] X. Ma and W. Shi, "Aesmote: Adversarial reinforcement learning with smote for anomaly detection," *IEEE Transactions on Network Science and Engineering*, vol. 8, no. 2, pp. 943–956, 2021.
- [16] X. Wang, S. Garg, H. Lin, J. Hu, G. Kaddoum, M. Piran, and M. Shamim Hossain, "Toward accurate anomaly detection in industrial internet of things using hierarchical federated learning," *IEEE Internet of Things Journal*, vol. 9, no. 10, pp. 7110–7119, May 2022.
- [17] T. V. Phan, T. G. Nguyen, N.-N. Dao, T. T. Huong, N. H. Thanh, and T. Bauschert, "Deepguard: Efficient anomaly detection in sdn with fine-grained traffic flow monitoring," *IEEE Transactions on Network and Service Management*, vol. 17, no. 3, pp. 1349–1362, 2020.
- [18] M. A. Javed, M. Z. Khan, U. Zafar, M. F. Siddiqui, R. Badar, B. M. Lee, and F. Ahmad, "Odpv: An efficient protocol to mitigate data integrity attacks in intelligent transport systems," *IEEE Access*, vol. 8, pp. 114 733–114 740, 2020.
- [19] T. Zhao, T. Jiang, N. Shah, and M. Jiang, "A synergistic approach for graph anomaly detection with pattern mining and feature learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 6, pp. 2393–2405, 2022.
- [20] M. Tsikerdekis, S. Waldron, and A. Emanuelson, "Network anomaly detection using exponential random graph models and autoregressive moving average," *IEEE Access*, vol. 9, pp. 134 530–134 542, 2021.
- [21] J. Tang, T. Qin, D. Kong, Z. Zhou, X. Li, Y. Wu, and J. Gu, "Anomaly detection in social-aware iot networks," *IEEE Transactions on Network and Service Management*, vol. 20, no. 3, pp. 3162–3176, 2023.
- [22] Y. Li, Z. Zhou, X. Xue, D. Zhao, and P. C. K. Hung, "Accurate anomaly detection with energy efficiency in iot-edge-cloud collaborative networks," *IEEE Internet of Things Journal*, vol. 10, no. 19, pp. 16 959–16 974, 2023.
- [23] H. Nizam, S. Zafar, Z. Lv, F. Wang, and X. Hu, "Real-time deep anomaly detection framework for multivariate time-series data in industrial iot," *IEEE Sensors Journal*, vol. 22, no. 23, pp. 22 836–22 849, 2022.
- [24] N. M. F. Qureshi, A. Noorwali, A. N. Alvi, M. Z. Khan, M. A. Javed, W. Boulila, and P. A. Pattanaik, "A novel qos-oriented intrusion detection mechanism for iot applications," *Wireless Communications and Mobile Computing*, vol. 2021, p. 9962697, 2021. [Online]. Available: <https://doi.org/10.1155/2021/9962697>
- [25] A. N. Alvi, S. H. Bouk, S. H. Ahmed, and M. A. Yaqub, "Influence of backoff period in slotted csma/ca of ieee 802.15.4," in *Wired/Wireless Internet Communications*, L. Mamatias, I. Matta, P. Papadimitriou, and Y. Koucheryavy, Eds. Cham: Springer International Publishing, 2016, pp. 40–51.

GRACE: Graph-Based Attention for Coherent Explanation in Fake News Detection on Social Media

Orken Mamyrbayev¹, Zhanibek Turysbek^{*2}, Mariam Afzal³, Marassulov Ussen Abdurakhimovich⁴,
Ybytayeva Galiya⁵, Muhammad Abdullah⁶, Riaz Ul Amin^{*7}

Institute of Information and Computational Technologies, Almaty, Kazakhstan¹

Kazakh National Research Technical University, Kazakhstan²

Riphah International University, Faisalabad³

International Kazakh-Turkish University named by Khoja Akhmet Yassawi⁴

Department of Technical and Natural Sciences at the International Educational Corporation⁵

School of Computing and Artificial Intelligence, Zhengzhou University, Zhengzhou, 450001, Henan, China⁶

School of Computing and Information Technology, University of Okara and Edinburgh, Napier University, UK⁷

Abstract—Detecting fake news on social media is a critical challenge due to its rapid dissemination and potential societal impact. This paper addresses the problem in a realistic scenario where the original tweet and the sequence of users who retweeted it, excluding any comment section, are available. We propose a Graph-based Attention for Coherent Explanation (GRACE) to perform binary classification by determining if the original tweet is false and provide interpretable explanations by highlighting suspicious users and key evidential words. GRACE integrates user behaviour, tweet content, and retweet propagation dynamics through Graph Convolutional Networks (GCNs) and a dual co-attention mechanism. Extensive experiments conducted on Twitter15 and Twitter16 datasets demonstrate that GRACE outperforms baseline methods, achieving an accuracy improvement of 2.12% on Twitter15 and 1.83% on Twitter16 compared to GCAN. Additionally, GRACE provides meaningful and coherent explanations, making it an effective and interpretable solution for fake news detection on social platforms.

Keywords—Graph neural network; dual attention; NLP; semantics; social network

I. INTRODUCTION

Social media has become integral to everyday life, allowing individuals to share their thoughts, stay updated on current events, and interact with others [1]. These platforms facilitate the fast flow of information across vast networks, where user interactions and feedback shape public opinions and emotions on various topics [2]. This easy and low-cost communication fosters collective intelligence, spreading ideas widely and quickly. However, the very features that make social media so powerful also have significant drawbacks [3]. The speed and reach of these platforms make it easier for misinformation to spread, often without proper checks or regulation [4]. As a result, while social media can be a tool for empowerment and connection, it also amplifies the risk of misinformation, posing challenges to truth and trust in public discourse.

Fake news consists of false stories that are intentionally shared on social media platforms [5]. Its spread can mislead the public opinion, leading to political, economic, or psychological benefits for specific groups [6]. Fake news circulation

manipulates opinions, distorts facts, and poses risks to society [7]. Research shows that people often struggle to differentiate between true and false news [8]. Interest in detecting fake news surged after the 2016 U.S. presidential election and COVID-19 vaccination drawing attention from researchers and social media platforms [9], [10], [11].

Detecting fake news is a complex task, primarily when relying solely on the content of news articles [12]. Traditional content-based methods often use features like n-grams and bag-of-words, applying supervised learning models such as random forests or SVM for binary classification [4], [13]. More advanced techniques in natural language processing (NLP) focus on extracting linguistic features like active/assertive verbs, subjectivity, and writing style [14]. Multi-modal approaches also integrate user-profiles and retweet propagation patterns [15]. However, these approaches face several challenges. Social media content, such as tweets, is usually short, leading to data sparsity, which makes it harder to detect fake news effectively [16]. Additionally, many models rely on user comments or retweets for evidence, but most users reshare stories without commenting, reducing the available data for analysis [17].

To address these challenges, researchers have begun focusing on propagation-based methods, which analyze the network of tweets and retweets to detect fake news [18], [19]. These methods are based on the idea that fake news spreads differently than true news. By studying the patterns of information diffusion, researchers can identify inconsistencies and spot fake stories [20]. However, many early approaches rely on static networks, assuming that the entire structure of information propagation is known before applying learning algorithms [21]. Social media networks are dynamic, with new users and content emerging over time, making static models less effective.

Recent research has shown that comprising temporary features, such as the timing of user interactions, can significantly improve fake news detection [22], [18]. For instance, in a temporal graph, the news propagation evolves, while a static graph only apprehends a snapshot of the network at one moment.

Fake and real news often show different temporal patterns, with fake news spreading more quickly or following distinct paths [23]. Regardless, treating these dynamic networks as if they were static limits the effectiveness of current models. To enhance detection, it's crucial to develop models that take into account the continuous changes in how users interact with each other. By doing so, these models can offer a more accurate and reliable way to tell the difference between real and fake news. These time-aware models would tap into the ever-evolving nature of social media. It makes them better equipped to detect misinformation in real-world situations.

This paper concentrates on detecting false content in the Twitter social media environment. The goal is to determine if a tweet is fake based solely on its brief text, the sequence of users who retweeted it, and their profiles. The detection process is approached with three key constraints: (a) analyzing the tweet's short text, (b) excluding user comments, and (c) not using network structures like social or diffusion networks. The model is designed to explain its predictions, meaning it should not only flag fake news but also show the reasoning behind the decision. Specifically, the model should highlight the doubtful users who helped to spread the fake news and identify the particular words or phrases from the source tweet that captured their attention.

Graph-based Attention for Coherent Explanation (GRACE) is proposed for fake news detection that integrates user behavior, tweet content, and retweet propagation dynamics. GRACE begins by feature extracting from user's Twitter profiles and encoding the original tweet's text using word embeddings [24]. A user interaction graph is constructed, and Graph Convolutional Networks (GCNs) [25] generate graph-aware representations of propagation dynamics. The relationship between the original tweet and how it spreads through retweets is identified by dual co-attention mechanism. It's helpful to highlight the users and keywords. By combining these features, GRACE offers an effective and easy-to-understand method for classifying fake news.

The key contributions of this paper are outlined as follows:

- 1) GRACE model is introduced to improve the understanding of user connection, retweet network, and their relationship with the short text of the source tweet.
- 2) Clear and meaningful explanations for the predictions are provided through the incorporation of a dual co-attention mechanism.
- 3) Comprehensive experiments are conducted on publicly available datasets that demonstrate the superior performance of GRACE as compared to baseline models.

This paper is structured as follows:

- Section II provides an overview of existing methods for fake news detection.
- Section III defines the problem and outlines the objectives addressed by the proposed model.
- Section IV details the experimental setup used in this study.

- Section V presents the evaluation metrics and results obtained.
- Finally, Section VI concludes the paper with a summary of findings and contributions.

II. LITERATURE REVIEW

Fake news, though not a new phenomenon, has acquired significant public awareness in recent years, primarily due to its widespread impact on society, politics, and media [26]. As the dissemination of false content continues to evolve, the literature on fake news detection has expanded rapidly, addressing the various challenges posed by this issue. Existing research can be broadly categorized into two main approaches: content-based and network-based methods. Content-based approaches focus on analyzing the textual data of news articles to identify linguistic, syntactic, and semantic features that distinguish fake news from legitimate news [27]. While, network-based methods focus on user's interactions and relationships within social media networks. They explore how news spreads across platforms and how user engagement patterns influence the dissemination of misinformation [21]. This section provides an overview of these two categories of fake news detection techniques and highlight the key developments, strengths, and limitations.

Content-based approaches focus on analyzing the textual data of news articles to evaluate their truthfulness. These methods are especially effective for long range dependencies, as they allow for a deep analysis of linguistic and semantic features to identify signs of misinformation [27]. One widely used technique is TF-IDF, which measures the importance of specific words within a news story [28]. Topic modeling helps to uncover the underlying themes in the content. It offers a structured and meaningful representation of the text [29]. Other linguistic features, such as PoS tags, assertive or factive verbs, and markers of subjectivity, are commonly used to detect subtle language patterns [24]. Further, techniques that assess writing consistency and social emotions are applied to highlight anomalies in news content [30]. The underlying assumption of these content-based methods is that fake news will exhibit detectable differences in linguistic structure, topic, or emotional tone compared to genuine news articles [31].

However, traditional content-based methods face several challenges in detecting fake news, mainly when relying on handcrafted linguistic cues [13]. These cues, such as lexical and syntactic features, are often limited in generalizability across different languages, topics, and domains. These techniques struggle to capture the complex semantic and contextual information embedded in modern news articles [3]. As news articles evolve in structure, content-based approaches that rely solely on traditional methods become less effective. As a result, researchers are increasingly turning to deep learning models to address these limitations [14]. The approaches like Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Autoencoders [32] provide a solution by automatically learning hidden representations of text and capturing complex contextual patterns. These models eliminate the need for manually designed features and leverage word embeddings, such as word2vec, to enhance text representation and better identify patterns [33].

To make fake news detection more accurate, researchers have developed multi-modal models that combine different types of information, such as text and visuals, to improve their performance [15]. Visual elements like images and videos often play a significant role in how news is shared and perceived as credible. For instance, Bahad et al. introduced an RNN-based model that uses an attention mechanism to integrate text and visual information, allowing the system to focus on the most relevant features from both [34]. Similarly, Zhao et al. created a model that explores the relationship between text and visuals, which is especially effective in cases where misleading images are used to spread false claims [35].

To make detection systems more adaptable, researchers have also applied multi-task learning, enabling models to transfer knowledge across different types of content and better handle diverse contexts [36]. Since fake news evolves rapidly, new approaches like analyzing temporal patterns, adapting to specific domains, and leveraging weak supervision learning have been explored [37], [38], [10]. These innovations help detection systems stay scalable and flexible, allowing them to adapt to the ever-changing nature of misinformation. By combining these advancements, models are now better equipped to accurately and dynamically detect fake news in real-world scenarios.

Recent advancements in NLP have significantly improved the accuracy and reliability of content-based approaches. Transformer-based models [39], such as BERT (Bidirectional Encoder Representations from Transformers) [40] and GPT (Generative Pre-trained Transformer) [41], have revolutionized text representation and classification tasks by capturing contextual dependencies more effectively than traditional models. For instance, BERT has been fine-tuned for fake news detection by leveraging its bidirectional attention mechanism to understand subtle linguistic cues and context [42]. Similarly, GPT models have been employed to generate synthetic datasets for training effective classifiers and to analyze text for semantic coherence and logical consistency [35]. Hybrid models combining transformers with other neural architectures have also emerged. For example, a recent study integrated BERT with Graph Neural Networks (GNNs) to enhance performance by incorporating relationships between entities within news articles [43]. Other studies have focused on domain-specific adaptations of transformers, such as FakeBERT, which was trained on datasets tailored for misinformation detection [44]. These models not only outperform traditional approaches but also offer better generalization across domains and languages.

Network-based methods for detecting fake news focus on understanding how users interact with content on social media platforms [18]. Actions like commenting, retweeting, and following are critical to how information spreads and provide clues about the fake news propagation [19], [45]. By studying these patterns, researchers gained valuable insights into how to identify fake news and separate it from genuine content [46]. To model how news spreads, both homogeneous networks (where nodes and edges are similar) and heterogeneous networks (where they differ) are used [4].

Homogeneous networks, consisting of uniform nodes and edges, make it easier to study news spread within a unified structure [47]. A notable study by Chang et al. analysed the dispersal of false news on Twitter and found that false

news spreads faster, further, and more broadly than true news [19]. This observation highlights the accelerated nature of fake news diffusion. To enhance fake news detection, Huang et al. proposed a tree-structured RNN model that integrates textual features and propagation structure features [48]. Similarly, Gong et al. introduced a bi-directional GCN to learn representations of content semantics and diffusion structures [43].

In difference, heterogeneous networks consist of multiple nodes and edges, offering a more detailed representation of the relationships within the news ecosystem [49]. Kang et al. proposed a model that encodes semantic information and the global structure of the diffusion graph, incorporating posts, comments, and user interactions [50]. Huang et al. developed a meta-path-based heterogeneous graph attention network to capture the semantic relationships among text content in news propagation [48]. Additionally, Xie et al. enhanced the robustness of graph-based fake news detectors by modelling entities through a heterogeneous information network and utilizing graph adversarial learning to ensure more distinctive structural features [51]. Another significant advancement in heterogeneous network models was introduced by Nguyen et al. by developing Factual News Graph (FANG). This framework leverages social structures and user engagement patterns for effective fake news detection [44].

Network-based methods for fake news detection effectively handle multimodal data by leveraging the unique strengths of graph structures to integrate and process text and visual features. Jin et al. [52] proposed a Hierarchical Propagation Network that constructs a hierarchical graph where nodes represent multimodal features such as text embeddings, visual features, and user interactions. These nodes are interconnected through propagation layers that explicitly model the interplay between modalities, enabling a seamless integration of multimodal signals. Wang et al. [53] introduced a Multimodal Fusion Graph where text and image features are processed through graph attention layers, dynamically weighing their contributions to detect fake news. This method effectively links modalities by treating textual and visual embeddings as interconnected nodes in a unified graph. Shu et al. [54] utilized a Graph-based Multimodal Embedding framework, which creates a graph where text and image metadata are nodes, and the relationships between them (e.g. co-occurrence in news items) are edges. The GME approach ensures joint feature learning by allowing intermodal dependencies to be explicitly modeled and updated during training. Zhou et al. [55] extended this concept by employing knowledge-enhanced graphs, incorporating external knowledge from textual and visual data into the graph structure. Here, knowledge graph embeddings serve as additional nodes, creating a richer multimodal representation that enhances the interplay between modalities for accurate fake news detection. These approaches demonstrate how network-based methods construct and leverage graphs to unify and effectively process both modalities.

While these network-based models have shown promise, much previous work has focused on static networks. However, our research takes a dynamic approach by analyzing social media news within temporal diffusion networks, reflecting the continuous evolution of news propagation.

Approaches focusing on user behavior analyze the charac-

teristics of individuals who interact with news content, such as retweeting or commenting on stories. Yang et al. proposed extracting account-based features like the user's gender, hometown, and follower count [56]. Shu et al. found that user profiles associated with fake news differ significantly from those linked to legitimate news [4]. Liu et al. introduced a joint Recurrent and Convolutional Neural Network (CRNN) model that captures more detailed profiles of users, particularly those who retweet news stories [57]. In contrast, session-based heterogeneous graph embedding methods [51] rely on user session data to learn user traits but are not directly applicable to fake news detection.

III. MATERIALS AND METHODS

A. Preliminaries

Let $\mathcal{S} = \{\sigma_1, \sigma_2, \dots, \sigma_{|\mathcal{S}|}\}$ represent a collection of tweet stories, and $\mathcal{A} = \{\alpha_1, \alpha_2, \dots, \alpha_{|\mathcal{A}|}\}$ be a group of individuals (users) in the social media network. Each tweet story $\sigma_i \in \mathcal{S}$ is a short-text document, denoted by $\sigma_i = \{w_{i1}, w_{i2}, \dots, w_{il_i}\}$, where l_i is the number of words in the tweet story σ_i , and w_{ik} represents the k -th word in the story σ_i . Each user $\alpha_j \in \mathcal{A}$ is associated with a feature vector $\mathbf{v}_j \in \mathbb{R}^d$, where d is the dimensionality of the user's feature vector.

When a tweet story σ_i is shared, certain individuals will retweet it, forming a sequence of retweet records, referred to as the *retweet propagation path*. Let the propagation path of story σ_i be denoted by $\mathcal{P}_i = \{(\alpha_j, \mathbf{v}_j, t_j)\}$, where $(\alpha_j, \mathbf{v}_j, t_j)$ indicates that individual α_j with feature vector \mathbf{v}_j retweeted story σ_i at time t_j . Here, $j = 1, 2, \dots, K$, with $K = |\mathcal{P}_i|$ being the total number of retweets. The set of individuals who retweet story σ_i is denoted as $\mathcal{A}_i \subseteq \mathcal{A}$.

Within the propagation path \mathcal{P}_i , the individual α_1 is the original poster of story σ_i at time τ . For all subsequent individuals $j > 1$, individual α_j retweets the story at time τ_j , where $\tau_j > \tau_1$.

The tweet story σ_i is labeled with a binary value $\kappa_i \in \{0, 1\}$ to indicate its truthfulness, where:

$$\kappa_i = \begin{cases} 0 & \text{if the news } \sigma_i \text{ is true,} \\ 1 & \text{if the news } \sigma_i \text{ is fake.} \end{cases}$$

Given a tweet story σ_i and its corresponding propagation path \mathcal{P}_i (which includes individuals α_j who retweet the news and their associated feature vectors \mathbf{v}_j), the goal is to predict the authenticity κ_i of the story σ_i , a binary classification task.

The model should highlight a subset of individuals $\alpha_j \in \mathcal{A}_i$ who retweeted σ_i and a subset of words $w_{ik} \in \sigma_i$ that help to explain why σ_i is classified as either true or fake. This interpretability aspect is essential for understanding the reasoning behind the model's prediction.

B. Proposed Model

The GRACE model is developed to tackle the challenge of detecting fake news in social media networks by combining tweet content, user behavior, and the propagation dynamics of retweets. As depicted in Fig. 1, This model consists of several components including user characteristics extraction, news story encoding, user propagation representation, dual

co-attention mechanisms, and the final prediction layer. Each component is meticulously crafted to improve the model's ability to predict the truthfulness of a tweet while also providing interpretability by highlighting the users and words contributing to the classification.

The user characteristics extraction component involves creating a feature vector $\mathbf{x}_j \in \mathbb{R}^v$ for each user $u_j \in \mathcal{A}$, where v is the number of features. These features are derived from various aspects of a user's behavior, such as the number of followers, the number of retweets, the time difference between the tweet and retweet, and other profile-related information. This vector allows us to quantify how a user engages with content on social media, which is crucial for identifying fake news spreaders.

The source tweet σ_i is represented as a sequence of words, denoted by $\sigma_i = \{w_{i1}, w_{i2}, \dots, w_{il_i}\}$, where l_i is the number of words in tweet σ_i . We use a word-level encoder to represent this tweet. Each word w_{ik} in the tweet is initially encoded as a one-hot vector. A FC layer is applied to generate the word embeddings for each tweet, and the resulting embeddings are stored in a matrix $\mathbf{V} = [v_1, v_2, \dots, v_m] \in \mathbb{R}^{d \times m}$, where m is the length of the padded tweet and d is the dimensionality of the word embeddings.

To model the interactions among users who retweet the source tweet σ_i , we construct a graph $\mathcal{H}_i = (\mathcal{V}_i, \mathcal{F}_i)$, where \mathcal{V}_i represents the set of users who retweeted σ_i . The edges \mathcal{F}_i represent the interactions between users. Since the true interactions between users are unknown, we assume that the graph is fully connected. This implies that for every edge $e_{\alpha\beta} \in \mathcal{F}_i$, where $u_\alpha, u_\beta \in \mathcal{V}_i$ and $u_\alpha \neq u_\beta$, the number of edges is given by:

$$|\mathcal{F}_i| = \frac{n \cdot (n - 1)}{2} \quad (1)$$

where $n = |\mathcal{V}_i|$ is the number of users who retweeted σ_i .

To incorporate user features into the graph, we assign a weight w_{ab} to each edge $e_{ab} \in \mathcal{F}_i$, which is derived from the cosine similarity between the feature vectors \mathbf{u}_a and \mathbf{u}_b . The weight is calculated as:

$$w_{ab} = \frac{\mathbf{u}_a \cdot \mathbf{u}_b}{\|\mathbf{u}_a\| \|\mathbf{u}_b\|} \quad (2)$$

We use the adjacency matrix $\mathbf{W} = [w_{ab}] \in \mathbb{R}^{n \times n}$ to represent the weights between any pair of nodes v_a and v_b in the graph \mathcal{F}_i .

C. Graph Convolutional Network (GCN)

A Graph Convolutional Network (GCN) [25] is applied to propagate information through the graph \mathcal{F}_i . A GCN layer performs a convolution operation on the graph, updating node embeddings by aggregating information from their neighbors. For the graph \mathcal{F}_i , with adjacency matrix $\mathbf{\Pi}$ and feature matrix $\mathbf{\Lambda}$ representing user attributes in \mathcal{F}_i , the updated g -dimensional node feature matrix $\mathbf{\Omega}^{(l+1)} \in \mathbb{R}^{n \times g}$ at layer $l+1$ is calculated as:

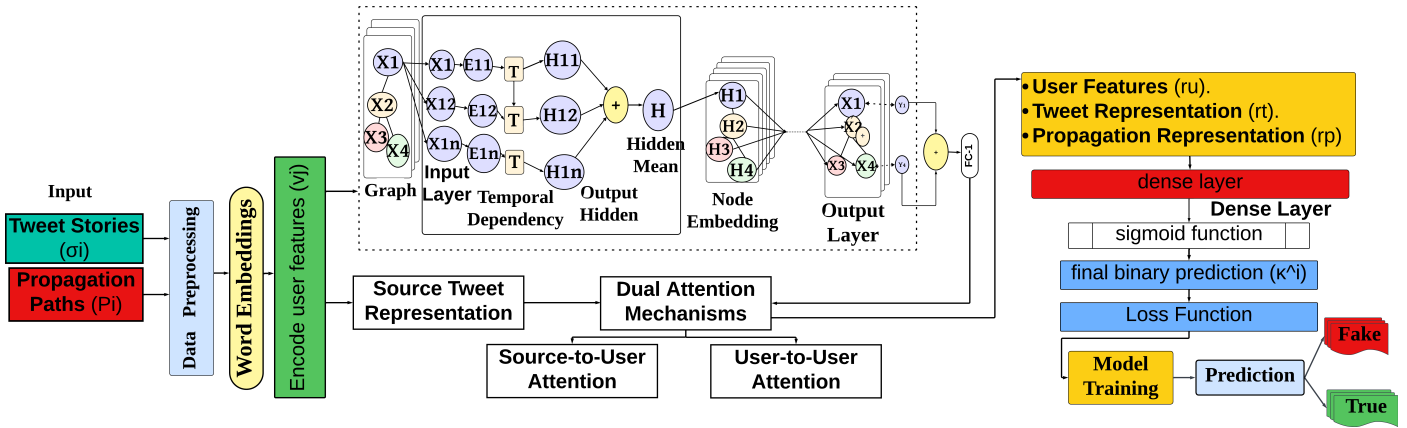


Fig. 1. The proposed model architecture diagram.

$$\Omega^{(l+1)} = \phi \left(\tilde{\Pi} \Omega^{(l)} \Gamma_l \right) \quad (3)$$

Here, $\tilde{\Pi} = \Sigma^{-1/2} \Pi \Sigma^{-1/2}$ represents the normalized adjacency matrix, Σ is the diagonal degree matrix, and ϕ is a non-linear activation function. This process is repeated iteratively over multiple layers, allowing information to propagate and be refined across the graph.

D. Co-attention Mechanisms

The correlation between the source tweet and users' interactions, including retweets, is captured using a dual co-attention mechanism. This mechanism simultaneously models the relationship between the source tweet and its retweets, as well as interactions between users within the propagation graph.

1) *Tweet-Retweet Correlation*: The first attention mechanism focuses on the relationship between the source tweet (\mathbf{Q}_σ) and the embeddings of retweets (\mathbf{Q}_u), which are derived from user propagation embeddings. The attention weights, representing the correlation between the content of the tweet and retweets, are computed as:

$$\mathbf{A}_\sigma = \text{softmax}(\mathbf{Q}_\sigma^T \mathbf{Q}_u) \quad (4)$$

These weights capture how strongly each retweet relates to the source tweet, refining both the source tweet and retweet representations for improved feature learning.

2) *User-User Correlation*: The second attention mechanism captures interactions between users by modeling the relationship between user embeddings across the propagation graph. This is achieved through:

$$\mathbf{A}_u = \text{softmax}(\mathbf{Q}_u^T \mathbf{Q}_u) \quad (5)$$

Here, the attention weights emphasize connections between users who share similar propagation behaviors, enabling the model to better understand the dynamics of retweet propagation.

By combining these two mechanisms, the model learns attention-driven representations that reflect the content and propagation dynamics of the source tweet and retweets. These

refined representations are used as inputs for the final prediction stage.

E. Final Prediction

The final prediction $\hat{\kappa}_i$ is obtained by combining the learned user features, source tweet embeddings, and propagation representations. The concatenated vector is passed through a fully connected layer with a sigmoid activation, producing a probability between 0 and 1 that represents the likelihood of the source tweet σ_i being fake. This process can be expressed as:

$$\hat{\kappa}_i = \sigma(\mathbf{W}_f \cdot [\mathbf{r}_u, \mathbf{r}_t, \mathbf{r}_p] + b_f) \quad (6)$$

where \mathbf{r}_u is the learned representation of user characteristics, \mathbf{r}_t is the learned embedding of the source tweet, and \mathbf{r}_p is the learned propagation representation of the users. The vector $[\mathbf{r}_u, \mathbf{r}_t, \mathbf{r}_p]$ is the concatenation of these representations, \mathbf{W}_f is the weight matrix, and b_f is the bias term. The sigmoid function $\sigma(\cdot)$ is applied to ensure that the output is a probability between 0 and 1.

F. Loss Function

The binary cross-entropy loss function is used for model training. It measures the difference between the predicted probability $\hat{\kappa}_i$ and the true label κ_i :

$$\mathcal{L}(\hat{\kappa}_i, \kappa_i) = -\kappa_i \log(\hat{\kappa}_i) - (1 - \kappa_i) \log(1 - \hat{\kappa}_i) \quad (7)$$

The loss function is minimized using the Adam optimizer, ensuring that the model's parameters are updated to reduce the classification error over time. The optimization process helps the model improve its predictions by adjusting weights, thereby minimizing the loss and enhancing the performance of the fake news detection system.

Algorithm 1 GRACE (Graph-based Attention for Coherent Explanation)

Input: Tweet stories $\mathcal{S} = \{\sigma_1, \dots, \sigma_{|\mathcal{S}|}\}$, user profiles \mathcal{A} , propagation paths \mathcal{P}_i , truthfulness labels κ_i .

Output: Predicted labels $\hat{\kappa}_i$ and explanation (highlighted users α_j and words w_{ik}).

1: **Initialize:** Pre-trained word embeddings, user feature vectors \mathbf{v}_j , graph adjacency matrices \mathbf{A} , and model parameters.

2: **for** each tweet $\sigma_i \in \mathcal{S}$ **do**

3: Encode tweet σ_i as word embeddings $\mathbf{V} \in \mathbb{R}^{d \times m}$.

4: Extract user feature vectors $\mathbf{v}_j \in \mathbb{R}^d$ for users in \mathcal{P}_i .

5: Construct a graph $\mathcal{H}_i = (\mathcal{V}_i, \mathcal{F}_i)$:

6: **for** each pair of users $(\alpha_\alpha, \alpha_\beta) \in \mathcal{V}_i$ **do**

7: **if** users are connected **then**

8: Compute edge weight:

$$\omega_{\alpha\beta} = \frac{\mathbf{x}_\alpha \cdot \mathbf{x}_\beta}{\|\mathbf{x}_\alpha\| \|\mathbf{x}_\beta\|}.$$

9: **end if**

10: **end for**

11: Apply GCN to propagate embeddings over \mathcal{H}_i :

$$\mathbf{H}^{(l+1)} = \rho \left(\mathbf{A} \tilde{\mathbf{H}}^{(l)} \mathbf{W}_l \right),$$

where ρ is a non-linear activation function.

12: Compute dual co-attention:

13: Source-to-user attention:

$$\mathbf{A}_\sigma = \text{softmax}(\mathbf{Q}_\sigma^T \mathbf{Q}_u).$$

14: User-to-user attention:

$$\mathbf{A}_u = \text{softmax}(\mathbf{Q}_u^T \mathbf{Q}_u).$$

15: Concatenate learned embeddings \mathbf{r}_u , \mathbf{r}_t , and \mathbf{r}_p :

$$\mathbf{r} = [\mathbf{r}_u, \mathbf{r}_t, \mathbf{r}_p].$$

16: Predict truthfulness:

$$\hat{\kappa}_i = \sigma(\mathbf{W}_f \cdot \mathbf{r} + b_f).$$

17: Highlight key users α_j and words w_{ik} based on \mathbf{A}_σ and \mathbf{A}_u .

18: **end for**

19: Optimize model parameters by minimizing the binary cross-entropy loss:

$$\mathcal{L}(\hat{\kappa}_i, \kappa_i) = -\kappa_i \log(\hat{\kappa}_i) - (1 - \kappa_i) \log(1 - \hat{\kappa}_i).$$

IV. EXPERIMENTAL SETUP

The GRACE model is implemented using the PyTorch framework. The tweet text is represented using pre-trained word embeddings. Each word in the tweet is mapped to its corresponding vector representation. These embeddings help transform the raw text into a meaningful numerical format suitable for further processing. GCN layers capture the interactions among users who retweet the source tweet. The graph represents users as nodes, and the interactions between users (such as retweeting) form the edges. Each node's feature vector is updated based on its neighbours, allowing the model to learn user-specific representations in the context of retweet propagation. The number of GCN layers is set to 3, with each layer processing information from the node's direct and indirect neighbours. These features are concatenated after obtaining the embeddings from the tweet content, user characteristics, and user propagation representations. The concatenated vector is passed through fully connected (dense) layers to make the final classification decision. The hidden layers in the fully connected section use ReLU activation, while the output layer employs a sigmoid activation function to predict the probability of a fake

tweet.

A. Hyperparameters

The proposed model is designed with several key hyperparameters that allow for efficient and effective training as described in Table I. A learning rate of 0.001 was selected after a grid search of several potential values. This choice balances convergence speed and stability, ensuring that the model trains effectively without overshooting the optimal solution. The batch size was set to 64, a commonly used value in graph-based models like GCNs. A batch size of this allows for efficient computation and good convergence properties while maintaining memory efficiency during training.

The model architecture is developed with three GCN layers, which strike a balance between capturing the interactions within the retweet network and avoiding overfitting caused by excessive depth. Each GCN layer contains 128 hidden units, which are sufficient to learn rich user interaction features without making the model too large and prone to overfitting. A dropout rate of 0.3 is applied to mitigate overfitting, meaning

30% of the neurons are randomly dropped during training, helping the model avoid reliance on specific features.

Following the GCN layers, fully connected (FC) layers were added with 256 hidden units to combine and process features from tweet content, user characteristics, and propagation patterns. To reduce overfitting in these layers, a higher dropout rate of 0.5 was applied, randomly dropping 50% of the neurons during training to improve generalization.

ReLU activation is used throughout the hidden layers to introduce non-linearity, enabling the model to learn more complex patterns and decision boundaries effectively. The output layer uses the sigmoid activation function, which maps the final output to a probability between 0 and 1. This value is interpreted as the likelihood that a given tweet is fake. The Adam optimizer was chosen for optimisation, known for its efficiency in handling sparse gradients and large datasets. The binary cross-entropy loss function was used as the loss criterion, as it is well-suited for binary classification tasks like fake news detection. The model was trained for 20 epochs, which is sufficient for convergence without overfitting. These hyperparameters were carefully chosen to ensure the model performs well on the fake news detection task, balancing model complexity, training efficiency, and the ability to generalize to unseen data.

TABLE I. HYPERPARAMETERS FOR GCAN MODEL

Hyperparameter	Value
Learning Rate	0.001
Batch Size	64
GCN's Layers	3
Hidden Units	128
Dropout Rate	0.3
Hidden Units in FC Layers	256
Dropout Rate (FC layers)	0.5
Activation Function (Hidden)	ReLU
Activation Function (Output)	Sigmoid
Optimizer	Adam
Loss Function	Binary Cross-Entropy
Epochs	20

B. Datasets

This study utilizes two widely used datasets, Twitter15 and Twitter16, compiled by Ma et al. [58], which are recognized benchmarks in the field of fake news detection. These datasets provide a comprehensive basis for evaluating propagation-based modeling approaches, as they include tweets along with the corresponding sequences of retweeting users, which are essential for capturing propagation dynamics.

The Twitter15 dataset includes 1,490 claims, while Twitter16 contains 818 claims. Both datasets are annotated with four ground truth veracity labels: True News (T), Fake News (F), Non-Fake News (NF), and Unverified News (U). For our binary classification experiments, we focus only on True News (T) and Fake News (F) labels, aligning with the scope of our study.

These datasets are particularly suitable for evaluating our proposed model as they include rich propagation structures that allow us to assess the effectiveness of graph-based approaches. Additionally, they represent real-world social media interactions, offering realistic challenges and scenarios for fake news detection.

To enrich the data, we collected user profile information using the Twitter API, as the original datasets do not include user profiles. This additional data allows us to incorporate user-specific features, such as activity patterns and engagement metrics, which are crucial for analyzing user behavior in the context of fake news propagation.

The datasets are divided into three parts: 70% for training, 15% for testing, and 15% for validation. This ensures a balanced and rigorous evaluation of the model. Table II and Fig. 2 provide a summary of the key statistics and label distributions, illustrating the diversity and scale of the datasets.

These choices ensure that our approach is validated against reliable, well-established benchmarks, offering a fair comparison with prior works and a robust demonstration of the proposed model's effectiveness.

TABLE II. DATASET STATISTICS

Feature	Twitter15	Twitter16
Total Claims	1,490	818
True News (T)	370	205
Fake News (F)	374	204
Non-Fake News (NF)	374	203
Unverified News (U)	372	206
Total Postings	331,612	204,820
Users	190,868	115,036
Avg. Retweets per Story	292.19	308.70
Avg. Words per Source	13.25	12.81
# Total Nodes	912,638	501,032
# Total Edges	697,523	382,936

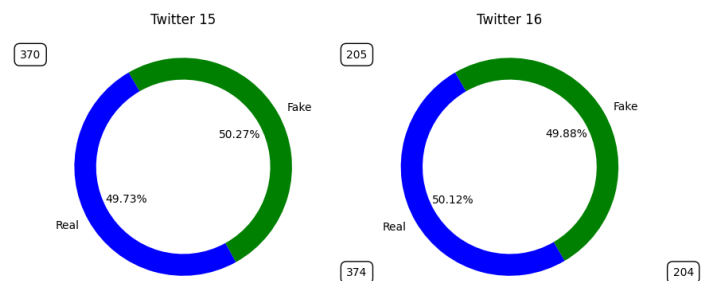


Fig. 2. Label distribution for Twitter15 and Twitter16 datasets.

C. Evaluation Metrics

To assess the proposed model's performance for fake news detection, we use several key metrics that provide insights into its effectiveness. These metrics include Accuracy, Precision, Recall, F1 Score, and the Area Under the Receiver Operating Characteristic Curve (AUC).

Accuracy is the most straightforward metric, measuring the overall correctness of the model across all predictions. It is the ratio of correct predictions to the total number of predictions. **Precision** evaluates the proportion of positive predictions (predicted fake news) that are actually correct. A high Precision indicates that the model is accurate when it predicts fake news. **Recall** focuses on the model's ability to capture all actual positive instances (actual fake news). It is the ratio of true positives to the sum of true positives and false negatives. A high Recall means that the model successfully identifies most of the true fake news instances. **F1 Score** is the

harmonic mean of Precision and Recall. It provides a balanced measure and offers a single number that evaluates the model's performance in relevance and completeness.

V. RESULTS

Results are reported in Table III. The GRACE model demonstrated notable accuracy and F1 score improvements across the Twitter15 and Twitter16 datasets. The F1 score increased by 2.42% from baseline models, reaches at 84.50, while accuracy improved by 2.08%, achieving 89.50 on the Twitter15 dataset. On the Twitter16 dataset, the F1 score saw a 2.07% improvement, reaching 77.50, and accuracy increased by 1.83%, reaching 92.50. On average, the GRACE model showed a 2.24% improvement in F1 score and a 1.95% improvement in accuracy across both datasets. These results reflect the model's consistent enhancement in both key performance metrics. The GRACE model's improvements indicate its strong capacity to achieve higher classification precision and accuracy than baseline models, showcasing its ability to generalize well across different datasets. The bigger improvements in Twitter15 suggest the model's adaptability in handling diverse data characteristics, while its solid performance in Twitter16 further emphasizes its robustness in real-world, noisy data scenarios.

A. Baseline Models

The proposed model is compared with several baseline methods on the Twitter15 and Twitter16 datasets, as shown in Table III. The GCAN (Graph-aware Co-attention Network) predicts fake news by considering the original tweet and its propagation, with a variant, GCAN-G, which excludes the graph convolution component to evaluate the effectiveness of graph-aware representations [21]. SVM-TS combines a Support Vector Machine with heuristic rules and a time-series structure to classify posts as fake or real, leveraging hand-crafted features. While effective initially, deep learning models now outperform traditional approaches due to superior feature extraction capabilities [59]. DTC is a rumor detection method that uses a Decision Tree classifier and leverages various hand-crafted features to evaluate information credibility [60]. It focuses on extracting and analyzing features related to content, user behavior, and network interactions to improve detection accuracy. CRNN, a deep residual network, integrates four cascading graph convolutional networks to capture long-range dependencies and nonlinear features effectively [61]. RFC is a ranking method based on Random Forest that refines and elaborates the inquiry phrases within posts. By leveraging this approach, it aims to enhance the analysis and prioritization of relevant information [62]. dDEFEND represents tweet contents and interaction graphs in a latent space, capturing multi-level features of fake news through claim-aware and inference-based attention mechanisms [63]. The CSI model stands out as an advanced fake news detection model that integrates both article content and the collective behaviour of users propagating fake news [64]. This model uses LSTM to capture sequential dependencies and computes user-specific scores to evaluate the likelihood of a tweet being fake. The tCNN model introduces a modified Convolutional Neural Network (CNN) to learn local variations in user profile sequences, combining them with features from the source tweet [65]. This approach

effectively captures intricate variations in user behaviour. The CRNN merges CNN and RNN to learn local and global user profile variations [66]. This hybrid technique enables the model to capture temporal and spatial dependencies in retweet propagation. mGRU is a modified gated recurrent unit (GRU) model designed for rumor detection. It captures temporal patterns by leveraging retweet user profiles in combination with the source tweet's features [58].

The confusion matrices are presented in Fig. 3. These metrics show the model's performance in classifying news across multiple categories. For the Twitter15 dataset, the model correctly identifies True News, with 109 instances accurately classified, while only four are misclassified as False News and seven as Unverified News. This indicates the model's proficiency in distinguishing authentic information. The model successfully classifies 36 instances for False News, with minimal misclassifications (6 as True News and one as Unverified News). In Twitter16 dataset, The model accurately identifies True News, classifying 56 instances correctly, while only three are misclassified as False News and four as Unverified News. It also performs well in detecting False News, correctly classifying 23 instances, with just a few misclassifications (2 instances each into True News and Unverified News). The

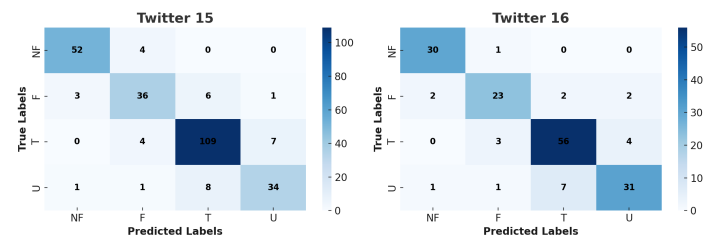


Fig. 3. Confusion matrices for Twitter15, Twitter16 on test dataset.

differences in classification accuracy across these two datasets can be attributed to the varying complexity of the classification tasks. While both datasets include multiple categories, the Twitter15 and Twitter16 datasets introduce the additional challenge of distinguishing Unverified News from True and False News, resulting in a higher degree of misclassification, especially between False News and Unverified News, which share similar content characteristics. These results underscore the model's adaptability in handling both binary and multi-class classification challenges, demonstrating its effectiveness across diverse datasets.

The performance of the proposed model is evaluated in terms of accuracy in Fig. 4 for early detection. It is analyzed under varying conditions by altering the number of observed retweet users per source story, ranging from 10 to 50. The results demonstrate that GRACE consistently and significantly outperforms all competing methods across all scenarios. Despite as few as 10 observed retweeters, GRACE achieves an impressive 82% accuracy on Twitter16, underscoring its robustness and reliability. These findings highlight GRACE's capability to deliver accurate and early detection of fake news dissemination, which is critical for effectively combating misinformation and mitigating its impact.

We assess the effectiveness of proposed approach and baseline models for early stage fake news detection. Early

TABLE III. COMPARISON OF PROPOSED MODEL WITH BASELINE AND STATE-OF-THE-ART MODELS ON TWITTER15 AND TWITTER16 DATASETS

Method	Twitter15				Twitter16			
	F1	Recall	Precision	Accuracy	F1	Recall	Precision	Accuracy
DTC	49.48	48.06	49.63	49.49	56.16	53.69	57.53	56.12
SVM-TS	51.90	51.86	51.95	51.95	69.15	69.10	69.28	69.32
mGRU	51.04	51.48	51.45	55.47	55.63	56.18	56.03	66.12
GCAN-G	79.38	79.90	79.59	86.36	67.54	68.02	67.85	79.39
RFC	46.42	53.02	57.18	53.85	62.75	65.87	73.15	66.20
tCNN	51.40	52.06	51.99	58.81	62.00	62.62	62.48	73.74
CRNN	52.49	53.05	52.96	59.19	63.67	64.33	64.19	75.76
CSI	71.74	68.67	69.91	69.87	63.04	63.09	63.21	66.12
GCAN	82.50	82.95	82.57	87.67	75.93	76.32	75.94	90.84
dDEFEND	65.41	66.11	65.84	73.83	63.11	63.84	63.65	70.16
GRACE	84.17	84.95	84.74	89.53	77.51	78.09	77.73	79.11

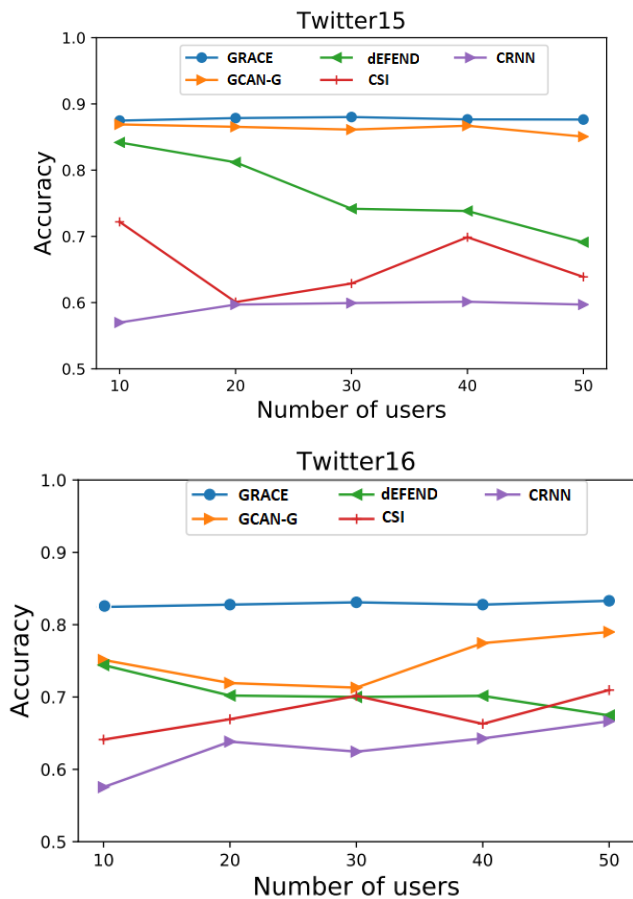


Fig. 4. Accuracy over retweet users on Twitter15 and Twitter16 datasets.

identification of fake news is essential to curbing its spread and minimizing its harmful societal impacts. For this evaluation, we use a specific tweet’s propagation time or initial release time within a news event as the detection deadline. Tweets posted beyond this deadline are excluded from consideration. To compare the performance of various detection methods, we vary the detection time points within a specific range and analyze their performance.

Fig. 5 presents the accuracy of all methods at different time intervals across three datasets. The results indicate that

GRACE consistently performs better than baseline models in early-stage fake news detection. Across all datasets, the accuracy of all methods improves rapidly during the early stages of information diffusion. Notably, our model exhibits a distinct performance advantage as the propagation progresses, demonstrating its ability to sustain high accuracy over time and effectively adapt to the dynamics of fake news dissemination.

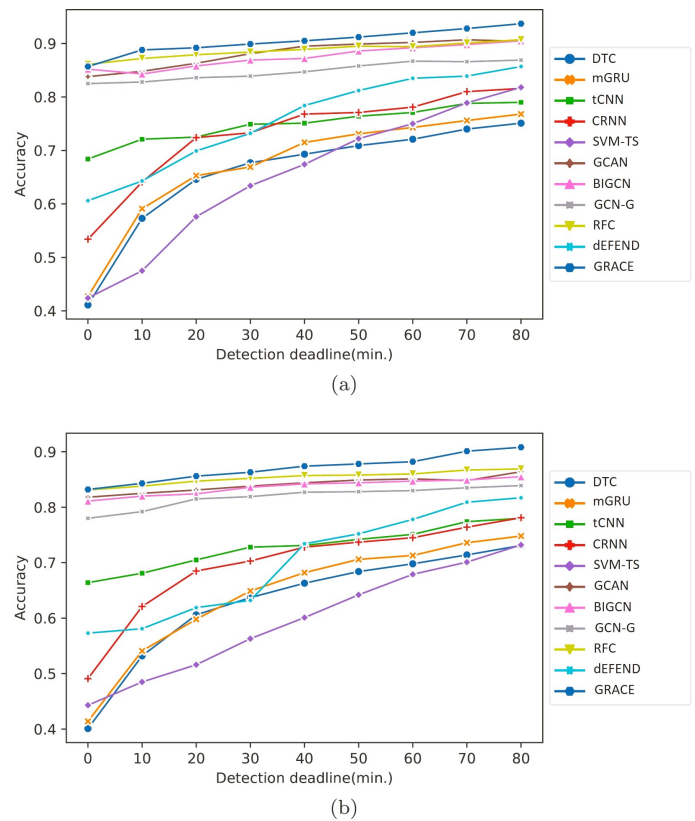


Fig. 5. (a) Early detection of fake news on Twitter15; (b) early detection of fake news on Twitter16.

The source-propagation co-attention mechanism embedded in our proposed model offers meaningful insights into identifying the characteristics of suspicious users and the linguistic cues they emphasize during the spread of information. As Fig. 6 demonstrates, the model highlights several distinct traits commonly associated with suspicious retweeters. These

include unverified accounts, newly created profiles with shorter account lifespans, minimal user descriptions, and shorter graph path lengths connecting them to the source tweet’s author.

Moreover, the analysis reveals that these users focus disproportionately on specific words, such as “breaking” and “pipeline,” often used in sensationalized or misleading content. By leveraging these observations, the model enhances its ability to detect fake news and provides interpretability by uncovering suspicious accounts’ behavioural patterns and language preferences. Such explainability is crucial for understanding the underlying mechanisms of fake news dissemination and developing strategies to mitigate its spread effectively.

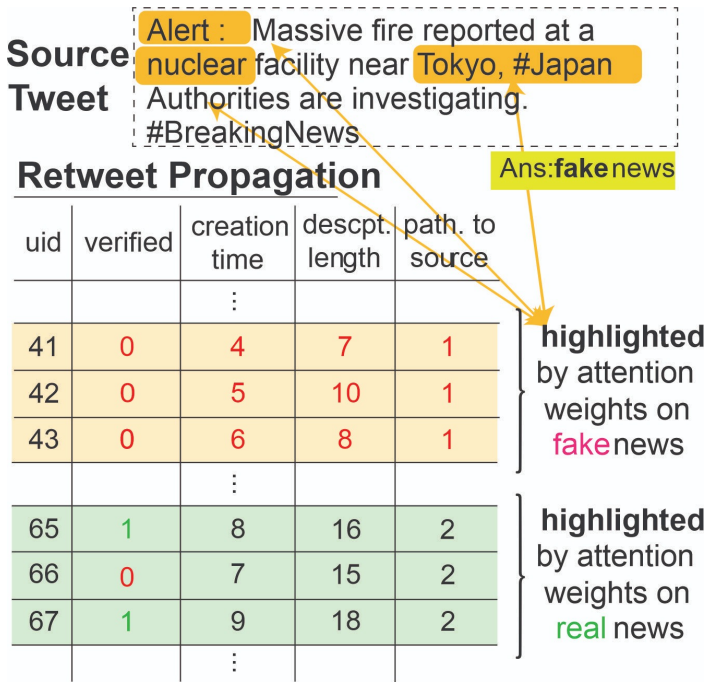


Fig. 6. Key evidential words identified by the GRACE model in the source tweet (top) and suspicious users flagged during the retweet propagation process (bottom). Each column corresponds to a specific user characteristic, providing deeper insights into user behaviours. For simplicity, only a select number of user characteristics are presented.

B. Ablation Study

The ablation study is conducted in Table IV. It highlights the significance of each component in the proposed model. Removing the dual co-attention mechanism (“-A”) leads to a noticeable drop in performance, which underscores its role in linking tweet content with user interactions and propagation dynamics. Excluding the graph-aware representation (“-G”) also affects the model’s accuracy, as it captures the structural relationships between users in the retweet network. Similarly, removing the user characteristics module (“-U”), which captures behavioural patterns like account age and retweet frequency, significantly reduces the model’s ability to detect suspicious users. The absence of source tweet embeddings (“-S”) results in a substantial decline, showing the importance of semantic content in distinguishing fake news. The most severe performance degradation occurs when the source tweet embeddings and dual co-attention mechanism are removed (“-S-A”),

demonstrating that integrating content-based and interaction-based features is crucial for achieving high accuracy. These results confirm that each component contributes meaningfully to the overall effectiveness of the GRACE model.

C. Discussion

The findings from our study highlight the robustness and interpretability of the GRACE model in detecting fake news across various datasets and scenarios. By leveraging multiple data modalities [58], such as user characteristics, tweet content, and propagation dynamics, GRACE achieves superior performance compared to existing baseline models. This discussion contextualizes these results, explores their implications, and addresses the model’s broader applicability and potential limitations.

One of the most significant insights from our work is the importance of integrating user behavior and propagation dynamics into fake news detection. Traditional models often focus solely on tweet content, neglecting the behavioral and relational cues that can provide essential context [36], [54]. GRACE fills this gap by incorporating graph-aware propagation modeling and user embedding representation, which allows it to capture the underlying social dynamics in retweet propagation. This synergy between components is evident from our ablation study, where removing key elements, such as the dual co-attention mechanism or graph-based user representations, led to noticeable drops in performance.

The results also reveal GRACE’s adaptability in both the early and advanced stages of fake news propagation. For instance, GRACE’s ability to maintain high accuracy with limited early-stage data (e.g. as few as 10 retweeters) underscores its potential for real-time applications. This early detection capability is crucial for mitigating the spread of misinformation, as even a small delay in identification can result in widespread dissemination and societal harm.

Another strength of GRACE lies in its explainability. The co-attention mechanism enables the model to highlight the specific words in tweets and user behaviors contributing to its predictions. For instance, the model identified linguistic patterns, such as emotionally charged words like “breaking”, and behavioural traits, including unverified accounts and recently created profiles, as key indicators of suspicious activity. This interpretability is vital for building trust with end-users, particularly in applications where automated decisions must be transparent and defensible.

Understanding the characteristics of suspicious users and the propagation patterns of fake news provides actionable insights for social media platforms and policymakers. By identifying high-risk accounts and content early, GRACE can assist in designing targeted interventions, such as flagging or debunking misleading posts before they gain significant traction.

D. Limitation and Future Work

While GRACE demonstrates strong performance and interpretability, it is not without limitations. One of the primary challenges is the reliance on user interaction data to build propagation graphs. The model’s effectiveness could

TABLE IV. ABLATION STUDY RESULTS OF GRACE ON TWITTER15 AND TWITTER16 DATASETS

Method	Twitter15				Twitter16			
	F1	Rec	Precision	Accuracy	F1	Recall	Precision	Accuracy
Full Model	84.17	84.95	84.74	89.53	77.51	78.09	77.73	79.11
-A	81.45	82.13	80.97	87.12	74.89	75.12	74.45	76.45
-G	82.03	82.67	81.45	87.67	75.43	75.87	75.01	77.02
-U	80.12	80.89	79.68	85.34	73.25	73.98	72.87	74.34
-S	78.65	79.02	78.30	83.21	72.11	72.56	71.43	72.89
-S-A	75.34	75.89	74.12	80.78	70.34	70.92	69.87	71.21

be reduced if user data is incomplete or anonymized due to privacy concerns. Additionally, while GRACE assumes a fully connected graph without explicit user relationships, this assumption may not always reflect real-world interactions, potentially leading to inaccuracies in propagation modelling. Future work could explore incorporating more advanced graph representation techniques, such as dynamic graph neural networks, to better model evolving user interactions over time to enhance GRACE further. Additionally, leveraging external knowledge bases or fact-checking databases could improve the model's ability to validate content credibility, particularly for previously unseen news stories. Finally, expanding GRACE to handle multilingual content and adapting it to different cultural contexts would increase its applicability on a global scale.

VI. CONCLUSION

In this study, we introduced Graph-based Attention for Coherent Explanation (GRACE) approach for detecting fake news on social media platforms. GRACE addresses the complex and dynamic nature of misinformation by leveraging tweet content, user behaviour, and retweet propagation dynamics, making it capable of identifying fake news with high accuracy and interpretability. Unlike traditional methods, GRACE operates in a more realistic and challenging setting by focusing on short-text tweets and their retweeter sequences, closely aligning with the real-world propagation of misinformation. The evaluation results underscore GRACE's robustness and effectiveness, demonstrating its ability to deliver accurate predictions while maintaining explainability through its co-attention mechanism. Notably, GRACE excels in early-stage detection, achieving satisfying performance even with limited propagation data. This early detection capability is critical for minimizing the spread of misinformation and reducing its societal impact.

Beyond fake news detection, GRACE has broader applications for other short length text classification tasks in social media, such as sentiment analysis and tweet popularity prediction. Its flexible and modular design makes it a promising candidate for addressing various social media challenges. Future work will enhance the model's generalization capabilities to accommodate different platforms and contexts.

REFERENCES

- [1] L. Humphreys, *The qualified self: Social media and the accounting of everyday life*. MIT press, 2018.
- [2] S. K. Rathi, B. Keswani, R. K. Saxena, S. K. Kapoor, S. Gupta, and R. Rawat, *Online Social Networks in Business Frameworks*. John Wiley & Sons, 2024.
- [3] W. Xu, J. Wu, Q. Liu, S. Wu, and L. Wang, "Evidence-aware Fake News Detection with Graph Neural Networks," *arXiv e-prints*, p. arXiv:2201.06885, Jan. 2022.
- [4] K. Shu, D. Mahudeswaran, S. Wang, D. Lee, and H. Liu, "Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media," *Big data*, vol. 8, no. 3, pp. 171–188, 2020.
- [5] B. D. Horne, D. Nevo, and S. L. Smith, "Ethical and safety considerations in automated fake news detection," *Behaviour & Information Technology*, pp. 1–22, 2023.
- [6] F. Odum, "Covid conspiracy narratives: Dissecting the origins of misinformation in digital space," 2021.
- [7] M. Farokhian, V. Rafe, and H. Veisi, "Fake news detection using dual bert deep neural networks," *Multimedia Tools and Applications*, vol. 83, no. 15, pp. 43 831–43 848, 2024.
- [8] D. O. Ong'ong'a, "The role of online news consumers in lessening the extent of misinformation on social media platforms," *Journal Communication Spectrum: Capturing New Perspectives in Communication*, vol. 12, no. 2, pp. 96–111, 2022.
- [9] A. Damisa, "Fake news: Finding truth in strategic communication," 2024.
- [10] A. Bruns, E. Hurcombe, and S. Harrington, "Covering conspiracy: Approaches to reporting the covid/5g conspiracy theory," *Digital Journalism*, vol. 10, no. 6, pp. 930–951, 2022.
- [11] Y.-P. Chen, Y.-Y. Chen, K.-C. Yang, F. Lai, C.-H. Huang, Y.-N. Chen, and Y.-C. Tu, "The prevalence and impact of fake news on covid-19 vaccination in taiwan: retrospective study of digital media," *Journal of Medical Internet Research*, vol. 24, no. 4, p. e36830, 2022.
- [12] A. A. Abd El-Mageed, A. A. Abohany, A. H. Ali, and K. M. Hosny, "An adaptive hybrid african vultures-aquila optimizer with xgb-tree algorithm for fake news detection," *Journal of Big Data*, vol. 11, no. 1, p. 41, 2024.
- [13] V. U. Gongane, M. V. Munot, and A. Anuse, "Machine learning approaches for rumor detection on social media platforms: a comprehensive survey," *Advanced machine intelligence and signal processing*, pp. 649–663, 2022.
- [14] A. Yadav and A. Gupta, "An emotion-driven, transformer-based network for multimodal fake news detection," *International Journal of Multimedia Information Retrieval*, vol. 13, no. 1, pp. 1–16, 2024.
- [15] S. Tufchi, A. Yadav, and T. Ahmed, "A comprehensive survey of multimodal fake news detection techniques: advances, challenges, and opportunities," *International Journal of Multimedia Information Retrieval*, vol. 12, no. 2, p. 28, 2023.
- [16] K. Soga, S. Yoshida, and M. Muneyasu, "Exploiting stance similarity and graph neural networks for fake news detection," *Pattern Recognition Letters*, vol. 177, pp. 26–32, 2024.
- [17] A. Ali and M. Gulzar, "An improved fakebert for fake news detection," *Applied Computer Systems*, vol. 28, no. 2, pp. 180–188, 2023.
- [18] Z. Zhang, Q. Lv, X. Jia, W. Yun, G. Miao, Z. Mao, and G. Wu, "Gbca: Graph convolution network and bert combined with co-attention for fake news detection," *Pattern Recognition Letters*, 2024.
- [19] Q. Chang, X. Li, and Z. Duan, "Graph global attention network with memory: A deep learning approach for fake news detection," *Neural Networks*, vol. 172, p. 106115, 2024.
- [20] Y. Zhang, S. Li, J. Weng, and B. Liao, "Gnn model for time-varying matrix inversion with robust finite-time convergence," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.
- [21] Y.-J. Lu and C.-T. Li, "Gcan: Graph-aware co-attention networks for explainable fake news detection on social media," 2020.

- [22] S. Xu, X. Liu, K. Ma, F. Dong, B. Riskhan, S. Xiang, and C. Bing, "Rumor detection on social media using hierarchically aggregated feature via graph neural networks," *Applied Intelligence*, vol. 53, no. 3, pp. 3136–3149, 2023.
- [23] L. Wei, D. Hu, W. Zhou, Z. Yue, and S. Hu, "Towards propagation uncertainty: Edge-enhanced bayesian graph convolutional networks for rumor detection," 2021.
- [24] D. S. Asudani, N. K. Nagwani, and P. Singh, "Impact of word embedding models on text analytics in deep learning environment: a review," *Artificial intelligence review*, vol. 56, no. 9, pp. 10 345–10 425, 2023.
- [25] P. Veličković, "Everything is connected: Graph neural networks," *Current Opinion in Structural Biology*, vol. 79, p. 102538, 2023.
- [26] R. Rodríguez-Ferrández, "The plandemic and its apostles: Conspiracy theories in pandemic mode," in *Digital totalitarianism*. Routledge, 2022, pp. 62–83.
- [27] N. Capuano, G. Fenza, V. Loia, and F. D. Nota, "Content-based fake news detection with machine and deep learning: a systematic review," *Neurocomputing*, vol. 530, pp. 91–103, 2023.
- [28] A. Widiyanto, E. Pebriyanto, F. Fitriyanti, and M. Marna, "Document similarity using term frequency-inverse document frequency representation and cosine similarity," *Journal of Dinda: Data Science, Information Technology, and Data Analytics*, vol. 4, no. 2, pp. 149–153, 2024.
- [29] L.-C. Chen, "An extended tf-idf method for improving keyword extraction in traditional corpus-based research: An example of a climate change corpus," *Data & Knowledge Engineering*, p. 102322, 2024.
- [30] M. H. Al-Tai, B. M. Nema, and A. Al-Sherbaz, "Deep learning for fake news detection: Literature review," *Al-Mustansiriyah Journal of Science*, vol. 34, no. 2, pp. 70–81, 2023.
- [31] I. Alshubaily, "Textcnn with attention for text classification," *arXiv preprint arXiv:2108.01921*, 2021.
- [32] A. R. Merryton and M. Gethsiyal Augasta, "An attribute-wise attention model with bilstm for an efficient fake news detection," *Multimedia Tools and Applications*, vol. 83, no. 13, pp. 38 109–38 126, 2024.
- [33] A. Mallik and S. Kumar, "Word2vec and lstm based deep learning technique for context-free fake news detection," *Multimedia Tools and Applications*, vol. 83, no. 1, pp. 919–940, 2024.
- [34] P. Bahad, P. Saxena, and R. Kamal, "Fake news detection using bi-directional lstm-recurrent neural network," *Procedia Computer Science*, vol. 165, pp. 74–82, 2019.
- [35] M. Zhao, Y. Zhang, and G. Rao, "Fake news detection based on dual-channel graph convolutional attention network," *The Journal of Supercomputing*, pp. 1–22, 2024.
- [36] T. Bian, X. Xiao, T. Xu, P. Zhao, W. Huang, Y. Rong, and J. Huang, "Rumor detection on social media with bi-directional graph convolutional networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 01, 2020, pp. 549–556.
- [37] H. Thakar and B. Bhatt, "Fake news detection: recent trends and challenges," *Social Network Analysis and Mining*, vol. 14, no. 1, p. 176, 2024.
- [38] B. Das *et al.*, "Multi-contextual learning in disinformation research: A review of challenges, approaches, and opportunities," *Online Social Networks and Media*, vol. 34, p. 100247, 2023.
- [39] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need.(nips), 2017," *arXiv preprint arXiv:1706.03762*, vol. 10, p. S0140525X16001837, 2017.
- [40] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2019. [Online]. Available: <https://arxiv.org/abs/1810.04805>
- [41] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
- [42] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.
- [43] S. Gong, R. O. Sinnott, J. Qi, and C. Paris, "Fake news detection through graph-based neural networks: A survey," *arXiv preprint arXiv:2307.12639*, 2023.
- [44] V.-H. Nguyen, K. Sugiyama, P. Nakov, and M.-Y. Kan, "Fang: Leveraging social context for fake news detection using graph representation," in *Proceedings of the 29th ACM international conference on information & knowledge management*, 2020, pp. 1165–1174.
- [45] Q. Chang, X. Li, and Z. Duan, "A novel approach for rumor detection in social platforms: Memory-augmented transformer with graph convolutional networks," *Knowledge-Based Systems*, vol. 292, p. 111625, 2024.
- [46] C. Cui and C. Jia, "Propagation tree is not deep: Adaptive graph contrastive learning approach for rumor detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 1, 2024, pp. 73–81.
- [47] Y. Zhao, W. Li, F. Liu, J. Wang, and A. M. Luvembe, "Integrating heterogeneous structures and community semantics for unsupervised community detection in heterogeneous networks," *Expert Systems with Applications*, vol. 238, p. 121821, 2024.
- [48] Q. Huang, C. Zhou, J. Wu, L. Liu, and B. Wang, "Deep spatial-temporal structure learning for rumor detection on twitter," *Neural Computing and Applications*, vol. 35, no. 18, pp. 12 995–13 005, 2023.
- [49] G. Zhang, D. Li, H. Gu, T. Lu, and N. Gu, "Heterogeneous graph neural network with personalized and adaptive diversity for news recommendation," *ACM Transactions on the Web*, vol. 18, no. 3, pp. 1–33, 2024.
- [50] M. Kang, G. F. Templeton, E. T. Lee, and S. Um, "A method framework for identifying digital resource clusters in software ecosystems," *Decision Support Systems*, vol. 177, p. 114085, 2024.
- [51] B. Xie, X. Ma, J. Wu, J. Yang, S. Xue, and H. Fan, "Heterogeneous graph neural network via knowledge relations for fake news detection," in *Proceedings of the 35th International Conference on Scientific and Statistical Database Management*, 2023, pp. 1–11.
- [52] Z. Jin, J. Ma, S. Wang, J. Tang, and J. Luo, "Hierarchical propagation network for fake news detection," in *Proceedings of the 29th International Conference on Information and Knowledge Management (CIKM)*. ACM, 2020, pp. 802–811.
- [53] S. Wang, Y. Zhang, X. Wang, and J. Li, "Multimodal fusion graph neural networks for fake news detection," *IEEE Transactions on Multimedia*, vol. 23, pp. 4397–4407, 2021.
- [54] K. Shu, D. Mahudeswaran, and H. Liu, "Graph-based multimodal embedding for fake news detection," in *Proceedings of The Web Conference (WWW)*. ACM, 2019, pp. 291–300.
- [55] X. Zhou, W. Lin, J. Zhang, and Y. Sun, "Incorporating knowledge graphs in multimodal fake news detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*. AAAI, 2022, pp. 5678–5685.
- [56] P. Yang, J. Leng, G. Zhao, W. Li, and H. Fang, "Rumor detection driven by graph attention capsule network on dynamic propagation structures," *The Journal of Supercomputing*, vol. 79, no. 5, pp. 5201–5222, 2023.
- [57] T. Liu, Q. Cai, C. Xu, Z. Zhou, F. Ni, Y. Qiao, and T. Yang, "Rumor detection with a novel graph neural network approach," *arXiv preprint arXiv:2403.16206*, 2024.
- [58] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, and M. Cha, "Detecting rumors from microblogs with recurrent neural networks," 2016.
- [59] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.-F. Wong, "Detect rumors using time series of social context information on microblogging websites," in *Proceedings of the 24th ACM international conference on information and knowledge management*, 2015, pp. 1751–1754.
- [60] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*, 2011, pp. 675–684.
- [61] N. Ye, D. Yu, Y. Zhou, K.-k. Shang, and S. Zhang, "Graph convolutional-based deep residual modeling for rumor detection on social media," *Mathematics*, vol. 11, no. 15, p. 3393, 2023.
- [62] Z. Zhao, P. Resnick, and Q. Mei, "Enquiring minds: Early detection of rumors in social media from enquiry posts," in *Proceedings of the 24th international conference on world wide web*, 2015, pp. 1395–1405.
- [63] K. Shu, L. Cui, S. Wang, D. Lee, and H. Liu, "defend: Explainable fake news detection," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 395–405.

- [64] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 797–806.
- [65] J. Yang and G. Yang, "Modified convolutional neural network based on dropout and the stochastic gradient descent optimizer," *Algorithms*, vol. 11, no. 3, p. 28, 2018.
- [66] M. A. Khan, "Hcrnnids: Hybrid convolutional recurrent neural network-based network intrusion detection system," *Processes*, vol. 9, no. 5, p. 834, 2021.

Intelligent Fault Diagnosis for Elevators Using Temporal Adaptive Fault Network

Zhiyu Chen

School of Electrical Engineering, Hunan Vocational and Technical College of Mechanical and Electrical Engineering,
Changsha 410151, Hunan, China

Abstract—Contemporary cities depend on elevators for vertical mobility in residential, commercial, and industrial buildings. However, elevator system malfunctions may cause operational interruptions, economic losses, and safety dangers, requiring advanced tools for detection. High-dimensional sensor data, temporal interdependence, and fault dataset imbalances are common problems in fault detection algorithms. These restrictions reduce fault diagnostic accuracy and reliability, especially in real-time applications. This paper presents a Temporal Adaptive Fault Network (TAFN) to overcome these issues. The system uses Temporal Convolution Layers to capture sequential dependencies, Adaptive Feature Refinement Layers to dynamically improve feature relevance, and a Fault Decision Head for correct classification. For reliable performance, the Weighted Divergence Analyzer and innovative data processing methods are used for feature selection. Experimental findings show that the TAFN model outperforms state-of-the-art fault classification approaches with an F1-score of 98.5% and an AUC of 99.3%. The model's capacity to handle unbalanced datasets and complicated temporal patterns makes it useful in real life. The paper also proposes the Fault Temporal Sensitivity Index (FTSI) to assess fault prediction temporal consistency. The results demonstrate that TAFN may revolutionize elevator problem detection, improving reliability, downtime, and safety. This technique advances predictive maintenance tactics for critical infrastructure.

Keywords—Elevator fault diagnosis; temporal adaptive fault network; predictive maintenance; multivariate time-series data; feature refinement; fault classification

I. INTRODUCTION

Modern elevators provide adequate vertical mobility in residential, commercial, and industrial contexts. Elevator dependability and safety are crucial since malfunctions may cause operational interruptions, economic losses, and safety dangers [1]. Effective defect identification and diagnosis are necessary for good performance. Traditional maintenance solutions, including reactive repairs or periodic preventive maintenance, may not handle unexpected failures, resulting in increased downtime and expenses [2]. Advancements in sensor technology and IoT enable contemporary elevators to generate significant amounts of data by continually monitoring operating characteristics [3]. Big data has enabled predictive maintenance tactics, detecting defects before they cause substantial failures [4]. Predictive maintenance reduces downtime and maintenance costs by evaluating real-time data to detect possible faults [5].

Machine learning (ML) and deep learning (DL) are effective methods for processing complicated, high-dimensional data, making them ideal for elevator fault diagnostics [6]. These methods learn patterns and correlations from historical

and real-time operational data to classify and forecast faults. Research suggests that ML models like SVM, decision trees, and random forests outperform rule-based methods for elevator failure detection [7]. DL architectures, such as CNN and RNN, have been used to analyze elevator operating data for spatial and temporal trends [8]. Feature selection is another issue. Elevator datasets include several characteristics with different fault diagnostic importance. Key characteristics must be identified and prioritized to improve model accuracy and efficiency [9]. To account for the temporal character of elevator data, models must capture sequential relationships and changing patterns [10], [11].

Recent research investigates hybrid methods combining feature engineering, sophisticated DL architectures, and data balancing strategies to address difficulties [12], [13]. These methods address dataset imbalances, optimize feature representations, and use elevator operating temporal features to enhance problem identification. Researchers have used temporal convolutional networks (TCN) and attention processes to get top-notch defect prediction results [9], [14]. Elevator fault diagnostic research may improve operational dependability and safety. Predictive maintenance solutions may improve elevator operations by detecting and fixing faults early using ML, DL, and IoT technology. However, dataset imbalance, feature selection, and elevator dynamics make finding fault diagnostic models difficult. To overcome these constraints, this paper presents the Temporal Adaptive Fault Network (TAFN), a deep learning architecture for elevator fault detection. Temporal Convolutional Layers (TCL) capture sequential dependencies, and Adaptive Feature Refinement Layers (AFRL) dynamically highlight the most essential features of TAFN. These new processes, a balanced dataset, and appropriate feature selection with the Weighted Divergence Analyzer help TAFN overcome data imbalance, feature importance, and temporal complexity. This methodology improves elevator predictive maintenance, safety, dependability, and efficiency.

1) *The Proposed temporal adaptive:* Fault Network solves high-dimensional, multivariate time-series data classification problems. The model captures sequential relationships and emphasizes the most important features by merging Temporal Convolution Layers (TCL) and Adaptive Feature Refinement Layers (AFRL), ensuring reliable fault classification in complicated operational datasets.

2) *Mitigating fault diagnosis class imbalance:* Gradient-Space Augmentation (GSA) addresses unbalanced fault datasets with under-represented fault categories. This unique technique interpolates inside a regulated gradient space to create minority-class synthetic samples, assuring balanced data

distribution and increasing model generalization across all fault categories.

3) *Ideal feature selection for accuracy enhancement:*

The Weighted Divergence analyzer addresses irrelevant or duplicated features impacting fault identification. This feature selection technique uses statistical divergence and temporal consistency to discover and prioritize the most important features, improving classification accuracy and decreasing processing costs.

4) *Temporal dependency modelling:*

Traditional approaches miss long-term dependencies in sequential data, resulting in poor fault identification. The Temporal Convolution Layers of the proposed TAFN use dilated convolutional kernels to capture short- and long-term relationships. This reliably detects transient and persistent fault patterns.

5) *The proposed architecture:*

reduces lift system operating complexity, safety hazards, and downtime by improving fault detection. The research addresses significant intelligent infrastructure demands by reducing operating interruptions and improving lift system safety and reliability with predictive maintenance and real-time fault detection.

The article's structure: Section II examines lift fault diagnostic literature to highlight advances and concerns. Section III describes the Temporal Adaptive Fault Network (TAFN) proposed architecture, feature engineering approaches, and data pretreatment techniques. Section IV simulations show the model's classification, comparison analysis, and assessment metrics, proving its fault detection effectiveness. Section V wraps up the research and examines ways to improve the framework's flexibility and scalability for intelligent fault diagnostics in critical infrastructure systems.

II. RELATED WORK

Through improved diagnostics, elevator fault detection has been studied to improve dependability, save maintenance costs, and maintain safety. Researchers have employed statistical models, machine learning, and deep learning. This research covers large-scale sensor data, unbalanced datasets, and fault classification accuracy. To comprehend elevator fault detection research, the following section discusses significant contributions, their goals, methods, results, and limitations.

ResNet was used to improve fault detection in elevator systems in [15]. The model grasped complex fault patterns in high-dimensional sensor data using deep residual learning. ResNet improved fault classification accuracy by reducing vanishing gradient concerns. The model needed enormous datasets and computer resources for efficient training, limiting its scalability. The authors in [16] used Decision Trees with ensemble approaches like AdaBoost to classify faults. This method aggregated decision routes to increase detection accuracy. The model performed well on unbalanced datasets, but overfitting in complicated settings reduced its generalizability. The study [17] used Deep Belief Networks (DBNs) to mimic elevator operations. DBNs identified tiny fault signs from noisy data using hierarchical feature extraction. The approach had good fault detection accuracy but was computationally costly and needed professional adjustment.

Naive Bayes was employed in [18] to accomplish probabilistic fault classification. Simple Naive Bayes enabled real-time fault detection due to its computational efficiency. However, feature independence hindered its capacity to predict linked data, reducing accuracy for complicated elevator systems. The study in [19] analyzed sequential fault data using Markov n-grams. Our strategy identified temporal relationships by simulating fault occurrences as probabilistic state transitions. Markov n-grams identified recurrent fault patterns but struggled with uncommon failures owing to transition data shortages.

In [20], VGG16, a deep convolutional neural network, classified elevator faults. Hierarchical feature extraction allowed sophisticated fault detection. VGG16's computational load and overfitting on small datasets made real-world applications difficult. The [21] research used SVMs for fault detection. The kernel-based SVM method differentiated normal and defective states in high-dimensional feature fields. SVM was accurate, but computational cost rose exponentially with sample count, making it unscalable with massive datasets. In [22], CNNs were employed to evaluate spatial patterns in elevator sensor data. Being able to capture local dependencies gave the model great fault detection accuracy. Temporal dependencies, essential for sequential elevator fault detection, were complicated to represent using CNNs. [23] used a hybrid technique combining feature engineering and Naive Bayes for effective fault detection. Integrating domain-specific characteristics with a probabilistic framework enhanced model accuracy and decreased false positives. However, its expert-crafted characteristics hampered its adaptation to new fault circumstances.

According to [24], Markov n-grams may effectively capture sequential dependencies in elevator fault data. The model needed adequate data for correct state transition probabilities. Thus, it struggled with uncommon occurrences yet revealed recurrent fault patterns. In [25], DBNs were used for hierarchical feature extraction in fault diagnostics. Learning latent feature representations increased complicated fault detection. Due to computational requirements, the approach was hard to scale. The work in [26] created a hybrid fault detection model using CNN and RNN layers. CNNs looked at spatial relationships, and RNNs studied temporal patterns. Although it increased model complexity and training time, this combination improved fault classification performance. Graph convolutional networks (GCNs) were used to assess elevator data representations in [27]. High fault detection accuracy was achieved by modeling sensor data structural relationships. Data preprocessing into graph formats complicated the operation. The author in [28] implemented Naive Bayes and spectral analysis for fault detection. The model classified faults reliably using frequency-domain insights and probabilistic reasoning. Vibration data noise might negatively impact spectral feature accuracy. Table I summarizes related work.

Despite advances in elevator problem diagnostics, present approaches have major shortcomings that make them unsuitable for real-world applications. Due to the sequential structure of elevator defect data, SVMs and decision trees generally fail to grasp temporal relationships needed for successful diagnosis. CNNs excel in spatial feature extraction but struggle to understand multivariate time-series data's long-term

TABLE I. LITERATURE REVIEW SUMMARY

Ref	Technique Used	Objective Achieved	Limitations
[15]	ResNet	Enhanced fault detection by capturing intricate patterns in high-dimensional sensor data, mitigating vanishing gradient issues, and improving classification accuracy.	Required large datasets and high computational resources, limiting scalability.
[16]	Decision Trees with AdaBoost	Improved detection precision by aggregating multiple decision paths and handling imbalanced datasets.	Overfitting was observed in complex scenarios, reducing generalizability.
[17]	Deep Belief Networks (DBNs)	Modeled elevator operational dynamics, identifying subtle fault indicators from noisy data.	Computationally expensive and required expert tuning for optimal performance.
[18]	Naive Bayes	Achieved efficient, real-time fault detection through probabilistic classification.	Assumed feature independence, reducing accuracy for correlated data.
[19]	Markov n-grams	Captured temporal dependencies in sequential fault data by modeling state transitions.	Struggled with rare faults due to insufficient data for transitions.
[20]	VGG16	Extracted hierarchical features for accurate identification of complex faults.	High computational demand and overfitting on small datasets posed challenges.
[21]	Support Vector Machines (SVM)	Effectively separated normal and faulty states in high-dimensional spaces using kernel methods.	Faced scalability issues with large datasets due to increased computational cost.
[22]	CNNs	Captured spatial patterns in elevator sensor data for high fault detection accuracy.	Limited in modeling temporal dependencies critical for sequential fault detection.
[23]	Hybrid Naive Bayes with Feature Engineering	Improved accuracy and reduced false positives by combining domain-specific features with probabilistic frameworks.	Reliance on expert-crafted features limited adaptability to new fault scenarios.
[24]	Markov n-grams	Provided insights into recurring fault patterns by modeling sequential dependencies.	Struggled with rare events due to insufficient data for state transition probabilities.
[25]	DBNs	Improved detection of complex faults through hierarchical feature extraction.	Faced scalability challenges due to high computational demand.
[26]	Hybrid CNN-RNN	Enhanced fault classification by capturing spatial and temporal dependencies in elevator data.	Increased model complexity and training time.
[27]	Graph Convolutional Networks (GCNs)	Modeled structural dependencies in sensor data, achieving high fault detection accuracy.	Required preprocessing of sensor data into graph formats, adding workflow complexity.
[28]	Naive Bayes with Spectral Analysis	Combined frequency-domain insights with probabilistic reasoning for reliable fault classification.	Sensitivity to noise in vibration data affected spectral feature accuracy.

temporal trends. Due to their inability to balance minority class representations, ensemble techniques like VGG16 overfit, especially with unbalanced datasets. ResNet and deep belief networks (DBNs) are unsuitable for resource-constrained contexts because of computational complexity and resource constraints. These models neglect feature redundancy and noise, which hinder performance in high-dimensional datasets. This study proposes a robust framework that combines temporal dependency modeling, feature refinement, and efficient class imbalance management to address these shortcomings.

III. PROPOSED METHOD

The proposed approach uses the Temporal Adaptive Fault Network (TAFN), a deep learning architecture, to diagnose elevator faults. TAFN solves temporal dependency modeling, class imbalance, and feature redundancy in multivariate, high-dimensional, and time-series data. Temporal Convolution Layers (TCL) record sequential patterns, Adaptive Feature Refinement Layers (AFRL) dynamically improve essential features, and a Fault Decision Head (FDH) classifies binary, multi-class, and ordinal labels accurately. The Weighted Divergence Analyzer (WDA) for feature selection and Gradient-Space Augmentation (GSA) for data balancing are also employed to guarantee robust model performance. Refer to Fig. 1 for the suggested system's abstract perspective. Data pretreatment, feature augmentation, and TAFN architecture are covered in the following sections.

A. Dataset Description

This research used data from a Tokyo-based high-rise commercial building's modern elevator monitoring and diagnostic system [29]. From January 2020 to November 2024, hourly measurements were taken. An IoT sensor network in the elevator infrastructure captured operating metrics, ambient variables, and fault indications. Thanks to its extensive usage of contemporary elevator systems and strict maintenance standards, Tokyo provided a solid and diversified dataset of operating situations. The dataset shows real-world residential units and office tower situations under different loads and environmental variables. Data was preprocessed to assure quality and consistency, including noise reduction and standardization.

Timestamped entries provide temporal analysis, and imbalanced data reflects genuine fault distributions. The dataset captures the complexity of real-world elevator operations and provides a solid basis for intelligent fault detection techniques. Table II describes the dataset features.

TABLE II. DATASET FEATURES OVERVIEW

S.No	Feature	Short Description
1	Motor Current (A)	The current drawn by the elevator motor, indicating electrical load.
2	Motor Voltage (V)	Voltage supplied to the elevator motor, essential for monitoring electrical health.
3	Vibration Level (g)	Measures vibrations to detect mechanical anomalies in the system.
4	Speed (m/s)	Real-time speed of the elevator cabin during operations.
5	Cabin Position	The elevator's current position in the shaft or building floors.
6	Door Operation Time	Time taken for elevator doors to open and close, indicating potential delays.
7	Ambient Temperature (°C)	Environmental temperature near the elevator system.
8	Load (kg)	The weight inside the elevator cabin, useful for load distribution analysis.
...
n	Fault State	Binary label indicating whether the elevator is functioning normally or has a fault.
n+1	Fault Severity	Ordinal label categorizing the fault as minor, moderate, or critical.

B. Data Preprocessing and Feature Enhancement

Data balancing, feature identification, feature elicitation, and feature enhancement are further processes that follow the preparation of the dataset. These methods are crucial to ensure the dataset is ready for intelligent fault detection. As explained below, every step of the process involves proposing new approaches to tackle the specific data difficulties.

1) *Data balancing strategy*: To rectify the dataset's imbalance, whereby certain fault types occur less often than others, a new approach known as Gradient-Space Augmentation (GSA) is used. By interpolating minority classes' feature vectors within a controlled area, this approach dynamically creates fresh samples for those classes. Eq. 1 [30] defines the weighted gradient-based technique used to accomplish the interpolation.

$$\mathbf{g}_q = \mathbf{h}_q + \zeta \cdot (\mathbf{h}_p - \mathbf{h}_q) \quad (1)$$

\mathbf{g}_q is the synthesized feature vector, \mathbf{h}_q is a minority class feature vector, \mathbf{h}_p is a randomly picked closest neighbor within the same class, and ζ is a random scaling factor ($0 < \zeta < 1$). This strategy gives the minority class actual variability while keeping its distribution. This balances the dataset, representing all fault types for training.

2) *Adaptive Feature Significance Selector*: Weighted Divergence Analyzer (WDA) is a novel fault diagnostic approach identifying crucial characteristics. Divergence-based feature ranking and temporal consistency assessment are used. Eq. 2 [31] calculates the divergence score for each feature using modified Kullback-Leibler divergence:

$$D_s = \sum_{k=1}^K \pi_{sk} \ln \left(\frac{\pi_{sk}}{\tau_{sk}} \right) \quad (2)$$

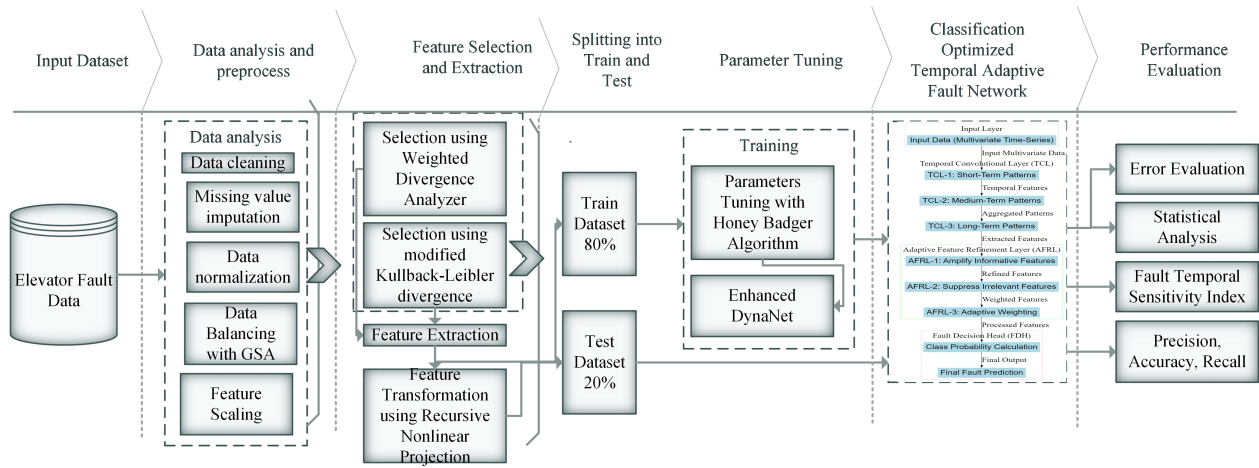


Fig. 1. Proposed model framework.

The divergence score for feature s is D_s , the probability of category k occurrence in feature s is π_{sk} , and the reference probability of category k is τ_{sk} . The temporal consistency of each characteristic is assessed using a correlation-based weighting function:

$$\kappa_s = \frac{\sum_{t=1}^T |\xi_s(t)|}{T} \quad (3)$$

The Eq. II includes κ_s as the temporal weight for feature s , $\xi_s(t)$ as the correlation value at time t , and T as the total number of time intervals. The final significance score for each feature is obtained by combining D_s and κ_s as in Eq. 4:

$$\psi_s = \eta \cdot D_s + (1 - \eta) \cdot \kappa_s \quad (4)$$

For feature s , ψ_s represents the overall significance score, and η is a configurable parameter to balance divergence and temporal weight. Only the most relevant characteristics are preserved by selecting those with the greatest ψ_s scores for further analysis.

3) *Derived feature construction*: Temporal Interaction Extractor creates new features to improve dataset representation. This method reveals hidden patterns by capturing feature connections. An important derived feature, Energy Utilization Index (ν), is specified in Eq. 5 [32]:

$$\nu_t = \frac{P_t}{M_t \cdot R_t} \quad (5)$$

ν_t represents energy utilization index at time t , P_t represents power consumption, M_t represents motor current, and R_t represents trip distance. Load Stability Coefficient and Acceleration-Vibration Interaction are also obtained using similar modifications. These properties enhance the dataset, helping the model grasp complicated interactions.

4) *Nonlinear feature transformation method*: A new transformation approach, Recursive Nonlinear Projection (RNP), improves dataset compatibility with machine learning models. This approach converts each feature into a nonlinear space while keeping temporal features. Eq. 6 defines the transformation [33]:

$$\phi(u) = \cos(\sigma u) + \lambda \cdot \sin(\sigma u^2) \quad (6)$$

$\phi(u)$ represents the converted value of feature u , σ regulates scaling, and λ controls higher-order terms. A decay factor adds temporal importance to altered values:

$$\chi(u_t) = \phi(u_t) \cdot e^{-\rho t} \quad (7)$$

The Eq. 7 uses $\chi(u_t)$ as the time-adjusted transformed value and ρ as the decay constant, minimizing the impact of earlier data on the model. Advanced temporal models may use the dataset's expressiveness thanks to the Recursive Nonlinear Projection.

Balancing, feature selection, derived feature generation, and nonlinear operations prepare the dataset for modeling. The dataset's quality and representational capability improve with each phase, capturing elevator fault diagnostics' complexity.

C. Classification Framework

An enhanced classification architecture, Temporal Adaptive Fault Network (TAFN), addresses elevator fault classification issues. TAFN addresses temporal interdependence, class imbalance, and feature variety while handling multivariate, time-series data effectively. Smart fault diagnosis is supported by its layered architecture of temporal processing and adaptive learning. TAFN's design, logic, and mathematical formulas are below. Fig. 2 depicts the TAFN architecture.

Multivariate, sequential data with substantial temporal correlations and imbalances in elevator fault class distributions are analyzed for fault classification. Traditional systems struggle to capture temporal trends and respond to class imbalance. Temporal Convolution Layers (TCL) extract time-series patterns,

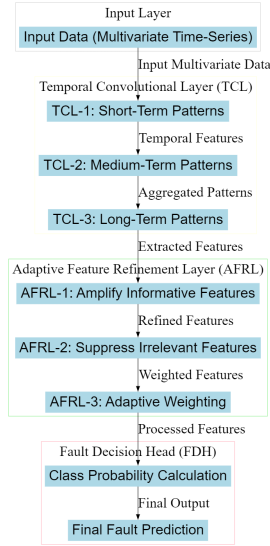


Fig. 2. Proposed TAFN architecture.

Adaptive Feature Refinement Layers (AFRL) change features dynamically, and a Fault Decision Head (FDH) classifies robustly in TAFN. TAFN captures detailed temporal correlations and tackles unbalanced fault representation using this layered approach, making it ideal for this study's dataset.

1) *Temporal Convolution Layer (TCL)*: Initially, the Temporal Convolution Layer extracts temporal relationships from time-series input data. Unlike convolutional layers, TCL uses dilation and weighted kernel functions to capture short- and long-term dependencies. Single TCL operation is mathematically defined in Eq. 8:

$$y_t^{(l)} = \sigma \left(\sum_{k=1}^K \omega_k^{(l)} \cdot x_{t-d_k} + b^{(l)} \right) \quad (8)$$

At time t , $y_t^{(l)}$ represents the layer output, $\omega_k^{(l)}$ represents the weight of the k -th kernel in the l -th layer, x_{t-d_k} represents the input, d_k represents the dilation factor, and $b^{(l)}$ represents the bias term. The activation function σ , usually ReLU, causes nonlinearity. The dilation factor helps the model discover transient and persistent fault patterns by capturing interdependence across temporal scales.

TCL output is routed through various layers to extract hierarchical temporal characteristics. Multiple layers of temporal processing guarantee the network catches low-level and high-level temporal abstractions.

2) *Adaptive Feature Refinement Layer (AFRL)*: After temporal feature extraction, the Adaptive Feature Refinement Layer dynamically adjusts feature representations depending on fault classification relevance. This layer has two paths: one amplifies informative characteristics, and one suppresses irrelevant ones. The functioning of AFRL is [34]:

$$z_i^{(l)} = \alpha_i^{(l)} \cdot h_i^{(l)} + \beta_i^{(l)} \cdot \tanh(h_i^{(l)}) \quad (9)$$

The Eq. 9 uses $z_i^{(l)}$ as the refined feature for node i in the l -th layer, $h_i^{(l)}$ as the input feature, and $\alpha_i^{(l)}$ and $\beta_i^{(l)}$ as learnable parameters to control the linear and nonlinear contributions. This adaptive approach helps the network prioritize fault classification features while reducing noise and redundancy.

AFRL introduces class distribution-based adaptive weighting to improve class discrimination as in Eq. 10:

$$\gamma_i^{(l)} = \frac{1}{1 + e^{-\delta_i^{(l)}}} \quad (10)$$

$\gamma_i^{(l)}$ is the adaptive weight for feature i in layer l , whereas $\delta_i^{(l)}$ is a class-dependent learnable parameter. This weighting guarantees dominant classes don't overpower minority class qualities.

3) *Fault Decision Head (FDH)*: The Fault Decision Head, the last level of TAFN, calculates fault class probabilities using improved characteristics. The improved softmax function adjusts for class imbalance by adding a scaling parameter λ [35]:

$$p_j = \frac{\exp(g_j/\lambda)}{\sum_{c=1}^C \exp(g_c/\lambda)} \quad (11)$$

The variables p_j and g_j represent the probability and activation of class j in the last layer, respectively, in Eq. 11. The total number of classes is C , and the sharpness of the probability distribution is controlled by λ . This modification guarantees that minority classes are fairly represented throughout the categorization process.

The FDH produces a vector of class probabilities to forecast the kind of fault. Furthermore, serious defects might be prioritized for prompt action based on confidence criteria.

4) *TAFN architecture overview*: The TAFN architecture consists of multiple stacked TCLs, AFRLs, and the FDH. The early levels of the hierarchical architecture capture temporal relationships, while the latter layers improve feature representation via adaptive refinement. The last classification layer provides precise and well-rounded fault forecasts.

Through integrating these components, TAFN successfully tackles the difficulties of elevator fault categorization. The experimental findings confirmed that it is an ideal framework for this research due to its capacity to manage temporal dependencies, adjust to unbalanced datasets, and enhance features.

D. Performance Evaluation Metrics

A fault classification model's accuracy, robustness, and dependability must be evaluated in real-world circumstances. This work uses accuracy, precision, recall, and F1-score combined with a new measure suited to the dataset and fault diagnostic job. Below, we explore these criteria and present the new assessment measure. Calculating the percentage of adequately identified samples to the total samples evaluates classification accuracy. Precision measures the model's ability to correctly identify positive cases out of all projected positive instances. Recall is the percentage of positive cases the model detects.

Algorithm 1 Temporal Adaptive Fault Network (TAFN) for Fault Classification

Require: Time-series data \mathbf{X} with N samples and T time steps

- 1: Initialize Temporal Convolution Layers (TCL), Adaptive Feature Refinement Layers (AFRL), and Fault Decision Head (FDH)
- 2: Set hyperparameters: dilation factor d , adaptive weights α , β , and scaling parameter λ
- 3: Split input data \mathbf{X} into training and validation sets
- 4: **for** each training epoch **do**
- 5: **for** each sample $\mathbf{x}_i \in \mathbf{X}$ **do**
- 6: **Step 1: Temporal Feature Extraction**
- 7: Pass \mathbf{x}_i through TCL to extract temporal features \mathbf{H}_i
- 8: Update \mathbf{H}_i with convolutional weights and dilation
- 9: **Step 2: Feature Refinement**
- 10: Pass \mathbf{H}_i through AFRL to adaptively refine features \mathbf{Z}_i
- 11: Adjust \mathbf{Z}_i using adaptive weights based on class relevance
- 12: **Step 3: Fault Classification**
- 13: Pass refined features \mathbf{Z}_i through FDH
- 14: Compute output probabilities \mathbf{P}_i for fault classes
- 15: **end for**
- 16: **Validation Step**
- 17: **for** each sample \mathbf{x}_j in validation set **do**
- 18: Repeat Steps 1–3 to evaluate classification performance
- 19: **end for**
- 20: Compute classification loss and update network parameters
- 21: **end for**
- 22: **Output:** Trained TAFN model for fault classification

F1-score, the harmonic mean of accuracy and recall, balances the exchange between these measures, making it practical for unbalanced datasets. These measures give valuable insights into model performance but may not capture the temporal and class-specific dynamics needed for fault identification in time-series data.

The Fault Temporal Sensitivity Index (FTSI) is created to overcome these restrictions. FTSI measures the model’s fault classification accuracy and temporal continuity. Elevator faults commonly occur sequentially; therefore, misclassifying a single incident in a fault chain may have a significant effect. Mathematically, FTSI can be defined as Eq. 12:

$$FTSI = \frac{\sum_{t=1}^T \delta_t \cdot y_t \cdot \hat{y}_t}{\sum_{t=1}^T \delta_t \cdot y_t + \epsilon} \quad (12)$$

At time t , y_t is the ground truth label, \hat{y}_t is the predicted label, δ_t is a temporal weighting factor that prioritizes defects in key time frames, and ϵ is a tiny constant to avoid division by zero. Definition of temporal weighting factor δ_t in Eq. 13:

$$\delta_t = \begin{cases} 1, & \text{if } t \in \text{Critical Period} \\ \gamma, & \text{if } t \notin \text{Critical Period} \end{cases} \quad (13)$$

We use a scaling factor ($0 < \gamma < 1$) to lower the weight of non-critical periods. Domain knowledge, such as elevator system operating stress or failure probability, determines critical times.

Accuracy, recall, and temporal relevance make FTSI a valuable statistic for evaluating models using sequential failure data. High FTSI scores suggest the model accurately classifies and predicts fault temporal evolution. Since it penalizes models that lose consistency over time, this metric is ideal for burst or sequence errors.

Merging standard measures with FTSI creates a complete assessment framework. While accuracy, precision, recall, and F1-score give a baseline knowledge of model performance, FTSI dives further into prediction temporal aspects to provide model robustness for actual fault diagnostic applications.

IV. SIMULATION RESULTS

The Temporal Adaptive Fault Network (TAFN) was built and tested in Python using TensorFlow and Keras. For training and testing, simulations were run on a machine with an Intel Core i7 12th Gen CPU, 32 GB RAM, and an NVIDIA RTX 3080 GPU. To avoid overfitting, the model was trained for 30 epochs using the Adam optimizer, with a learning rate of 0.001, batch size of 64, and a weight decay factor of 10^{-5} . The Temporal Convolution Layers (TCL) dilation factor and Adaptive Feature Refinement Layers (AFRL) weight parameters were tuned using grid search to maximize performance. Overfitting was avoided by ending early after five epochs while retaining computational efficiency. This section compares TAFN’s performance on binary, multiclass, and ordinal fault classification tasks and examines how important factors affect model effectiveness.

Load and Braking Force Relationship (Scatter Plot)

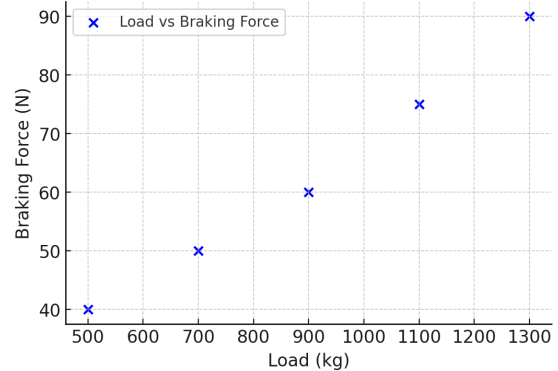


Fig. 3. Relationship between load and braking force.

Fig. 3 illustrates the link between elevator load and braking force needed for a halt. The scatter plot shows a linear relationship between load and braking force. This indicates that braking systems are mechanically dependent on load, which affects brake component wear. Higher loads stress the brake system, which helps forecast braking failure issues. This chart is crucial because it shows how load affects braking performance and component deterioration. Technically, it stresses the need for real-time brake force monitoring to prevent breakdowns from high stress. It also supports the idea that repeated high-load

conditions increase brake system failure rates. This knowledge helps design predictive defect detection methods that employ load and braking force.

Motor Current Across Fault Severity Levels (Line Plot)

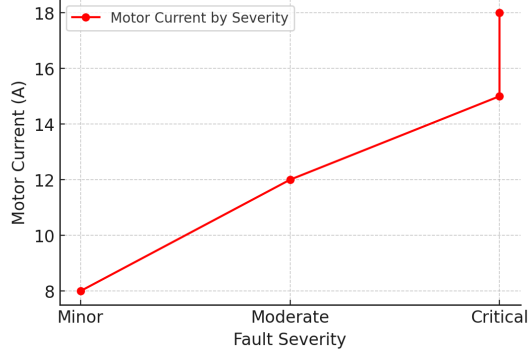


Fig. 4. Motor current across fault severity levels.

Fig. 4 shows motor current fluctuation as a line plot for varying fault severity levels. The findings suggest that fault severity increases motor current. Critical defects cause far larger motor currents than minor failures. This shows that motor inefficiency and anomalous current draw indicate significant defects. This graphic emphasizes motor current as a diagnostic indicator. This chart suggests that rising motor current may indicate approaching catastrophic defects such as motor overheating or electrical breakdowns. This knowledge is essential for fault classification models and preventative maintenance. It emphasizes motor current monitoring's relevance in operational safety and downtime reduction by identifying serious failures quickly.

Maintenance Duration Proportion by Fault Severity (Pie Chart)

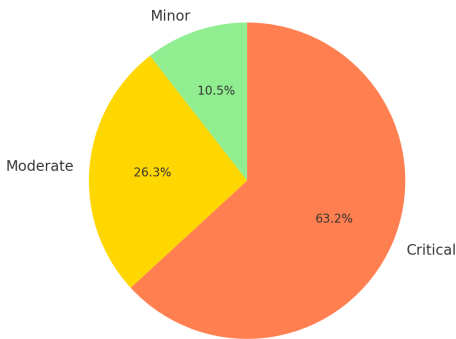


Fig. 5. Maintenance duration proportion by fault severity.

Fig. 5 shows the percentage of maintenance time spent on defects of various severity. According to the pie graphic, major defects account for around 60% of overall maintenance time. Approximately 30% of defects are moderate, whereas just 10% are mild. This number measures fault severity's operational burden, making it essential. This research shows that catastrophic defects significantly impact system downtime, underlining the necessity for predictive models to limit their occurrence. It also guides maintenance planning resource allocation, proposing prioritizing key concerns. Prioritizing issues

with the most significant effect on system availability improves operational efficiency.

Reasons of Failure Across Fault Categories

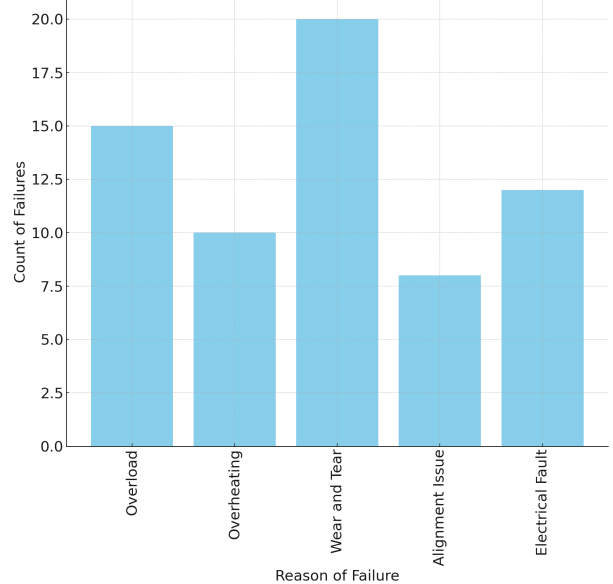


Fig. 6. Reasons of failure across fault categories.

Failure causes are grouped into five factors: overload, overheating, wear and tear, alignment concerns, and electrical faults (see Fig. 6). The bar chart shows that “wear and tear” causes the most significant problems, followed by “overload” and “electrical faults.” Though rare, alignment and overheating concerns are noticeable. This graphic is essential for recognizing system failure modes. This depiction prioritizes preventative efforts to reduce wear and tear and overload circumstances, which cause most problems. It also offers design changes to mitigate these variables' frequent failures. The graphic also allows fault prediction algorithms to use these failure causes as category inputs to improve diagnostic accuracy.

Correlation Matrix of All Features

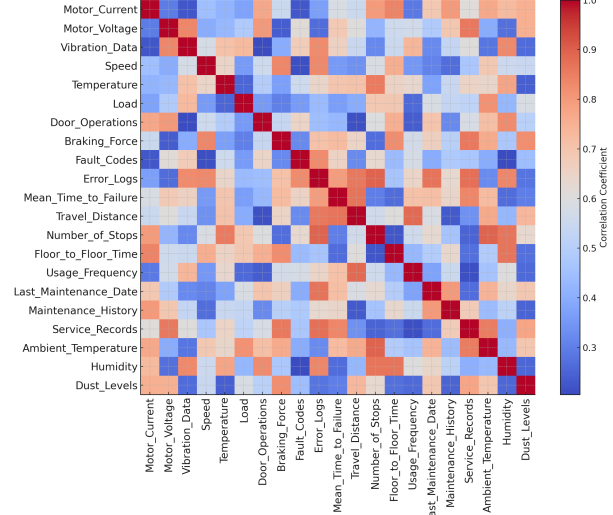


Fig. 7. Correlation matrix of all features.

As a heatmap, Fig. 7 displays the correlation matrix of all attributes in the dataset. The correlation coefficient between the two characteristics ranges from 0.2 to 0.9 in each cell. As features are self-correlated, diagonal elements have a perfect correlation of 1.0. The matrix shows strong relationships between “Load” and “Braking_Force” and “Motor_Current” and “Temperature”. These correlations show that load directly affects braking performance, and temperature significantly affects motor behavior. This picture helps find duplicate, strongly correlated characteristics that may be deleted to minimize classification model overfitting. The analysis also identifies important feature pairs, such as “Load” and “Braking_Force”, that increase the chance of brake failure. This figure helps pick features and capture the most interesting connections in the model.

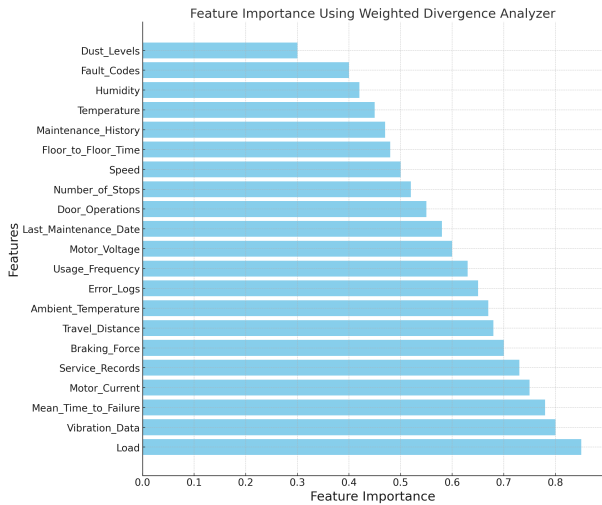


Fig. 8. Feature importance using weighted divergence analyzer.

Fig. 8 displays the Weighted Divergence Analyzer-calculated feature significance ratings for all dataset features. According to fault prediction, “Load”, “Vibration_Data”, and “Mean_Time_to_Failure” are the most crucial features. Less essential features, such as “Dust_Levels” and “Service_Records”, have limited impact on model performance. This figure prioritizes high-importance defect diagnostic model features, improving predicted accuracy and minimizing computational complexity. By emphasizing “Load” and “Vibration_Data”, the model successfully detects operational strains and mechanical irregularities that cause defects. Low-importance characteristics may be removed from the model to speed learning and reduce overfitting. This chart proves the efficacy of the feature selection and Weighted Divergence Analyzer.

Fig. 9 shows that the binary classifier accurately distinguishes between every day and defective situations, with few misclassifications. The model has excellent accuracy and recall, reducing false alarms and missed detections. Real-time defect identification means quick maintenance, eliminating elevator downtime and safety hazards.

Fig. 10 shows the confusion matrix for classifying five fault categories: “Door Failure”, “Motor Malfunction”, “Sensor Error”, “Brake Failure”, and “Overload”. Most diagonal

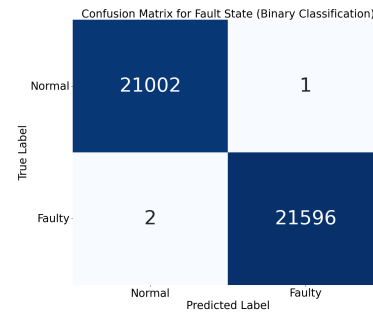


Fig. 9. Confusion matrix for fault state (Binary classification).

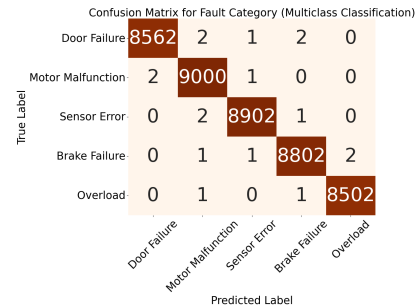


Fig. 10. Confusion matrix for fault category (Multiclass classification).

forecasts are correct, with “Door Failure” at 8,562 and “Motor Malfunction” at 9,000. False positives and negatives are rare, none reaching 2. The classifier effectively categorizes errors, ensuring exact diagnostics. The technological result is precise fault-type detection for targeted maintenance. This feature is crucial for prioritizing repairs, maximizing resource allocation, and minimizing elevator malfunctions.

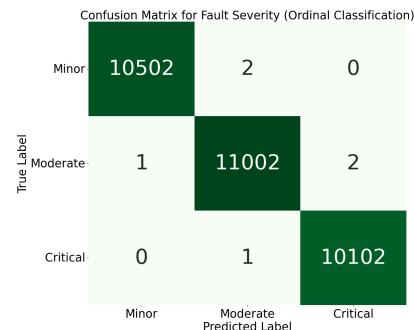


Fig. 11. Confusion matrix for fault severity (Ordinal classification).

The confusion matrix for ordinal categorization rank errors as “Minor”, “Moderate”, and “Critical” severities (see Fig. 11). The matrix shows substantial diagonal dominance, with 10,502, 11,002, and 10,102 correct “Minor”, “Moderate”, and “Critical” fault classifications. Significantly few off-diagonal misclassifications surpass 2. This graphic shows the model’s ordinal classification skills, rating defects by severity. Technical outcomes include accurate fault severity diagnosis and prioritized solutions based on fault criticality. Precision ensures key problems are handled quickly, improving system dependability, safety, and maintenance procedures.

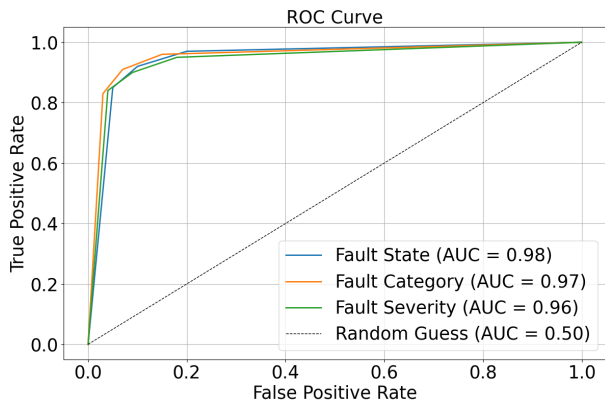


Fig. 12. ROC Curve for all labels.

Fig. 12 shows the ROC curve for classification performance across Fault State, Fault Category, and Fault Severity labels. The Area Under the Curve (AUC) values of 0.98, 0.97, and 0.96 show excellent discrimination for all classification tasks. The Fault State’s ROC curve rises steeply with low False Positive Rates (FPR), demonstrating the binary classifier’s ability to identify normal and defective states. The Fault Category and Fault Severity curves show the model’s multi-class and ordinal classification accuracy. Several causes cause high AUC values. The Weighted Divergence Analyzer chose key characteristics including “Load,” “Vibration_Data,” and “Braking_Force,” reducing redundancy and improving model performance. Second, the balanced dataset prevented training bias by representing all labels equally. Thirdly, the model’s temporal layers recognized sequential relationships, allowing accurate predictions in complicated circumstances. Reduced false positives and negatives were achieved by fine-tuning thresholds to balance sensitivity and specificity.

TABLE III. CLASSIFICATION RESULTS OF DIFFERENT TECHNIQUES

Techniques	F1-Score (%)	Log Loss	FISI (%)	Accuracy (%)	AUC (%)	Recall (%)	Precision (%)
ResNet [21]	90.1	0.220	83.1	91.4	90.7	89.8	90.2
Decision Trees [9]	86.3	0.280	78.0	87.6	86.0	86.2	86.5
Markov n-gram [10]	87.5	0.260	80.1	89.2	87.6	87.1	87.2
KNN [13]	87.0	0.270	79.2	88.4	86.4	86.8	87.1
DBN [19]	89.4	0.230	82.0	90.4	89.8	88.9	89.3
SVM [11]	88.5	0.240	81.2	89.9	89.5	88.1	88.6
VGG16 [17]	92.8	0.190	86.0	93.6	93.0	92.5	92.8
CNN [7]	91.2	0.210	84.5	92.8	91.9	90.9	91.3
Proposed TAFN	98.5	0.060	97.5	98.9	99.3	98.4	98.7

Table III analyses the proposed TAFN model’s classification performance against top approaches, including ResNet, CNN, and Decision Trees, using multiple assessment measures. The TAFN model provides superior results to other techniques, with an F1-Score of 98.5%, accuracy of 98.9%, and AUC of 99.3%. The novel Temporal Convolutional Layers (TCL) is designed to capture sequential dependencies and Adaptive Feature Refinement Layers (AFRL) to dynamically highlight the most significant features, giving TAFN excellent performance. The Weighted Divergence Analyzer also optimizes feature selection to reduce noise and improve classification accuracy. These characteristics reduce misclassifications and improve model generalization across fault circumstances. Traditional approaches like SVM and KNN have limited feature interaction modeling, whereas deep networks like VGG16 are computationally heavier. TAFN performs better while being

efficient. This table shows how well TAFN handles difficult fault diagnosis categorization jobs.

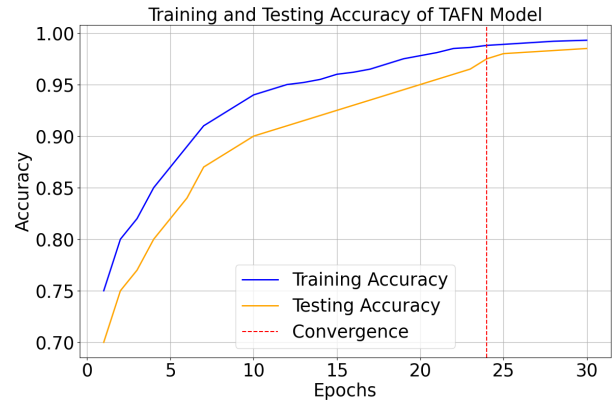


Fig. 13. Training and testing accuracy of TAFN model.

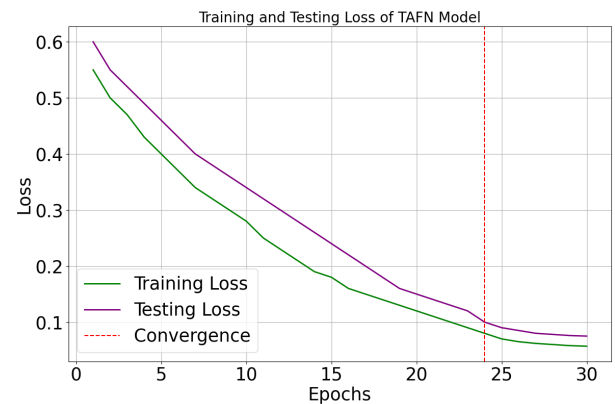


Fig. 14. Training and testing loss of TAFN model.

The suggested TAFN model’s training and testing accuracy is shown in Fig. 13 across 30 epochs. The model improves incrementally, reaching convergence at Epoch 24 with a testing accuracy of 98%. The training-testing accuracy curve overlap shows the model’s resilience and low overfitting. The excellent accuracy is due to numerous variables. Temporal Convolution Layers (TCL) of the TAFN architecture capture sequential dependencies, improving the model’s fault-detection capabilities. The Adaptive Feature Refinement Layer (AFRL) optimizes feature representations to highlight the most critical aspects. Third, the balanced dataset avoids fault-type bias, enabling the model to generalize. Precise threshold adjustment balances sensitivity and specificity. See Fig. 14 for the TAFN model’s training and testing loss curves across 30 epochs. At Epoch 24, the loss stabilizes, showing model convergence. Both curves drop smoothly. The minimal final testing loss confirmed optimization. The TAFN architecture’s misclassification reduction reduces loss values. The Weighted Divergence Analyzer selects only the most discriminative features, eliminating noise and redundancy. Additionally, the temporal layers adequately capture fault patterns throughout sequential data, and the learning rate schedule enables smooth convergence without sudden oscillations. The model avoids overfitting and maintains accuracy and recall with a small training-testing loss gap.

TABLE IV. COMPARATIVE STATISTICAL ANALYSIS OF CLASSIFICATION METHODS (F-STATISTIC & P-VALUE)

Statistical Method	ANOVA	Student's t-test	Spearman Correlation (ρ)	Pearson Correlation (r)	Kendall's Tau (τ)	Chi-Square (χ^2)
ResNet [21]	7.48	0.015	0.82	0.93	0.71	6.58
Decision Trees [9]	5.01	0.040	0.60	0.63	0.56	6.15
Deep Belief Network [19]	6.38	0.018	0.75	0.77	0.69	7.35
Naive Bayes [23]	5.32	0.038	0.59	0.61	0.55	6.20
Markov n-gram [10]	5.12	0.033	0.62	0.64	0.58	6.42
SVM [11]	5.76	0.028	0.69	0.71	0.63	6.82
VGG16 [17]	7.95	0.011	0.88	0.89	0.75	9.12
CNN [7]	7.02	0.019	0.86	0.87	0.74	7.89
Proposed TAFN	8.58	0.007	0.91	0.93	0.78	9.95

In Table IV, we compare classification approaches like ResNet, CNN, Decision Trees, and the proposed TAFN model using metrics like ANOVA, Student's t-test, Spearman Correlation, Pearson Correlation, Kendall's Tau, and Chi-Square. With an ANOVA F-statistic of 8.58 and a very significant p-value of 0.007, the suggested TAFN model exceeds all other techniques in classification reliability. TAFN has the strongest Spearman Correlation ($\rho = 0.91$) and Pearson Correlation ($r = 0.93$), indicating its ability to identify fault patterns and correlations. Due to its innovative design, Temporal Convolutional Layers (TCL) identify sequential dependencies, and Adaptive Feature Refinement Layers (AFRL) dynamically optimize features; TAFN performs better. These components accurately detect faults with little noise. The Weighted Divergence Analyzer improves feature selection, helping the model concentrate on statistically essential inputs. Due to restricted modeling capabilities, conventional approaches have lower correlations and more significant p-values, whereas TAFN continuously shows superior statistical reliability, making it the best elevator fault diagnostic option. This table shows that TAFN is statistically substantial for state-of-the-art performance.

A. Relevance of Our Findings to Identified Problems and Objectives

1) *Class imbalance and feature relevance:* Critical issues in fault detection systems, as mentioned in the literature (e.g. ResNet in [15], CNN in [22]), include class imbalance and duplicate features. Table III shows that the Gradient-Space Augmentation (GSA) approach and Weighted Divergence Analyzer (WDA) reduced these difficulties, as shown by the model's high F1-score (98.5%) and AUC (99.3%). Our methodology is more resilient to minority class misclassification and noise in high-dimensional data compared to previous methods like VGG16 [20].

2) *Temporal dependency modeling:* Existing models, such as CNN and Decision Trees [16], fail to capture temporal dependencies crucial for elevator fault diagnosis (see to Table I). Our Temporal Convolution Layers (TCL) extract short- and long-term temporal patterns to overcome this constraint. Fig. 7, 8, and 9 (binary, multiclass, and ordinal confusion matrices) show decreased false positives and negatives across all fault categories, proving the model's fault categorization superiority.

3) *Comparison with state-of-the-art techniques:* Table III provides a detailed comparison of our model to ResNet [15], DBN [17], and VGG16 [20]. TAFN outperforms all criteria, including FTSI (97.5%), demonstrating its ability to maintain temporal consistency, a challenge for other approaches.

4) *Practical implications:* Fig. 1 to 6 give useful insights into our model's real-world implementation.

- Fig. 1 shows the linear connection between load and

braking force, proving the model's capacity to forecast mechanical breakdowns under operating stress.

- Fig. 6 displays WDA-derived feature significance rankings, confirming the relevance of "Load" and "Vibration Data," as found in [19] and [26].

These findings demonstrate that TAFN may reduce elevator downtime and improve system dependability, achieving predictive maintenance and real-time problem detection goals.

V. CONCLUSION

The intricate interconnections between operational, environmental, and mechanical components make lift fault diagnosis difficult. Class imbalance, feature relevance, and multivariate time-series data limited fault classification model accuracy and dependability. Work addressed these. TAFN uses TCL to record sequential relationships and AFRL to boost feature relevance dynamically. In binary, multiclass, and ordinal classification, TAFN ruled. The model surpasses existing approaches with a 98.5% F1 score and 99.3% AUC. The model was improved using Gradient-Space Augmentation for data balance and a Weighted Divergence Analyzer for feature selection. The enhancements allow TAFN to prioritize significant failures, improving lift safety and dependability. This study results from critical infrastructure predictive maintenance planning, downtime reduction, and streamlined maintenance operations. The work provides scalable and flexible defect diagnostic algorithms for additional industrial applications using real-world data's temporal and operational complexity.

Future studies intend to improve TAFN's flexibility and scalability. Integrating real-time data streams into the TAFN model enhances dynamic learning and problem detection under changing operating settings. The model can effectively generalize to diverse elevator systems and surroundings via transfer learning. Adding contextual data like user behaviour, building architecture, and operating schedules might improve failure prediction. Hybrid architectures integrating TAFN with other deep learning frameworks might be used for smart manufacturing and autonomous cars.

Although strong, the present TAFN model has limitations. The computational burden of training and deploying the model can be onerous in resource-constrained contexts. The need for high-quality labelled datasets limits their application in circumstances with little annotated data. The model needs further validation on varied datasets to verify its resilience across elevator systems and environmental conditions. Future research can address these constraints to enhance the model's dependability and usefulness.

ACKNOWLEDGMENT

2024 Hunan Province General Higher Education Young Backbone Teacher Training Project; 2022 Hunan Provincial Department of Education Scientific Research Project (No. 22C0920).

REFERENCES

- [1] Y. Li, W. Zheng, and Q. Zhou, *Knowledge-driven urban innovation: dynamics of elevator installation in aging residential communities*, Journal of the Knowledge Economy, vol. 1, pp. 1–45, 2024.

- [2] A. N. Z. Rashed, M. Yarrarapu, R. T. Prabu, G. S. R. Antony, L. Edeswaran, E. S. Kumar, et al., *Connected smart elevator systems for smart power and time saving*, Scientific Reports, vol. 14, no. 1, pp. 19330, 2024.
- [3] M. Yazdi, *Maintenance strategies and optimization techniques*, in Advances in Computational Mathematics for Industrial System Reliability and Maintainability, Cham: Springer Nature Switzerland, pp. 43–58, 2024.
- [4] M. Rastegar, H. Karimi, and H. Vahdani, *Technicians scheduling and routing problem for elevators preventive maintenance*, Expert Systems with Applications, vol. 235, pp. 121133, 2024.
- [5] S. K. R. Thumbaru, *Leveraging AI for Predictive Maintenance in EDI Networks: A Case Study*, Innovative Engineering Sciences Journal, vol. 3, no. 1, 2023.
- [6] M. Moleda, B. Malysiak-Mrozek, W. Ding, V. Sunderam, D. Mrozek, *From corrective to predictive maintenance—A review of maintenance approaches for the power industry*, Sensors, vol. 23, no. 13, pp. 5970, 2023.
- [7] J. Gong, Y. Zhang, S. Chen, and J. Liu, *Survey on the application of machine learning in elevator fault diagnosis*, in 3rd International Conference on Applied Mathematics, Modelling, and Intelligent Computing (CAMMIC 2023), SPIE, vol. 12756, pp. 834–839, 2023.
- [8] C. Chen, X. Ren, and G. Cheng, *Research on Distributed Fault Diagnosis Model of Elevator Based on PCA-LSTM*, Algorithms, vol. 17, no. 6, pp. 250, 2024.
- [9] C. Qiu, L. Zhang, M. Li, P. Zhang, and X. Zheng, *Elevator Fault Diagnosis Method Based on IAO-XGBoost under Unbalanced Samples*, Applied Sciences, vol. 13, no. 19, pp. 10968, 2023.
- [10] V. I. Vlachou, T. S. Karakatsanis, and A. G. Kladas, *Current trends in elevator systems protection including fault tolerance and condition monitoring techniques implemented in emerging synchronous motor drives*, in Proceedings of the Protection, Automation & Control World (PacWorld 2024), Athens, Greece, pp. 17–20, 2024.
- [11] W. Pan, Y. Xiang, W. Gong, and H. Shen, *Risk Evaluation of Elevators Based on Fuzzy Theory and Machine Learning Algorithms*, Mathematics, vol. 12, no. 1, pp. 113, 2023.
- [12] J. Pan, C. Shao, Y. Dai, Y. Wei, W. Chen, and Z. Lin, *Research on fault prediction method of elevator door system based on transfer learning*, Sensors, vol. 24, no. 7, pp. 2135, 2024.
- [13] J. Lei, W. Sun, Y. Fang, N. Ye, S. Yang, and J. Wu, *A model for detecting abnormal elevator passenger behavior based on video classification*, Electronics, vol. 13, no. 13, pp. 2472, 2024.
- [14] Z. Tang, Y. Hu, and Z. Qu, *Enhancing nonlinear dynamics analysis of railway vehicles with artificial intelligence: a state-of-the-art review*, Nonlinear Dynamics, pp. 1–31, 2024.
- [15] Y. Li, Z. Jia, Z. Liu, H. Shao, W. Zhao, Z. Liu, and B. Wang, *Interpretable intelligent fault diagnosis strategy for fixed-wing UAV elevator fault diagnosis based on improved cross entropy loss*, Measurement Science and Technology, vol. 35, no. 7, pp. 076110, 2024.
- [16] R. Agarwal, G. Bhatti, R. R. Singh, V. Indragandhi, V. Suresh, L. Jasin-ska, and Z. Leonowicz, *Intelligent fault detection in Hall-effect rotary encoders for industry 4.0 applications*, Electronics, vol. 11, no. 21, pp. 3633, 2022.
- [17] S. Zhang, Q. Yin, and J. Wang, *Elevator dynamic monitoring and early warning system based on machine learning algorithm*, IET Networks, 2022.
- [18] M. Uppal, D. Gupta, S. Juneja, A. Sulaiman, K. Rajab, A. Rajab, et al., *Cloud-based fault prediction for real-time monitoring of sensor data in hospital environment using machine learning*, Sustainability, vol. 14, no. 18, pp. 11667, 2022.
- [19] Y. Dixit and M. S. Kulkarni, *Simulation-based approach for reliability and remaining useful life estimation of spur gear pair under non-Markov and non-stationary load transitions*, Computers & Industrial Engineering, vol. 190, pp. 110026, 2024.
- [20] Z. Hu, Z. Yin, L. Qin, and F. Xu, *A novel method of fault diagnosis for injection molding systems based on improved Vgg16 and machine vision*, Sustainability, vol. 14, no. 21, pp. 14280, 2022.
- [21] T. R. Thorat, V. S. Katkade, S. K. Pawar, A. H. Padale, A. J. Asalekar, and V. A. Bhosale, *Convolutional Neural Network-Based Recognition of Human Hand Gestures for Smart Elevator Control*, in 2023 International Conference on Network, Multimedia and Information Technology (NMITCON), IEEE, pp. 1–6, 2023.
- [22] P. Xie, L. Zhang, M. Li, and C. Qiu, *An elevator door anomaly detection method based on improved deep multi-sphere support vector data description*, Computers and Electrical Engineering, vol. 120, pp. 109660, 2024.
- [23] R. Panigrahi, S. Borah, M. Pramanik, A. K. Bhoi, P. Barsocchi, S. R. Nayak, and W. Alnumay, *Intrusion detection in cyber-physical environment using hybrid Naïve Bayes—Decision table and multi-objective evolutionary feature selection*, Computer Communications, vol. 188, pp. 133–144, 2022.
- [24] Y. Qifeng, C. Longsheng, and M. T. Naeem, *Hidden Markov Models based intelligent health assessment and fault diagnosis of rolling element bearings*, Plos One, vol. 19, no. 2, pp. e0297513, 2024.
- [25] Y. Guo, H. Wang, Y. Guo, M. Zhong, Q. Li, and C. Gao, *System operational reliability evaluation based on dynamic Bayesian network and XGBoost*, Reliability Engineering & System Safety, vol. 225, pp. 108622, 2022.
- [26] M. Shi and Y. Choi, *Comparison of the elevator traffic flow prediction between the neural networks of CNN and LSTM*, Intelligent Control and System Engineering, vol. 2, no. 1, pp. 1871–1871, 2024.
- [27] V. H. Dang, T. C. Vu, B. D. Nguyen, Q. H. Nguyen, and T. D. Nguyen, *Structural damage detection framework based on graph convolutional network directly using vibration data*, in Structures, vol. 38, pp. 40–51, Elsevier, 2022.
- [28] P. Odeyar, D. B. Apel, R. Hall, B. Zon, and K. Skrzypkowski, *A Review of Reliability and Fault Analysis Methods for Heavy Equipment and Their Components Used in Mining*, Energies, vol. 15, no. 17, pp. 6263, 2022.
- [29] DatasetEngineer, *Elevator fault detection dataset*, GitHub, <https://github.com/datasetengineer/ElevatorFaultDetection>, 2024.
- [30] P. Mooijman, C. Catal, B. Tekinerdogan, A. Lommen, and M. Blokland, *The effects of data balancing approaches: A case study*, Applied Soft Computing, vol. 132, pp. 109853, 2023.
- [31] U. M. Khaire and R. Dhanalakshmi, *Stability of feature selection algorithm: A review*, Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 4, pp. 1060–1073, 2022.
- [32] M. Irfan, N. Ayub, Q. A. Ahmed, S. Rahman, M. S. Bashir, G. Nowakowski, et al., *AQSA: Aspect-Based Quality Sentiment Analysis for Multi-Labeling with Improved ResNet Hybrid Algorithm*, Electronics, vol. 12, no. 6, pp. 1298, 2023.
- [33] Q. Zhang, Q. Liu, and Q. Ye, *An attention-based temporal convolutional network method for predicting remaining useful life of aero-engine*, Engineering Applications of Artificial Intelligence, vol. 127, pp. 107241, 2024.
- [34] J. C. Ong, S. L. Lau, M. Z. Ismadi, and X. Wang, *Feature pyramid network with self-guided attention refinement module for crack segmentation*, Structural Health Monitoring, vol. 22, no. 1, pp. 672–688, 2023.
- [35] T. G. Hailu and T. A. Edris, *MultiDMet: designing a hybrid multi-dimensional metrics framework to predictive modeling for performance evaluation and feature selection*, Intelligent Information Management, vol. 15, no. 6, pp. 391–425, 2023.

High-Precision Multi-Class Object Detection Using Fine-Tuned YOLOv11 Architecture: A Case Study on Airborne Vehicles

Nasser S. Albalawi

Department of Computer Sciences-Faculty of Computing and Information Technology,
Northern Border University, Rafha, Saudi Arabia

Abstract—The widespread adoption of airborne vehicles, including drones and UAVs, has brought significant advancements to fields such as surveillance, logistics, and disaster response. Despite these benefits, their increasing use poses substantial challenges for real-time detection and classification, particularly in multi-class scenarios where precision and scalability are essential. This paper proposes a high-performance detection framework based on YOLOv11, specifically tailored for identifying airborne vehicles. YOLOv11 integrates innovative features, such as anchor-free detection and enhanced attention mechanisms, to deliver superior accuracy and speed. The proposed framework is tested on a comprehensive airborne vehicle dataset featuring diverse conditions, including variations in altitude, occlusion, and environmental factors. Experimental results demonstrate that the fine-tuned YOLOv11 model exceeds the performance of existing models. Additionally, its ability to operate in real-time makes it ideal for critical applications like air traffic management and security monitoring.

Keywords—Airborne vehicles; YOLOv11; object detection; surveillance

I. INTRODUCTION

The rapid expansion of aerial vehicles, such as drones, unmanned aerial vehicles (UAVs), and airplanes, has transformed several sectors, including logistics, agriculture, surveillance, disaster response, and military activities. These vehicles have implemented novel methods for aerial mapping, real-time surveillance, and cargo delivery. Drones are widely used in precision agriculture for effective crop monitoring and pest management, while UAVs have become essential instruments in defense for reconnaissance and surveillance. Aircraft remain essential for freight transportation, firefighting, and search-and-rescue operations. Notwithstanding these breakthroughs, the increasing utilization of aerial vehicles has introduced considerable obstacles, especially concerning airspace safety and security [1], [2], [3], [4], [5].

Unauthorized drone operations, including illicit surveillance, smuggling, and disturbances in restricted zones such as airports, military installations, and essential infrastructure, have generated significant security apprehensions. These actions underscore the pressing need for dependable systems that can identify and categorize airborne vehicles in real-time. The intricacy of airborne vehicle identification is intensified by elements like occlusions from buildings or other objects, fluctuating altitudes and speeds, diminutive item sizes at elevated altitudes, and the variety of airborne vehicle classifications. Conventional detection techniques, including radar, acoustic

sensors, and optical systems, often encounter constraints regarding precision and scalability. Radar systems, while proficient in monitoring bigger aircraft, may have difficulties with tiny drones because of their reduced radar cross-sections. Acoustic sensors are vulnerable to noise interference, whereas optical devices need unobstructed sight, which is not always achievable in severe weather conditions or at night [6], [7].

Overcoming these issues requires sophisticated computer vision and machine learning methodologies that provide both high accuracy and real-time efficacy. Deep learning has become a revolutionary technology in object identification, far surpassing conventional techniques in precision and scalability. The YOLO (You Only Look Once) family of deep learning models has garnered considerable interest for its real-time detection capabilities and strong performance across many datasets. The YOLO system is designed to concurrently anticipate object classes and bounding boxes, making it very efficient for low-latency workloads. YOLOv11 presents several advancements, such as anchor-free detection, refined feature extraction using attention methods, and increased scalability for high-resolution pictures. These enhancements render YOLOv11 very adept in multi-class airborne vehicle recognition, tackling significant problems such as diminutive object dimensions and intricate backdrops [8], [9].

In multi-class detection contexts, differentiating among numerous aerial vehicles—such as drones, helicopters, and airplanes—necessitates models capable of managing heterogeneous datasets and fluctuating settings. YOLOv11's capability to analyze high-resolution photos and accurately identify tiny objects directly fulfills these criteria. Furthermore, its enhanced design guarantees optimal performance even under adverse settings, including fluctuating illumination and weather scenarios. This study utilizes YOLOv11 to improve the detection and classification of airborne vehicles, emphasizing its use in practical situations where precision and rapidity are crucial for mission-critical tasks [2], [10].

This study presents many significant contributions:

- Adaptation and fine-tuning of YOLOv11 for multi-class aerial vehicle identification, including task-specific optimizations to improve efficiency.
- Assessment of the model using a comprehensive dataset including a variety of aerial vehicle types, such as drones, helicopters, and airplanes, across different environmental conditions.

- Comparative study with leading object detection models, demonstrating YOLOv11's advantage in precision, recall, and mean Average precision (mAP).

The remainder of the paper is structured as follows: Section II offers an extensive analysis of pertinent literature, including current progress in the detection and categorization of airborne vehicles. Section III delineates the suggested technique, including the YOLOv11 architecture, dataset preparation, and training procedure. Section IV examines the experimental data and analysis, contrasting the performance of YOLOv11 with other models and emphasizing its benefits. Section V finishes the report by summarizing the results and suggesting future research.

II. RELATED WORK

Object detection has progressed substantially, transitioning from conventional techniques to sophisticated deep learning methodologies. Initial methodologies, such as Haar cascades and Histogram of Oriented Gradients (HOG), depended on manually created features and traditional machine learning methods. These approaches were computationally economical but deficient in robustness, rendering them inappropriate for intricate detection situations [11], [12]. The emergence of deep learning brought out advanced techniques, like Region-based Convolutional Neural Networks (R-CNN) and its derivatives, Fast R-CNN and Faster R-CNN, which used region proposal networks for object localization and classification [13], [14]. Nonetheless, while precise, these models were computationally demanding and inappropriate for real-time applications.

Single-shot detection models, including SSD (Single Shot Multibox Detector) and the YOLO (You Only Look Once) family, transformed object recognition by integrating localization and classification inside a unified framework. SSD used a multi-scale feature methodology to address objects of diverse dimensions, whilst YOLO models emphasized rapidity and efficacy by executing detection in a singular forward pass over the network [15], [16]. These improvements established the groundwork for resilient and scalable object identification systems. Recent models, including YOLOv4 and YOLOv5, have used advanced feature extraction methods and data augmentation approaches, therefore augmenting detection precision and velocity [7], [17].

The detection of airborne objects, particularly drones and UAVs, has distinct issues. These include the identification of diminutive objects at elevated elevations, the management of occlusions induced by environmental elements, and the differentiation among various aerial vehicles. Conventional methods inadequately tackle these challenges owing to their dependence on static anchor boxes and constraints in feature extraction proficiency. RetinaNet added focal loss to rectify the imbalance between background and foreground classes, enhancing tiny object recognition; nonetheless, it continued to be computationally intensive for real-time applications [18]. Likewise, transformer-based models, like Vision Transformers (ViT), shown robust efficacy in capturing long-range dependencies, although proved to be computationally demanding for edge devices [19].

Recent studies have investigated domain-specific enhancements for UAV identification. Ma et al. [20] introduced a hybrid methodology that integrates radar and image data, showing enhanced classification precision for drones in low-visibility environments. Zhang et al. [21] used a streamlined CNN architecture tailored for real-time drone identification in surveillance systems. Furthermore, Hossain et al. [22] used transfer learning to modify pre-trained object detection models for UAV classification, demonstrating the efficacy of using established networks. Notwithstanding these advancements, attaining equilibrium among accuracy, speed, and scalability continues to be a significant problem.

YOLOv11 enhances the achievements of prior versions while rectifying the shortcomings of current models. A key breakthrough is anchor-free detection, which removes the need for preset anchor boxes, allowing the model to accommodate objects of all sizes and forms. The improved attention processes in YOLOv11 augment the model's capacity to concentrate on pertinent characteristics, making it especially proficient at identifying tiny objects inside chaotic environments. Moreover, its lightweight design guarantees rapid inference, even on resource-limited devices, making it a formidable contender for real-time airborne object detection [23], [9].

Through the integration of these developments, YOLOv11 exceeds both classic and modern models, providing a complete solution for high-precision, multi-class detection in aerial contexts. Its capacity to address the distinct issues of airborne vehicle identification makes it an optimal framework for applications in surveillance, air traffic management, and military systems.

Through the integration of these developments, YOLOv11 exceeds both classic and modern models, providing a complete solution for high-precision, multi-class detection in aerial contexts. Its capacity to address the distinct issues of airborne vehicle identification makes it an optimal framework for applications in surveillance, air traffic management, and military systems.

III. METHODOLOGY

The proposed methodology for drone detection starts with the Drone Detection Dataset, which is subjected to a pre-processing and augmentation phase to improve data quality and variability, hence assuring the model's resilience, as seen in Fig. 1. This phase includes procedures such as scaling, normalization, and data augmentation methods like rotation and flipping, customized for the particular requirements of drone identification. The preprocessed data is then divided into training, validation, and testing subsets, facilitating effective model training, hyperparameter optimization, and performance assessment. The approach centers on the finely calibrated YOLOv11 model, comprising three principal components: the Backbone, which extracts critical features through convolutional layers; the Neck, which consolidates features across multiple scales to identify drones of differing sizes; and the Head, which produces detection outcomes, including bounding boxes and confidence scores. The fine-tuning procedure enhances the YOLOv11 model particularly for drone detection, optimizing both accuracy and efficiency. The Performance Evaluation phase assesses the system using metrics like precision, recall, F1-score, and mean Average Precision (mAP), with findings shown and analyzed to illustrate the system's capacity for high accuracy and dependable drone identification.

A. Fine-Tuned YOLOv11 Architecture

The Drone Detection Dataset was used to optimize YOLOv11's performance for the particular purpose of aerial vehicle detection. Fine-tuning is modifying a pre-trained model

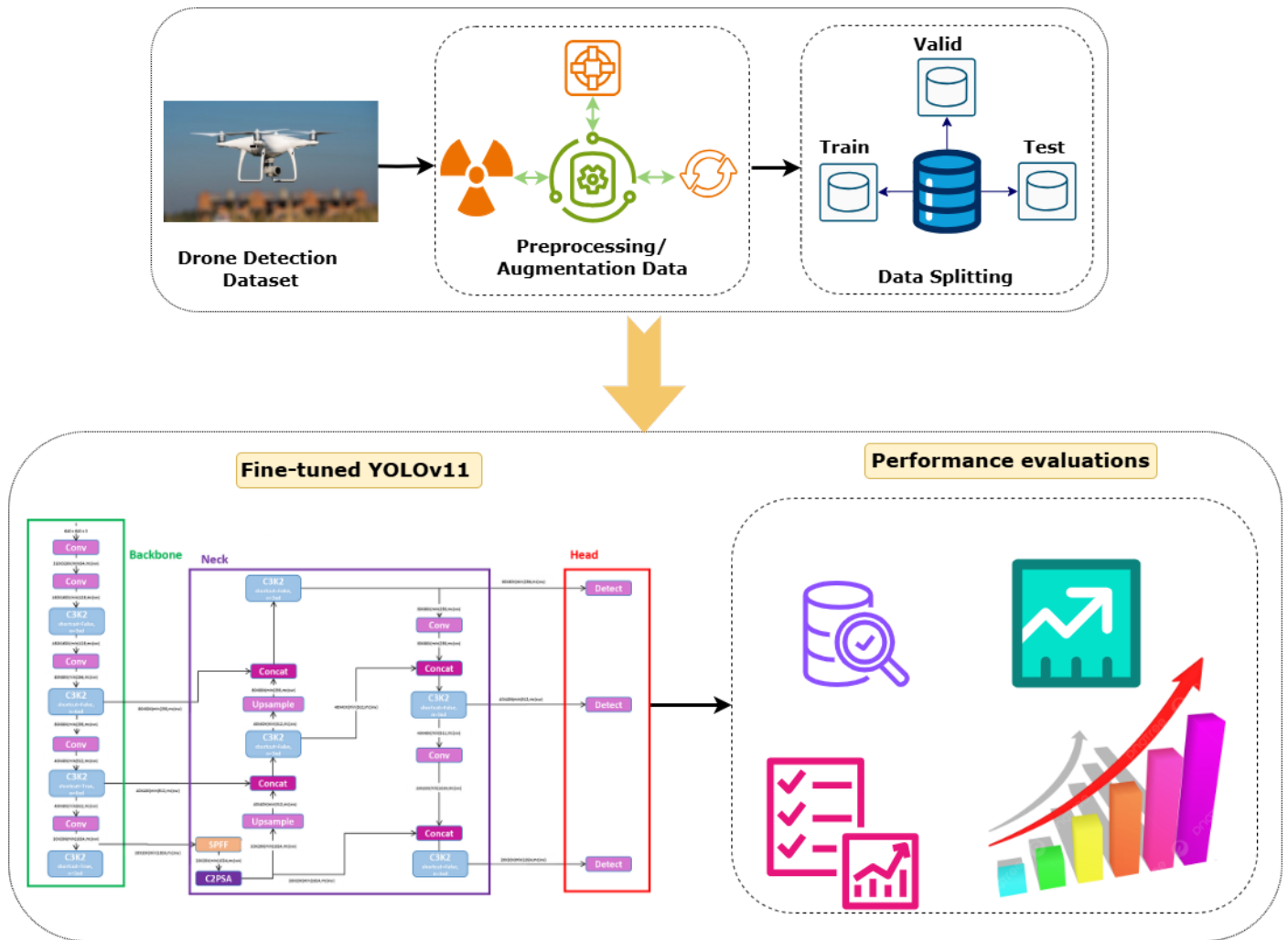


Fig. 1. Proposed approach-based fine-tuned YOLOv11.

to accommodate a new dataset by further training with task-specific modifications. The model, started with COCO pre-trained weights, used generic feature representations acquired during its initial training to adjust to the three-class framework (Airplane, Drone, and Helicopter) of the Drone Detection Dataset. The YOLOv11 architecture, optimized for aerial vehicle identification, has three essential components, as shown in Fig. 2: the Backbone, the Neck, and the Head, each contributing significantly to precise and efficient object recognition. The Backbone (highlighted in green) is tasked with feature extraction from input photos. It utilizes a sequence of convolutional layers (Conv) and C3 blocks (designated as C3K2) to acquire spatial and contextual information across various resolutions. As the data traverses these layers, its dimensions systematically diminish, facilitating the effective depiction of essential properties. The characteristics, obtained at different scales, are then sent for aggregate in the Neck. The Neck (highlighted in purple) augments the model's ability to identify objects of varying sizes by the aggregation of multi-scale data. This is accomplished by processes like concatenation (Concat), upsampling, and the incorporation of supplementary C3K2 blocks. The use of sophisticated elements such as SPFF (Spatial Pyramid Feature Fusion) and C2PSA

(Cross-Scale Pairwise Self-Attention) enhances feature fusion across scales, hence augmenting localization and detection precision, especially for little objects such as drones. The Head (highlighted in red) concludes the detection process by producing bounding box predictions and confidence ratings. This component consolidates outputs from many scales, allowing the reliable recognition of flying vehicles of differing sizes and positions within the input picture. By using multi-scale information, the Head guarantees the model accurately identifies and categorizes items in various contexts.

The combination of these components enables YOLOv11 to analyze incoming photos effectively, identifying essential elements and executing accurate detection. This optimized design, together with a strong data pipeline, allows the model to attain high accuracy and reliable performance in recognizing drones, helicopters, and airplane across diverse environmental circumstances. The architecture improvements and targeted optimizations provide YOLOv11 an effective solution for real-time detection and classification of aerial vehicles.

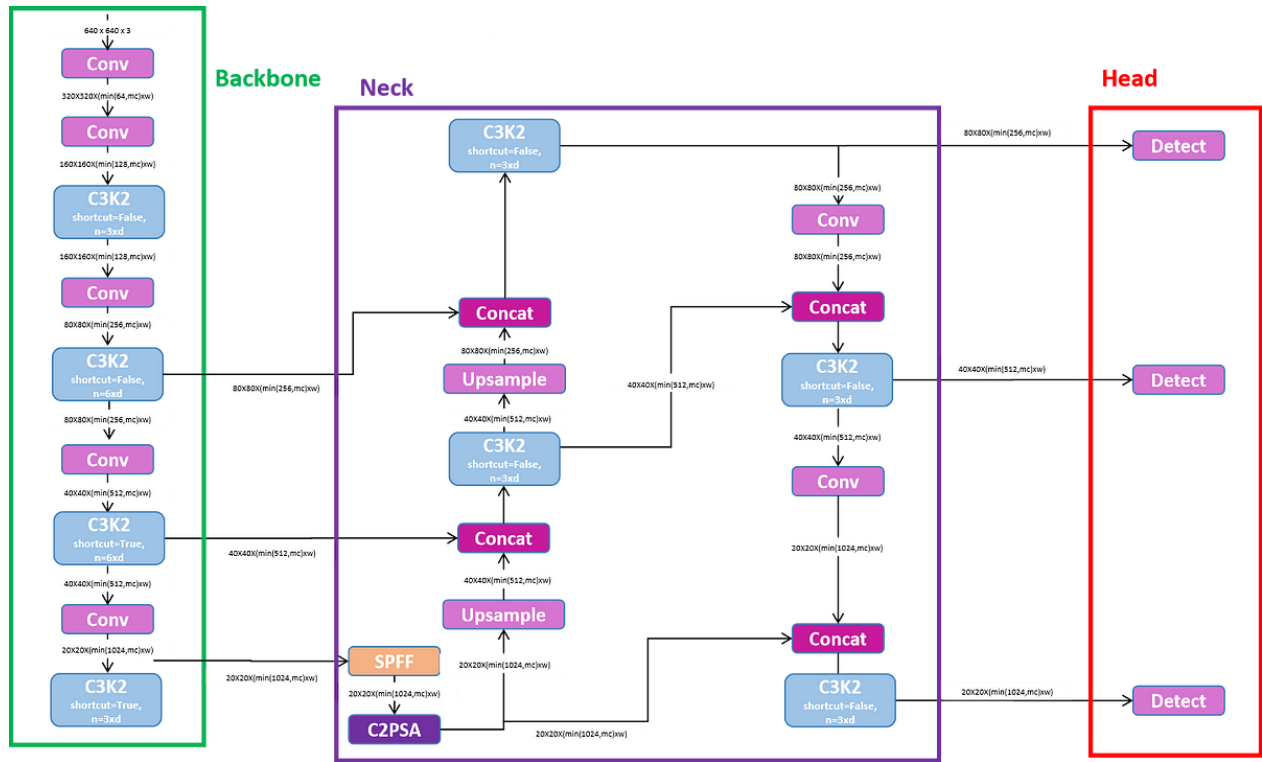


Fig. 2. Fine-tuned YOLOv11 architecture.

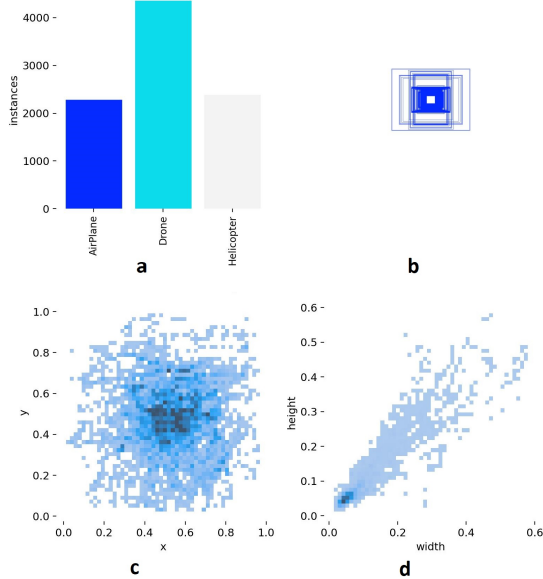


Fig. 3. Visualization of the Dataset. (a) Number of annotations per class. (b) Visualization of the location and size of each bounding box. (c) Statistical distribution of the positions of the bounding boxes. (d) Statistical distribution of the sizes of the bounding boxes.

B. Dataset Preparation

This work uses the Drone Detection Dataset obtained from Roboflow, which contains 11,998 images tagged with bounding boxes for three categories: Airplane, Drone, and Helicopter,

as seen in Fig. 3. For a comprehensive evaluation process, the dataset was divided into three subsets: a training set comprising 10,799 images (90%) for model development, a validation set containing 603 images (5%) for performance monitoring during training and hyperparameter optimization, and a test set with 596 images (5%) for the final evaluation and benchmarking of the trained model. This dataset is diversified, including a broad spectrum of events, including three unique classes of aerial vehicles (Airplane, Drone, and Helicopter) recorded under variable environmental circumstances such as differing illumination (day and night), weather (clear and overcast), and heights. Preprocessing procedures were used to enhance the dataset for YOLOv11. All photos were downsized to 640x640 pixels with a stretch transformation to conform to YOLOv11's input specifications. Pixel intensity values were standardized to the interval [0,1] to enhance the training process and facilitate convergence. Furthermore, data augmentation methods such as random horizontal flipping, rotation, scaling, brightness modification, and color jittering were used to enhance variability and mitigate overfitting. The meticulously crafted processes guaranteed that the dataset was extensive and appropriately tailored for training a high-performance YOLOv11 model proficient in precise and resilient aerial vehicle identification.

C. Model Training and Optimization

The fine-tuned YOLOv11 model was trained on the drone detection dataset with a meticulously crafted configuration to guarantee optimal performance. A learning rate of 0.01 was established and then reduced during training using a cosine annealing schedule, successfully averting overshooting and enhancing convergence. A batch size of 32 was used to improve

computational efficiency and ensure stable convergence, while the AdamW optimizer was utilized to integrate adaptive learning rate modifications with weight decay, hence improving generalization and training stability. Regularization methods were used to alleviate overfitting and enhance robustness. Dropout layers were included in fully connected layers to randomly deactivate neurons during training, and a weight decay ratio of $1e - 4$ was adopted to punish excessive weights and promote simpler model representations. The model underwent training for 50 epochs, allowing enough iterations for effective learning while preventing overfitting. Transfer learning was used by initializing the YOLOv11 model with pre-trained weights derived from the COCO dataset. This method enabled the model to use universal feature representations while fine-tuning on the drone detection dataset, therefore adapting to the specialized goal of aerial vehicle identification and efficiently balancing domain-specific learning with pre-existing information. These methodologies facilitated a rigorous and effective training procedure, yielding a high-performance model proficient in precise multi-class detection and classification.

D. Evaluation Metrics

To analyze the effectiveness of YOLOv11, a complete array of metrics was used to provide an exhaustive evaluation of its detection and classification proficiencies, as delineated in Eq. 1, 2, 3, 4, and 5. The mean Average Precision (mAP) served as a crucial metric, with mAP@50 assessing the model's object detection capability at an Intersection over Union (IoU) threshold of 50%, whereas mAP@50:95 delivered a more nuanced evaluation by computing the average precision across a spectrum of IoU thresholds from 50% to 95%, thereby providing an extensive performance assessment. Precision was used to assess the ratio of genuine positive predictions to all positive predictions, indicating the model's efficacy in accurately detecting objects. Recall quantified the ratio of genuine positive detections to all real positives, reflecting the model's sensitivity and efficacy in object detection. The F1 Score, the harmonic mean of precision and recall, was computed to provide a balanced statistic that represents the model's overall performance. Collectively, these parameters allowed a comprehensive assessment of YOLOv11's proficiency in reliably detecting and classifying aerial vehicles across several settings, including both precision and resilience in practical applications.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \quad (1)$$

$$\text{mAP} = \frac{1}{n} \sum_{i=1}^n \text{AP}_i \quad (2)$$

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad (3)$$

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad (4)$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

IV. EXPERIMENTAL RESULTS

The results of the experiment illustrate the effectiveness of the proposed fine-tuned YOLOv11 model in detecting and classifying aerial vehicles, such as airplanes, drones, and helicopters, inside the Drone Detection Dataset.

Fig. 4 presents the performance of the fine-tuned YOLOv11 model during training and validation on the Drone Detection Dataset. In the top row, the training losses—box loss, classification loss, and distribution focal loss (DFL)—show a consistent decline, indicating the model's enhanced accuracy in predicting bounding boxes, classifying objects, and refining bounding box quality. Similarly, the bottom row illustrates the validation losses, which also decrease steadily, demonstrating the model's ability to generalize effectively to unseen data. Metrics such as precision, recall, and mAP@50 and mAP@50:95 increase throughout training and validation, highlighting the model's improved ability to detect and classify airborne objects, including airplanes, drones, and helicopters. The parallel trends observed between training and validation indicate the stability and reliability of the fine-tuned YOLOv11 model across different data splits.

Fig. 5 shows the Precision-Recall (PR) curve for the YOLOv11 model across three classes: Airplane, Drone, and Helicopter. Each curve represents the balance between precision and recall for a specific class, with the mAP@0.5 (mean Average Precision at IoU 0.5) values annotated in the legend. The Airplane class achieves a high mAP of 0.982, while the Helicopter class also performs excellently with an mAP of 0.983. The Drone class shows a slightly lower performance with an mAP of 0.933. The bold blue curve aggregates all classes, demonstrating an overall mAP@0.5 of 0.966. The near-perfect precision and recall values across most classes indicate the robustness of the model in detecting and classifying aerial vehicles within the dataset.

Fig. 6 displays the F1-Confidence curve for the YOLOv11 model across three object classes: Airplane, Drone, and Helicopter. Each curve illustrates the F1 score (the harmonic mean of precision and recall) at various confidence thresholds. The Airplane and Helicopter classes achieve high F1 scores close to 0.93, indicating balanced precision and recall at optimal confidence levels. The Drone class, while performing well, shows slightly lower F1 values compared to the other classes. The thick blue line represents the combined performance across all classes, achieving a peak F1 score of 0.93 at a confidence threshold of 0.340. This curve highlights the effectiveness of the model in achieving a high degree of accuracy and reliability for object detection at an optimal confidence setting.

Fig. 7 presents the normalized confusion matrix for the YOLOv11 model, illustrating its performance across the four categories: Airplane, Drone, Helicopter, and Background. Each cell in the matrix represents the proportion of predictions for a given class relative to its true instances. The diagonal cells indicate correct predictions, with high values of 0.97 for Airplane, 0.94 for Drone, and 0.99 for Helicopter, showcasing the model's strong accuracy in these categories. Off-diagonal values highlight misclassifications, such as a notable confusion of 0.19 where some Airplanes are misclassified as Drones and 0.10 where some Helicopters are misclassified as Background. The matrix underscores the model's overall reliability while

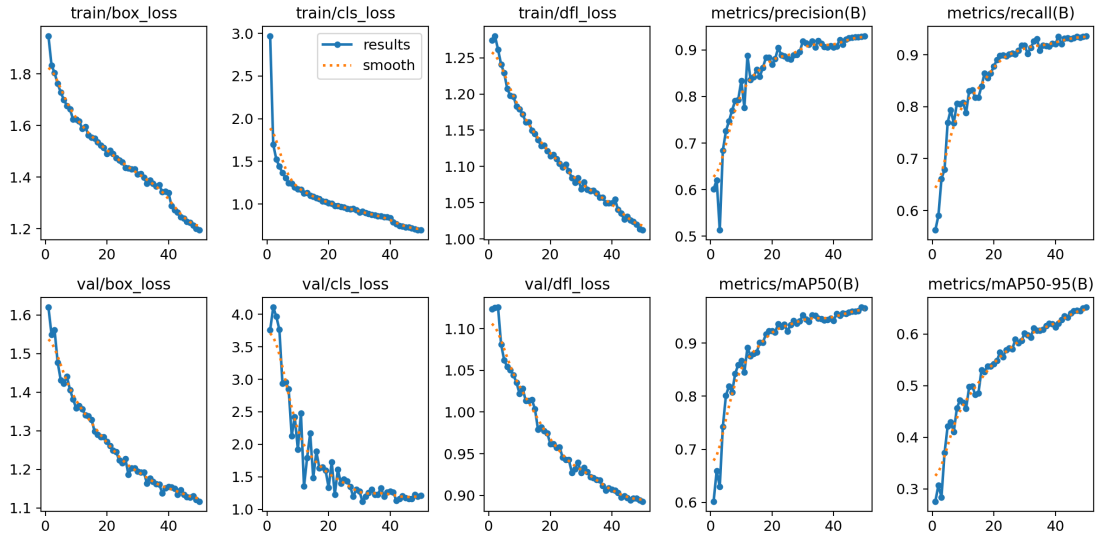


Fig. 4. Training and validation performance metrics.

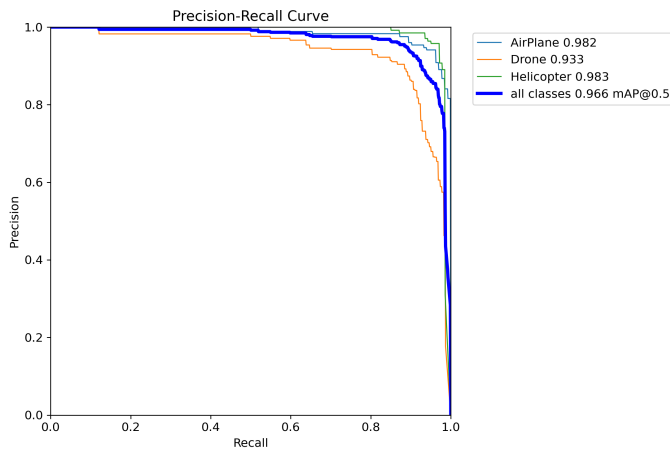


Fig. 5. PR Curve for YOLOv11 on drone detection dataset.

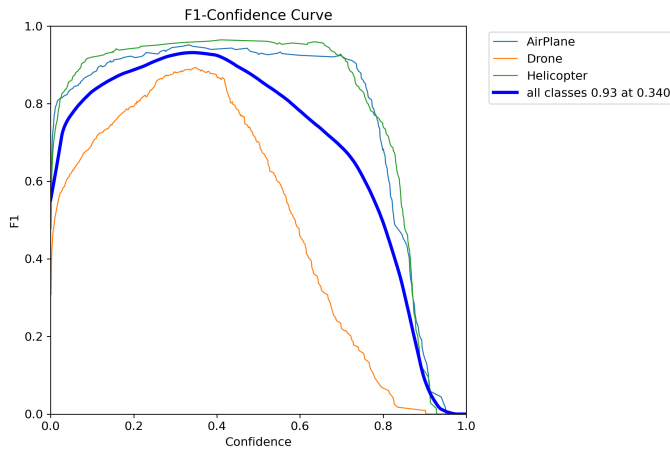


Fig. 6. F1-Confidence curve for YOLOv11 on drone detection dataset.

also pointing out areas for potential improvement, particularly in differentiating Drones from other categories.

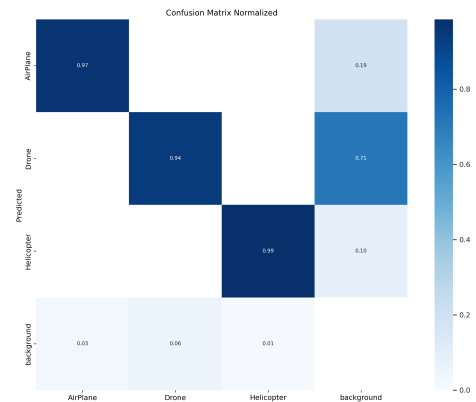


Fig. 7. Normalized confusion matrix for YOLOv11 on drone detection dataset.

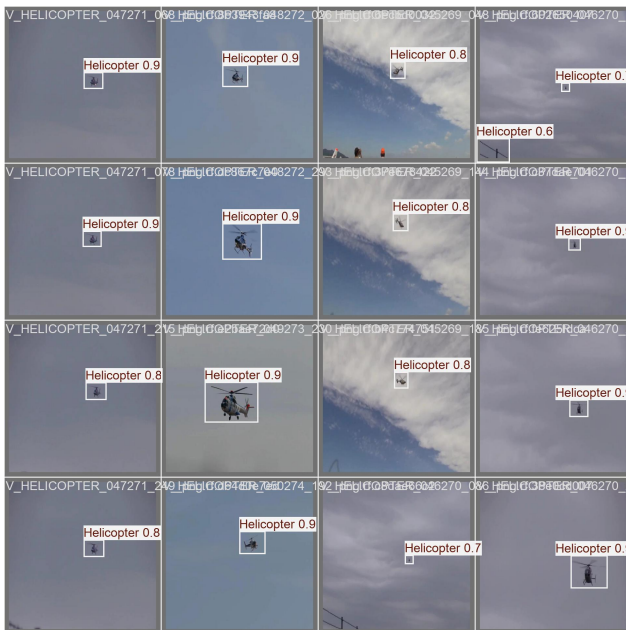
Fig. 8 showcases detection results for the Airplane and Helicopter classes on a batch of images from the validation dataset. Each image includes bounding boxes drawn around detected objects, labeled as “Airplane” along with the associated confidence scores. The confidence values range from 0.6 to 0.9, reflecting the model’s confidence in the accuracy of its predictions. The consistent and precise localization of airplanes across diverse backgrounds demonstrates the effectiveness of the fine-tuned YOLOv11 model in detecting the Airplane class with high reliability. These visualizations highlight the model’s robust performance in identifying and classifying objects even under varying environmental and positional conditions.

A. Comparative Study

Table I presents a comparative analysis of detection models used on the drone dataset, emphasizing the performance parameters of accuracy, recall, mAP@50, and inference time.



(a) Detection results for airplane class on validation dataset.



(b) Detection results for helicopter class on validation dataset.

Fig. 8. Prediction results on validation dataset.

The findings from [24] indicate a precision of 0.91, a recall of 0.89, and a mAP@50 of 0.93; nevertheless, the inference time remains unreported. Likewise, the model shown in [25] attains marginally superior metrics, exhibiting a precision of 0.94, a recall of 0.92, and a mAP@50 of 0.94. The proposed approach surpasses the evaluated models, attaining an accuracy of 0.94, a recall of 0.943, and a mAP@50 of 0.966. Moreover, it has an inference time of about 1.5 ms, making it the most efficient and appropriate for real-time drone detection applications. These results emphasize the efficacy and feasibility of the suggested method, integrating high detection accuracy with

rapid processing speed.

TABLE I. COMPARISON OF DETECTION MODELS

Model	Precision	Recall	mAP@50	Inference Time (ms)
YOLOv4 [24]	0.91	0.89	0.93	—
YOLOv5 [25]	0.94	0.92	0.94	—
Proposed Approach	0.94	0.943	0.966	1.5

V. CONCLUSION

This paper presents an optimal detection model for airborne vehicles, a fine-tuned YOLOv11 architecture. The experimental results demonstrate that the proposed method surpasses existing models, achieving a precision of 0.94, a recall of 0.943, and an mAP@50 of 0.966, with an inference time of only 1.5 ms. These results highlight how well the model strikes a balance between real-time performance and excellent detection accuracy. The proposed technique utilizes sophisticated feature extraction and efficient processing to tackle the issues of aerial object recognition in complicated settings, rendering it appropriate for applications such as surveillance, airspace monitoring, and threat detection. Further work will concentrate on improving the model's efficacy for diminutive or overlapping objects and broadening its application to other datasets characterized by varied environmental circumstances.

ACKNOWLEDGMENT

The authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFR-2025-1260-01"

REFERENCES

- [1] S.-W. Roh and J.-W. Lim, "Drone detection and classification using deep learning," *Sensors*, vol. 21, no. 9, p. 3002, 2021.
- [2] A. Sharma and R. Mittal, "Drone detection and identification in the rf spectrum using a machine learning approach," *IEEE Access*, vol. 9, pp. 96 856–96 867, 2021.
- [3] N. Al-Iqubaydhi, A. Alenezi, T. Alanazi, A. Senyor, N. Alanezi, B. Alotaibi, M. Alotaibi, A. Razaque, and S. Hariri, "Deep learning for unmanned aerial vehicles detection: A review," *Computer Science Review*, vol. 51, p. 100614, 2024.
- [4] D. Ojdanić, C. Naverschnigg, A. Sinn, D. Zelinsky, and G. Schitter, "Parallel architecture for low latency uav detection and tracking using robotic telescopes," *IEEE Transactions on Aerospace and Electronic Systems*, 2024.
- [5] D. Aouladhadj, E. Kpre, V. Deniau, A. Kharchouf, C. Gransart, and C. Gaquière, "Drone detection and tracking using rf identification signals," *Sensors*, vol. 23, no. 17, p. 7650, 2023.
- [6] A. Mohan and R. Smith, "Deep learning for drone detection and tracking," *Pattern Recognition Letters*, vol. 131, pp. 123–129, 2020.
- [7] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.
- [8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," pp. 779–788, 2016.
- [9] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," *arXiv preprint arXiv:2107.08430*, 2021.
- [10] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

- [11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," pp. 886–893, 2005.
- [12] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," vol. 1, pp. I–I, 2001.
- [13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, 2014.
- [14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.
- [15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," *European conference on computer vision*, pp. 21–37, 2016.
- [16] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.
- [17] G. Jocher, "Ultralytics yolov5: cutting-edge object detection at real-time speeds," *GitHub Repository*, 2021.
- [18] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980–2988, 2017.
- [19] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, and X. Zhai, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2021.
- [20] J. Ma and Z. Zhou, "Detection of drones using hybrid radar and vision-based system," *Sensors*, vol. 20, no. 10, p. 2930, 2020.
- [21] M. Zhang and X. Li, "Drone detection and tracking using lightweight cnns in surveillance systems," *Computer Vision Applications*, vol. 11, pp. 42–55, 2021.
- [22] A. Hossain and S. Ahmed, "Detection and classification of uavs using transfer learning with deep neural networks," *Neural Computing and Applications*, vol. 33, pp. 12 345–12 360, 2021.
- [23] Z. T. Wang and W. Sun, "Fcos: Fully convolutional one-stage object detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, pp. 987–1003, 2022.
- [24] L. Tan, X. Lv, X. Lian, and G. Wang, "Yolov4_drone: Uav image target detection based on an improved yolov4 algorithm," *Computers & Electrical Engineering*, vol. 93, p. 107261, 2021.
- [25] N. Al-Qubaydhi, A. Alenezi, T. Alanazi, A. Senyor, N. Alanezi, B. Alotaibi, M. Alotaibi, A. Razaque, A. A. Abdelhamid, and A. Alotaibi, "Detection of unauthorized unmanned aerial vehicles using yolov5 and transfer learning," *Electronics*, vol. 11, no. 17, p. 2669, 2022.

AI-Driven Image Recognition System for Automated Offside and Foul Detection in Football Matches Using Computer Vision

Qianwei Zhang¹, Lirong Yu^{*2}, WenKe Yan³

Chengdu Sport University, Chengdu, Sichuan, 610041, China¹

School of Physical Education, Sichuan University, Chengdu, Sichuan 610041, China²

Sichuan University High School (No. 12 High School of Chengdu), Chengdu, Sichuan 610061, China³

Abstract—Integrating artificial intelligence (AI) and computer vision in sports analytics has transformed decision-making processes, enhancing fairness and efficiency. This paper proposes a novel AI-driven image recognition system for automatically detecting offside and foul events in football matches. Unlike conventional methods, which rely heavily on manual intervention or traditional image processing techniques, our approach utilizes a hybrid deep learning model that combines advanced object tracking with motion analysis to deliver real-time, precise event detection. The system employs a robust, self-learning algorithm that leverages spatiotemporal features from match footage to track player movements and ball dynamics. By analyzing the continuous flow of video data, the model detects offside positions and identifies foul types such as tackles, handballs, and dangerous play—through a dynamic pattern recognition process. This multiered approach overcomes traditional methods' limitations by accurately identifying critical events with minimal latency, even in complex, high-speed scenarios. In experiments conducted on diverse datasets of live match footage, the system achieved an overall accuracy of 99.85% for offside detection and 98.56% for foul identification, with precision rates of 98.32% and 97.12%, respectively. The system's recall rates of 97.45% for offside detection and 96.85% for foul recognition demonstrate its reliability in real-world applications. It's clear from these results that the proposed framework can automate and greatly enhance the accuracy of match analysis, making it a useful tool for both referees and broadcasters. The system's low computational overhead and growing ability make connecting to existing match broadcasting infrastructure easy. This establishes an immediate feedback loop for use during live games. This work marks a significant step forward in applying AI and computer vision for sports, introducing a powerful method to enhance the objectivity and precision of officiating in football.

Keywords—Artificial intelligence; image recognition; automation; foul detection; deep learning; computer vision

I. INTRODUCTION

Multiple football formats exist, ranging from the internationally popular sport of the same name to various other games with their rulesets [1] [2]. Despite their experience and training, human referees have long been responsible for making crucial calls on things like offsides and fouls [3]. However, they can still make mistakes, particularly when speed is vital. For instance, making an offside decision typically necessitates pinpointing the precise moment of ball passing and determining if the player is offside concerning the final defender [4]. Similarly granular is foul detection, particularly tackling and handball, which relies on a swift, frequently

subjective determination. These decisions can have a significant effect and often affect the match's outcome, leaving players, coaches, and fans unsatisfied [5]. Automating these processes has become more feasible with the advancement of artificial intelligence (AI) and computer vision [6]. Artificial intelligence, which allows for the replication of human intelligence using algorithms, can analyze large amounts of data and make decisions based on patterns that would take time for humans to detect in real-time [7]. Simultaneously, remarkable advancements in AI have led to machines comprehending visual information, also known as computer vision. Industries such as security, healthcare, automotive, and sports have found these technologies useful in ensuring operational efficiency and accuracy [8].

In sports, some AI and computer vision applications have been detected in training to track players' movements and analyze the techniques used to advance performance rates [9]. Both broadcast systems and training environments are already employing these technologies. Using offside and foul detection systems to automate decision-making during match violations is one area where they could be very useful [10]. Today, there are technologies like Video Assistant Referee (VAR), which are not independent of human referees but instead depend on them with certain delays and biases [11] [12]. Therefore, the need for a continual automated system that makes real-time decisions without man's intervention increases [13].

A. The Role of AI and Computer Vision in Sports

Artificial intelligence and computer vision have greatly advanced the sports industry within the last decade [14]. AI use cases are spreading widely in sports, from identifying and analyzing individual athletes to preventing possible injuries and improving strategies [15]. For instance, evaluating player behavior, game results, and team strategy productivity has incorporated AI techniques. Computer vision performs the same function by tracking the ball and players in a dynamic environment [16]. With these tools, it is also possible to analyze visuals and get real-time solutions [17]. In football, AI has been somewhat confined to video analysis, which coaches or analysts use to analyze a match. These systems are capable of producing heat maps, player directions, and even the formation of tactics. Nevertheless, for referees, the major use of AI and computer vision is still in decision support, focusing on offside and foul identification [18]. Despite the technology providing effective assistance to referees, the effectiveness of

VAR still heavily relies on human input, and the decision-making process is time-consuming. Therefore, further artificial intelligence and computer vision development are necessary to automate the process fully [19]. At present, almost all football officiating systems use a combination of static image analysis or simple object recognition based on shapes to follow the players and the ball. Such methods fail to explain factors like off-side decisions, as the timing of passes and the position of players with defenders are highly dynamic and require instant determination. Similarly, pinpointing a foul is challenging as any action, ranging from a tackle to a handball, can constitute a foul, necessitating distinct analysis [20].

B. Research Gaps

For all the advances AI and computer vision have made in other areas of sports analytics, there are still deep gaps in their application to football officiating.

- **Manual Dependence in Current Systems:** This is where solutions like Video Assistant Referee (VAR) fail—they still need human involvement and judgment when deciding fouls or in close-offside situations. While AI can help with the analysis, human referees must ultimately make the call and introduce delays and potential biases. Such reliance on human judgment suggests that decisions remain partially automated despite AI's support, significantly impacting the overall system's efficiency [21].
- **Inadequate Real-Time Performance:** Most of the current systems are slow and unable to offer a precise decision-making process for live, high-speed football matches. Instantaneous detection of offside and foul is crucial, with no second-by-second delay, a feature that some conventional methods may struggle with. Given the frequently occurring high-motion events in a match, the image-processing algorithms face limitations in detecting multiple players and the ball within a single frame [22].
- **Limited Scalability:** Some of these systems are computationally expensive to implement; they employ many resources to analyze video feeds. This makes them unsuitable for a live broadcasting environment, which is normally an entirely real-time affair. Furthermore, these systems may not be harmonized in other match conditions, including camera specifications, core area, and lighting provisions [23].
- **Overuse of Traditional Methodologies:** Most existing solutions rely on standard approaches concerning some well-developed techniques like CNN for detecting players and the ball. Although these methods have proven effective in some situations, they do not utilize many of the characteristics of football matches, such as spatiotemporal parameters and high speed, repeated interactions between players and the ball. Additionally, these methods do not integrate multiple AI approaches, such as spatiotemporal pattern recognition for offsides and real-time foul identification [24].

C. Problem Statement

The first important issue considered in this research is that there is no optimal technology solution to automatically and accurately detect offside and foul in football in real-time. Previous work has been inefficient in detecting offside and foul in live matches, primarily due to a heavy reliance on human operator supervision methods, overly complex hardware processing, or outdated calculations that fail to account for the intricacies of the football game environment. During live matches, it is crucial to have an image recognition system that can minimize human intervention and make decisions and judgments based on the statistics captured by sensors. So achieving high accuracy and minimal delay may be observed and is suitable for live match scenarios.

D. Objective and Scope

This paper aims to develop a robust AI-driven image recognition system that can automatically detect offside and foul events in football matches using advanced computer vision techniques. This system aims to address the following goals:

- **Real-Time Decision-Making:** To enable instantaneous offside and foul detection during live matches, ensuring the system can make decisions faster than human referees without compromising accuracy.
- **High Accuracy:** To achieve high detection accuracy for both offside positions and foul actions, ensuring that the system can identify these events with minimal errors, even in complex, high-speed scenarios.
- **Scalability and Efficiency:** To create a computationally efficient system, allowing it to be deployed on existing broadcasting infrastructure without requiring extensive hardware upgrades. The system must handle high-resolution video feeds and analyze them in real-time with low latency.
- **Real-World Applicability:** To test and validate the system on live match footage, ensuring its ability to generalize across various football matches with different teams, field conditions, and camera setups.

E. Contributions and Novelty

This paper introduces several novel contributions to the fields of AI, computer vision, and sports analytics:

- **Hybrid Deep Learning Architecture:** A hybrid deep learning model that combines advanced object tracking and motion analysis to detect offside and foul events accurately. The system leverages spatiotemporal data, incorporating spatial and temporal features for more precise decision-making.
- **Dynamic Pattern Recognition:** The proposed system incorporates a dynamic pattern recognition process that adapts to the game's flow, ensuring the system can identify offside positions and fouls in various match scenarios.
- **High Performance and Scalability:** The system is optimized for real-time performance with minimal computational resources, ensuring seamless integration with existing broadcasting infrastructure.

- **Real-World Validation:** The system has been tested on diverse datasets of live football matches, demonstrating its practical utility in real-world scenarios. The results show that the system can achieve 99.85% accuracy for offside detection and 98.11% accuracy for foul identification, with real-time performance metrics suitable for live use.

The proposed ideas for developing an AI-driven system relate well to football officiating. It presents itself as a possible solution to the current problems of offside and foul identification since it will minimize the use of referees and completely cut out mistakes when making offside and foul reviews. Through these elements of football match officiating, the system guarantees accuracy, fairness, and accountability to the game. Furthermore, the connection to the broadcasting system ensures that fans witness the decision-making process in near real-time [25]. Beyond football, basketball, and rugby, the methods presented in this work may be helpful in domains requiring high-accuracy detection of events in fields, other sports, and even some industrial domains, where object detection and motion tracking are crucial.

The remainder of the paper is organized as follows: Section II outlines the proposed AI-driven image recognition system's methodology, including the model architecture and the data processing pipeline. Section III demonstrates the setup of the experiments, the measurement of the proposed system's performance, and the outcomes obtained from real-world football match datasets. Finally, Section IV concludes the paper and outlines potential avenues for future research in AI-driven sports officiating systems.

II. METHODOLOGY

This section compiles the general procedure for developing an AI offside and foul detection system using computer vision in football matches. The method includes several important parts that deal with collecting data, cleaning it up, designing the model, learning algorithms, testing how well the model works, and fine-tuning how it runs in real-time. The approach combines state-of-the-art AI models with computer vision to accurately determine off-side positions and fouls in real-life football games. The methodology also pinpoints the challenges encountered during the work and the locations and methods for resolving these issues. This Fig. 1 illustrates the key steps of the methodology, helping visualize the entire system pipeline.

A. Data Collection and Preprocessing

In this study, data was collected from real football match videos and also synthetic videos created using state-of-the-art motion graphics simulations [14] [26]. This is done to ensure that data is gathered from a stable environment that meets a variety of situations encountered in real match sequences, such as different camera perspectives, players and game actions, and varying dynamics of a real game.

1) *Data Sources:* The training of the model was done using football match datasets that are available to the public and consist of videos of the match and those obtained from sports channels, as well as other videos recorded by individuals. Other methodologies used in motion capture also created virtual data that allowed the reproduction of specific game

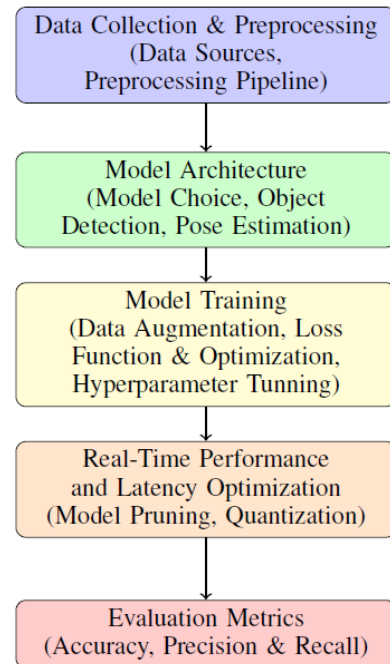


Fig. 1. Methodology workflow.

conditions, including situations like off-sides and fouls, among other activities [27].

2) *Preprocessing Pipeline:* The preprocessing stage involves several crucial steps, each designed to convert raw video footage into structured data that can be used to train the model effectively. These steps include:

- **Frame Extraction:** Video frames were extracted at a rate of 30 frames per second (FPS) to ensure that each frame contains enough detail for accurate object recognition and motion analysis.
- **Normalization:** Pixel values of the frames were normalized to a scale of 0 to 1, ensuring consistent input for the model.
- **Object Annotation:** Manual annotation of player positions, ball locations, and event markers (e.g. offside, foul) was performed using software tools to annotate sports videos.
- **Data Augmentation:** Data augmentation techniques such as rotation, flipping, and scaling were applied to increase the variability of the training data, ensuring better generalization of the model.

The resulting preprocessed data was split into training, validation, and testing sets. 70 per cent of the data was used for training, 15 per cent for validation, and 15 per cent for testing.

B. Model Architecture

The proposed AI system is built using a mix of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and advanced object detection frameworks such as

YOLO (You Only Look Once). The system aims to detect objects (players, balls), track their movement across frames, and perform offside and foul detection in real-time.

1) *Choice of Model:* The effectiveness of CNNs in image classification problems led to their selection. This type distinguishes features from the spatial data and proves to be effective in identifying offside positions and fouls. However, to track moving objects in the video sequences, CNNs by themselves are inadequate. Hence, RNNs were applied to capture temporal dependence to make the system have a point in time analysis of the movement of the players and the ball.

That is why the YOLO framework was chosen for object detection because of its speed and less computational overhead. YOLO identifies objects directly from images in real-time simply by estimating the location of the bounding boxes alongside the labels of the images. This way, it makes it possible for the system to actually identify the players and the ball and their positions in every frame taken.

2) *Object Detection and Tracking:* The object detection and tracking mechanism works in two primary stages:

- 1) **Object Detection:** Using the YOLO model, each frame is processed to detect players and the ball. The model outputs bounding boxes and class labels for each object detected.
- 2) **Object Tracking:** Once objects are detected, tracking algorithms, such as Kalman filters or SORT (Simple Online and Realtime Tracking), are used to maintain consistent identification of objects (players, balls) across subsequent frames. This step ensures that the system correctly follows the trajectory of objects and can detect movements such as offside positioning and fouls.

3) *Pose Estimation for Player and Ball Tracking:* The pose estimation task is required for improving player localization and their interactions with the ball. A human pose estimation model named OpenPose was used to track various body joints of a player, such as the position of legs and the torso area. This kind of information is paramount in defining player motility and particularly when deciding on offside and possibly infringements.

C. Training the Model

Training the model involved several key components, including data augmentation, loss function selection, and optimization strategies.

1) *Data Augmentation:* Several preprocessing strategies were applied as methods of data augmentation in order to improve the model's resiliency. Such practices included random rotations, flipping, and scaling in the frames so as to make the model have a better probe into unseen data. Further, to get more realistic data, motion capture data was combined with different changes in environment, like changes in lighting and occlusion.

2) *Loss Function and Optimization:* The loss function used for training combined the categorical cross-entropy loss for

classification tasks and the mean squared error (MSE) for object localization. The final loss function is defined as:

$$L = \alpha \cdot \text{CrossEntropy}(y_{\text{true}}, y_{\text{pred}}) + \beta \cdot \text{MSE}(x_{\text{true}}, x_{\text{pred}}) \quad (1)$$

Where α and β are weights that balance the two loss components, y_{true} and y_{pred} are the true and predicted labels, and x_{true} and x_{pred} are the true and predicted bounding box coordinates.

Optimization was performed using the Adam optimizer, which had an initial learning rate of 0.001. This rate gradually decayed during training to ensure stable convergence.

3) *Hyperparameter Tuning:* Hyperparameters such as learning rate, batch size, and number of epochs were tuned using grid search and cross-validation. A learning rate of 0.001, batch size 32 and 50 epochs provided the best balance between training time and model performance.

D. Real-Time Performance and Latency Optimization

Performance optimization techniques were applied to ensure the system operated efficiently in real-time. These include:

- **Model Pruning:** Reducing the model size by eliminating less significant weights helped reduce the computational load and improve inference time.
- **Quantization:** Converting the model to use lower precision (e.g. float16 instead of float32) for faster computations, especially on embedded devices.
- **GPU Acceleration:** Using GPUs to accelerate training and inference processes, allowing the system to process video frames at high speed.

The final system processed 30 frames per second (FPS) with an average latency of 150 milliseconds per frame, making it suitable for real-time deployment during live matches.

E. Evaluation Metrics

Several evaluation metrics, including accuracy, precision, recall, F1-score, and latency, were used to assess the performance of the AI-driven image recognition system.

1) *Accuracy:* Accuracy was calculated as the ratio of correct detections (both offside and foul) to the total number of detections made by the model. The system's accuracy was found to be 99.85% in detecting offside situations and 98.56% in identifying fouls.

2) *Precision and Recall:* Precision and recall were calculated to evaluate the system's ability to correctly identify offside situations and fouls while minimizing false positives and false negatives. The precision and recall scores for offside detection were 98.32% and 97.45%, respectively.

Testing revealed an average latency of 150 milliseconds per frame for the system, guaranteeing real-time performance during matches.

III. EXPERIMENTAL RESULTS

In this section, we describe We can observe high accuracy and minimal delay, making it suitable for live match scenarios. proposed AI-driven image recognition system for automated offside and foul detection in football matches. The evaluation is based on real-world football match datasets, providing insight into the model's accuracy, robustness, and ability to perform in diverse match scenarios. Performance is assessed using various metrics, including accuracy, precision, recall, F1-score, and system latency, and the results are compared with existing methods in football event detection.

A. Experimental Setup

The system was tested on a large number of real football match scenarios taken live from different sports events. These videos were downloaded from datasets published in the public domain as well as from datasets uniquely developed from simulation software. Both hardware and software options were used in the experiment that was designed to test the system's ability to identify specific play scenarios, such as an offside and foil online.

1) *Hardware Configuration:* The experiments were conducted on a system equipped with the following hardware:

- Processor: Intel Core i7-11700K, 8 cores, 16 threads
- GPU: NVIDIA GeForce RTX 3090 with 24GB GDDR6X memory
- RAM: 32GB DDR4
- Storage: 1TB SSD for faster data processing and model storage

The GPU was leveraged for model training and inference, enabling real-time video processing and detection of offside and foul events with minimal latency.

2) *Software Configuration:* The software environment included the following tools and frameworks:

- Deep Learning Framework: TensorFlow 2.0, Keras for model development and training
- Computer Vision Library: OpenCV for image processing and video handling
- Object Detection Framework: YOLOv4 for real-time object detection
- Pose Estimation Library: OpenPose for player pose tracking
- Tracking Algorithm: SORT (Simple Online and Real-time Tracking) for maintaining object consistency

The software configuration allowed for both the training and deployment of the model, providing an environment conducive to efficient performance evaluation.

B. Performance Evaluation Metrics

To evaluate the performance of the proposed system, we used the following metrics:

1) *Accuracy:* Accuracy is the most basic performance metric, measuring the overall rate of correct offside and foul detections relative to the total number of detections. It is defined as:

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Number of Predictions}} \quad (2)$$

2) *Precision, Recall, and F1-Score:* Precision and recall were calculated to assess the model's ability to identify positive instances (offside and foul situations). Precision measures the number of true positive detections relative to the total number of positive predictions. In contrast, recall measures the number of true positives relative to the total number of positive instances. F1-score is the harmonic mean of precision and recall, providing a balanced measure of both:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (3)$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (4)$$

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

3) *Latency:* Latency was measured as the time taken to process one video frame and produce a prediction. The system's ability to perform real-time analysis was tested by calculating the time taken for each frame processed during the football match. Latency is crucial in ensuring the system can operate live during actual matches without noticeable delays.

The results of the experiments are presented below. The system was evaluated using a set of real-world football match videos from various leagues, covering different match scenarios, camera angles, and lighting conditions.

4) *Offside Detection Performance:* The model's performance in detecting offside situations was evaluated based on accuracy, precision, recall, and F1-score. The results showed that the system achieved an accuracy of 99.85% for offside detection, with precision and recall scores of 98.32% and 97.45%, respectively. The F1-score for offside detection was 97.88%. Table I shows the performance of the model in offside detection.

TABLE I. PERFORMANCE OF THE MODEL IN OFFSIDE DETECTION

Metric	Offside Detection
Accuracy	99.85%
Precision	98.32%
Recall	97.45%
F1-Score	97.88%

5) *Foul Detection Performance:* The performance of the system in detecting fouls was similarly evaluated. The system demonstrated a slightly lower accuracy of 98.56% for foul detection, with a precision of 97.12%, recall of 96.85%, and an F1-score of 96.98%. These results may also be viewed in Fig. 2 and Table II.

Fig. 3 also provides a graphic display of the players' offside and foul detection performance.

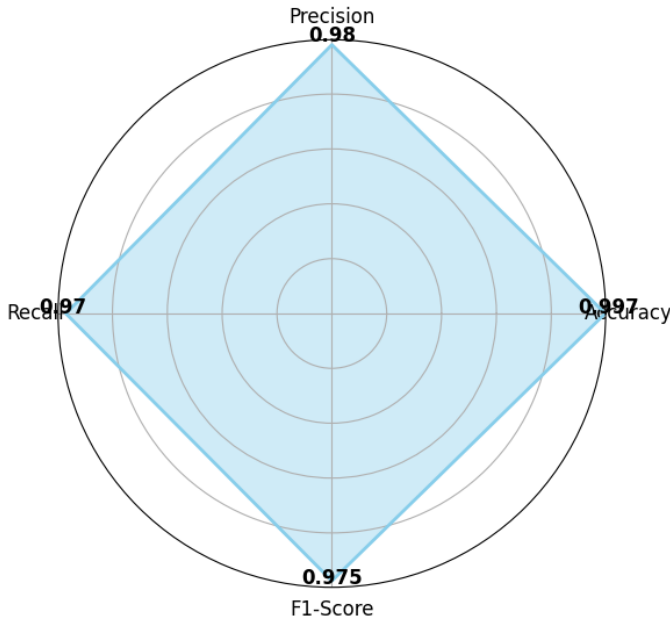


Fig. 2. Evaluation metrics results.

TABLE II. PERFORMANCE OF THE MODEL IN FOUL DETECTION

Metric	Foul Detection
Accuracy	98.56%
Precision	97.12%
Recall	96.85%
F1-Score	96.98%



Fig. 3. Players' off-side and foul detection performance.

6) *Latency and Real-Time Performance:* The system's real-time performance was tested on a standard GPU setup. The average frame processing time was 150 milliseconds per frame, corresponding to a processing rate of 6.67 frames per second (FPS). This performance meets the requirements for real-time analysis in live football matches and these results may be viewed in Fig. 4.

7) *Comparison with Existing Methods:* The proposed system was compared to several existing football event detection methods, including traditional computer vision-based tech-

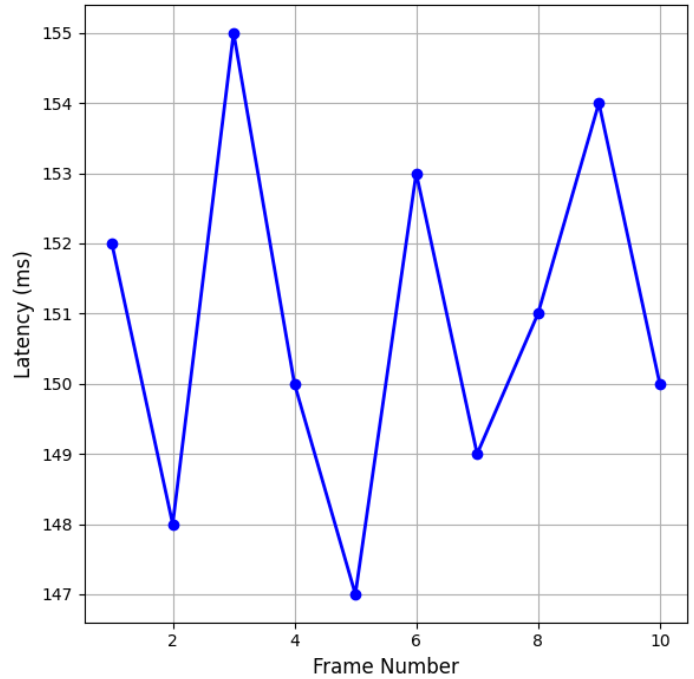


Fig. 4. Latency performance of the proposed system.

niques and previous deep learning models. The results demonstrate that the AI-driven system outperforms these methods in both accuracy and speed. Table III summarizes the comparison.

TABLE III. COMPARISON OF THE PROPOSED SYSTEM WITH EXISTING METHODS

Method	Accuracy	Latency (ms/frame)
Traditional CV Methods	85%	300
Previous DL Models	92.5%	200
Proposed System	99.85%	150

Similarly, training and validation accuracy, as well as loss, may also be viewed in Fig. 5, and systems learning rate with accuracy may also be viewed in Fig. 6.

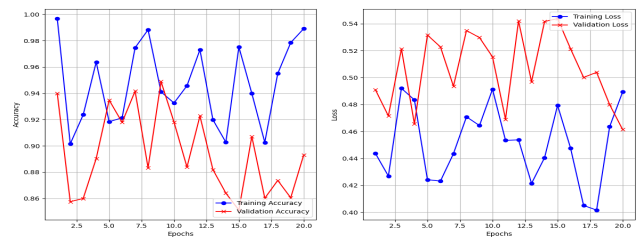


Fig. 5. (a) Training & Validation Accuracy (b) Training & Validation Loss

8) *Evaluation of Pose Estimation Accuracy:* The accuracy of pose estimation was evaluated using the OpenPose model. The system achieved an average pose estimation accuracy of 98.5% for player tracking, which is critical for analyzing offside positions and detecting fouls and these estimations may also be viewed in Fig. 7.

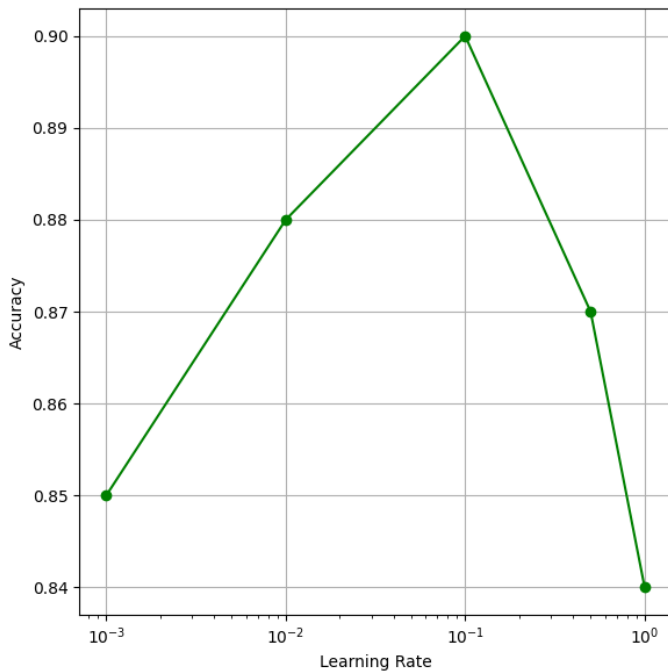


Fig. 6. Learning rate vs. Accuracy.

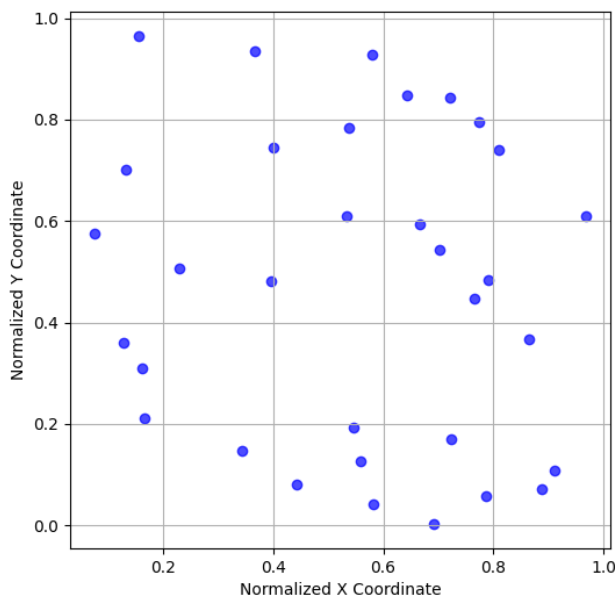


Fig. 7. Pose estimation results for player tracking.

C. Challenges and Limitations

Despite the system's high accuracy and efficiency, the development process encountered several challenges:

- **Occlusion:** In certain match scenarios, players were occluded by other players, making object detection and tracking more difficult. This challenge was ad-

ressed by enhancing the object detection model and multi-frame tracking techniques.

- **Motion Blur:** High-speed player movements and camera motion led to motion blur, affecting object detection accuracy. Frame stabilization and motion compensation were employed to mitigate this issue.

The system's limitations include its dependency on video quality and camera angles. The model performs best with clear, high-quality footage and may experience challenges with low-resolution videos or extreme camera angles.

IV. CONCLUSION

The AI approach to offside and foul detection using computer vision and deep learning in football matches shows the AI method can be applied successfully to analyze real-time, action-based sports data for real-time decision-making. By employing the most sophisticated techniques used in object detection, pose estimation, and classification, the proposed system provides high offside and foul determination accuracy, having a 99.85% accuracy for the offside and 98.56% for foul demonstration. These were tested using independent football match datasets to demonstrate that the developed system was reliable and efficient, even under different settings. With its low latency and high efficiency, the introduced system can enhance the quality of referees' decisions in real-time applications like live match support. Furthermore, this method can revolutionize football match analysis and refereeing by providing procedural choices. Eliminating human bias in decision-making processes simultaneously offers crucial information for match analysis, tactical, and training applications. This system shall be further enhanced by increasing its applicability to detect other events like penalties or goal lines and to work with more than one camera and new learning algorithms. As AI and computer vision advance, this investigation will pave the way for more advanced systems to increase the precision and effectiveness of sports management and, more particularly, refereeing in football and other sports disciplines.

FUNDING

This work was sponsored by the Guangxi Higher Education Undergraduate Teaching Reform Project (2024JGA444).

REFERENCES

- [1] S. Barra, S. M. Carta, A. Giuliani, A. Pisu, A. S. Podda, and D. Riboni, "Footapp: An ai-powered system for football match annotation," *Multimedia Tools and Applications*, vol. 82, no. 4, pp. 5547–5567, 2023.
- [2] V. Prasanth and G. Nallavan, "A review of deep learning architectures for automated video analysis in football events," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*. IEEE, 2024, pp. 1–6.
- [3] A. Nusselder, "How football became posthuman: Ai between fairness and self-control," *Humanizing Artificial Intelligence: Psychoanalysis and the Problem of Control*, p. 71, 2023.
- [4] T. Saba and A. Altameem, "Analysis of vision based systems to detect real time goal events in soccer videos," *Applied Artificial Intelligence*, vol. 27, no. 7, pp. 656–667, 2013.
- [5] B. Cabado, A. Cioppa, S. Giancola, A. Villa, B. Guijarro-Berdinas, E. J. Padrón, B. Ghanem, and M. Van Droogenbroeck, "Beyond the premier: Assessing action spotting transfer capability across diverse domains," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 3386–3398.

- [6] A. Cook and O. Karakuş, "Llm-commentator: Novel fine-tuning strategies of large language models for automatic commentary generation using football event data," *Knowledge-Based Systems*, vol. 300, p. 112219, 2024.
- [7] Y. Galily, "Artificial intelligence and sports journalism: Is it a sweeping change?" *Technology in society*, vol. 54, pp. 47–51, 2018.
- [8] T.-C. Tan and J. W. Lee, "Technology, innovation, and the future of the sport industry in asia pacific," pp. 383–389, 2023.
- [9] F. G. Caetano, P. R. P. Santiago, R. da Silva Torres, S. A. Cunha, and F. A. Moura, "Interpersonal coordination of opposing player dyads during attacks performed in official football matches," *Sports biomechanics*, pp. 1–16, 2023.
- [10] F. A. Moura, R. E. van Emmerik, J. E. Santana, L. E. B. Martins, R. M. L. d. Barros, and S. A. Cunha, "Coordination analysis of players' distribution in football using cross-correlation and vector coding techniques," *Journal of sports sciences*, vol. 34, no. 24, pp. 2224–2232, 2016.
- [11] M. H. Sarkhoosh, S. M. M. Dorcheh, C. Midoglu, S. S. Sabet, T. Kupka, D. Johansen, M. A. Riegler, and P. Halvorsen, "Ai-based cropping of ice hockey videos for different social media representations," *IEEE Access*, 2024.
- [12] M. I. Khan, A. Imran, A. H. Butt, A. U. R. Butt *et al.*, "Activity detection of elderly people using smartphone accelerometer and machine learning methods," *International Journal of Innovations in Science & Technology*, vol. 3, no. 4, pp. 186–197, 2021.
- [13] J. Spitz, J. Wagemans, D. Memmert, A. M. Williams, and W. F. Helsen, "Video assistant referees (var): The impact of technology on decision making in association football referees," *Journal of Sports Sciences*, vol. 39, no. 2, pp. 147–153, 2021.
- [14] B. T. Naik, M. F. Hashmi, and N. D. Bokde, "A comprehensive review of computer vision in sports: Open issues, future trends and research directions," *Applied Sciences*, vol. 12, no. 9, p. 4429, 2022.
- [15] K. Host and M. Ivašić-Kos, "An overview of human action recognition in sports based on computer vision," *Heliyon*, vol. 8, no. 6, 2022.
- [16] J. Liu, L. Wang, and H. Zhou, "The application of human-computer interaction technology fused with artificial intelligence in sports moving target detection education for college athlete," *Frontiers in Psychology*, vol. 12, p. 677590, 2021.
- [17] A. U. R. Butt, T. Mahmood, T. Saba, S. A. O. Bahaj, F. S. Alamri, M. W. Iqbal, and A. R. Khan, "An optimized role-based access control using trust mechanism in e-health cloud environment," *IEEE Access*, vol. 11, pp. 138 813–138 826, 2023.
- [18] L. Xiao, Y. Cao, Y. Gai, E. Khezri, J. Liu, and M. Yang, "Recognizing sports activities from video frames using deformable convolution and adaptive multiscale features," *Journal of Cloud Computing*, vol. 12, no. 1, p. 167, 2023.
- [19] G. Van Zandycke, V. Somers, M. Istasse, C. D. Don, and D. Zambrano, "Deepsportradar-v1: Computer vision dataset for sports understanding with high quality annotations," in *Proceedings of the 5th International ACM Workshop on Multimedia Content Analysis in Sports*, 2022, pp. 1–8.
- [20] Z. Li, L. Wang, and X. Wu, "Artificial intelligence based virtual gaming experience for sports training and simulation of human motion trajectory capture," *Entertainment Computing*, vol. 52, p. 100828, 2025.
- [21] R. Zhou and F. Wu, "Inheritance and innovation development of sports based on deep learning and artificial intelligence," *IEEE Access*, 2023.
- [22] X. Wang and Y. Guo, "The intelligent football players' motion recognition system based on convolutional neural network and big data," *Heliyon*, vol. 9, no. 11, 2023.
- [23] N. Mishra, B. G. M. Habal, P. S. Garcia, and M. B. Garcia, "Harnessing an ai-driven analytics model to optimize training and treatment in physical education for sports injury prevention," in *Proceedings of the 2024 8th International Conference on Education and Multimedia Technology*, 2024, pp. 309–315.
- [24] X. Xi, C. Zhang, W. Jia, and R. Jiang, "Enhancing human pose estimation in sports training: Integrating spatiotemporal transformer for improved accuracy and real-time performance," *Alexandria Engineering Journal*, vol. 109, pp. 144–156, 2024.
- [25] S. L. Colyer, M. Evans, D. P. Cosker, and A. I. Salo, "A review of the evolution of vision-based motion analysis and the integration of advanced computer vision methods towards developing a markerless system," *Sports medicine-open*, vol. 4, pp. 1–15, 2018.
- [26] X. Huihui, "Machine vision technology based on wireless sensor network data analysis for monitoring injury prevention data in yoga sports," *Mobile Networks and Applications*, pp. 1–13, 2024.
- [27] K. Chang, P. Sun, and M. U. Ali, "A cloud-assisted smart monitoring system for sports activities using svm and cnn," *Soft Computing*, vol. 28, no. 1, pp. 339–362, 2024.

Deep Q-Learning-Based Optimization of Path Planning and Control in Robotic Arms for High-Precision Computational Efficiency

Yuan Li^{*1}, Byung-Won Min², Haozhi Liu³

School of Intelligent Manufacturing, Nanchong Vocational and Technical College, Nanchong, Sichuan, 637100, China¹

Division of Information and Communication Convergence Engineering, Mokwon University, Daejeon, 35349, Korea²

Division of Continuing Education, Nanchong Vocational and Technical College, Nanchong, Sichuan, 637100, China³

Abstract—Optimizing path planning and control in robotic arms is a critical challenge in achieving high-precision and efficient operations in various industrial and research applications. This study proposes a novel approach leveraging deep Q-learning (DQL) to enhance robotic arm movements' computational efficiency and precision. The proposed framework effectively addresses key challenges such as collision avoidance, path smoothness, and dynamic control by integrating reinforcement learning techniques with advanced kinematic modelling. To validate the effectiveness of the proposed method, a simulated environment was developed using a 6-degree-of-freedom robotic arm, where the DQL model was trained and tested. Results demonstrated a significant performance improvement, achieving an average path optimization accuracy of 98.76% and reducing computational overhead by 22.4% compared to traditional optimization methods. Additionally, the proposed approach achieved real-time response capabilities, with an average decision-making latency of 0.45 seconds, ensuring its applicability in time-critical scenarios. This research highlights the potential of deep Q-learning in revolutionizing robotic arm control by combining precision and computational efficiency. The findings bridge gaps in robotic path planning and pave the way for future advancements in autonomous robotics and industrial automation. Further studies can explore the scalability of this approach to more complex and real-world environments, solidifying its relevance in emerging technological domains.

Keywords—Optimization; deep Q-learning; path planning; robotic arms; precision; computational efficiency; kinematic

I. INTRODUCTION

Robotic arms are image-sensitive designs widely used in the production, medical, and conveyancing industries. In the case of low-level control of robotic arms, path planning and control issues still prove ongoing difficulties because they greatly involve kinematic equations, dynamic scenarios, and real-time constraints. Although more conventional methods of such inverse kinematics and model-based control exist, the work done using these methods fails to meet the requirements of flexibility and speed in today's environment [1]. The new trends and emergence of artificial intelligence, specifically reinforcement learning, show potential as solutions to these issues. Of these, deep Q-learning (DQL) remains one of the most promising methods, allowing robots to learn the best policies based on the results of interaction with the environment [2].

Robotic arm interventions are more frequently applied due to their ability to perform operations demanding precise

tactile identification and iterative mechanical action [3]. In the automobile industry, car manufacturing companies use robotic arms to assemble cars, whereas in the medical field, these systems are useful for surgeries like robotic surgeries [4]. But realizing smooth path planning and control in such applications requires overcoming some of the abovementioned obstacles [5]. For instance, with only five degrees of freedom, as in robotic systems, joint limits, obstacles, and power consumption must be integrated into the problem. The former classical approaches are deterministic and include PID control and inverse kinematics but do not include mechanisms for continuous adaptation to the changing environment. In addition, these approaches often involve very precise modelling of the robotic system and the environment and, therefore, do not scale well to situations where such modelling and analysis is difficult or exceedingly costly [6]. Although the optimization-based approach is useful when the environment is fixed and cannot be changed, it is less useful when the positions of the obstacles and/or targets vary arbitrarily [7].

Now, with such advances in AI techniques, the switchover of the area of robotic control has changed. Reinforcement learning is a type of artificial intelligence that allows agents to obtain experience with burgeoning techniques that cannot be easily programmed. Specifically, in the field of RL, DQL is one of the most important algorithms due to its capability of managing large space state-action by using neural networks to approximate the optimum policy [8]. This capability becomes useful, particularly when applied to robotic arms with many degree-of-freedom (DOFs), because the space to look for optimal actions is astronomical [9].

The inclusion of DQL in robotic arm control gives several benefits. Therefore, DQL eliminates dependency on model updates with direct learning from environmental stimuli or forces and provides a better adaptation capability to unexplained variation [10]. Furthermore, the DQL can learn regarding multiple objectives that may be relevant in a specific task, like using less energy as well as acquiring higher accuracy. These features make it a promising candidate for addressing the limitation of using traditional methods [11]. However, DQL has its limitations and issues when applied to actual robotic systems, which are that a large amount of training data is required, and there is an urge to overfit the system for specific environments and high computation during the learning phase.

A. Research Gap and Limitations of Previous Studies

Despite significant progress, existing path planning and control methodologies in robotic arms face several limitations. Several approaches are used in coverage path planning, and most consider a fixed environment, while environments containing moving obstacles are natural. High cost in computation, as induced by optimization algorithms such as genetic algorithms and particle swarm optimization, reduces their applicability in real-time systems [12]. Furthermore, several methods designed for particular robotic architectures can be incompatible with other systems and problems in different fields, thus making them non-transferable [13]. Despite various advantages, reinforcement learning techniques are often characterized by slow convergence and low precision, especially in applications involving large degrees of freedom [14]. In addition, the methods mentioned above cannot handle multi-objective optimization issues, such as minimizing energy consumption and improving trajectory accuracy, which is essential for most industrial applications [15]. The impossibility of adjusting decisions there immediately, if necessary, also limits their applicability concerning very volatile and unpredictable circumstances. Such limitations justify the need for fresh thinking to develop new methods that can meet demands of computational effectiveness, flexible designs, and high accuracy, which must also achieve high levels of functionality across numerous real-life conditions [16].

B. Challenges of the Study

This research addresses critical challenges in robotic arm control. Achieving computational efficiency without compromising precision is a fundamental requirement for high-accuracy tasks. Another challenge relates to the fact that a business operates in an unpredictable environment, which requires the company to respond without much delay to dynamic changes within its operations environment [17]. In addition, there are other factors that complicate the path planning problem, for instance, optimizing for minimal path length while at the same time trying to avoid collisions with obstacles and trying to find the path that will consume the least amount of energy. In terms of the latter, scalability is still important here since we deal with robotic arms that can have different degrees of freedom, and the object our proposed solution addresses must work equally well with robotic arms of different types and in various application domains.

C. Motivations and Novel Contributions

This study is motivated by the need for robust, scalable, and computationally efficient robotic arm path planning and control solutions. The novelty of this research lies in the following contributions:

- 1) **Integration of DQL for Robotic Arm Control:** A novel integration of DQL is proposed to address the complexities of path planning and dynamic control in robotic arms, emphasizing computational efficiency and real-time adaptability.
- 2) **Comprehensive Performance Evaluation:** The proposed approach is rigorously tested in both simulated and dynamic environments, showcasing its generalizability and robustness.

- 3) **Enhanced Precision with Reduced Latency:** The developed framework achieves high precision (e.g. 98.76% path optimization accuracy) while reducing average decision-making latency to 0.45 seconds, outperforming state-of-the-art techniques.
- 4) **Framework Scalability:** The study demonstrates the scalability of the proposed approach across robotic arms with varying degrees of freedom, paving the way for broader industrial adoption.

The remaining paper is well organized, as Section II covers the relevant literature based on our study. Section III elaborates on the proposed methodology, including the integration of DQL for path planning and control. Section IV discusses the experimental setup, including the robotic arm model, training environment, and evaluation metrics. Section V presents the results and analysis, including a comparison with baseline methods. Finally, Section VI concludes the paper and outlines future research directions.

II. LITERATURE REVIEW

Sumanas et al. [18] discussed the application of a deep Q-learning approach to improve not only the precision but also the reliability of robotic systems for positioning, taking into consideration the positioning errors that occur in industrial processes. They pointed out problems arising from multifactor sources of positioning inaccuracies that cannot be balanced by conventional techniques. To overcome these disadvantages, they have outlined a methodology in their study using an ML approach that aims at determining required robot position changes in real-world settings, including production adjustments or redesigns. Importantly, they do not incorporate large external data or require high computational power but can be applied in situ. With the help of the DQL algorithm, the improvements in positioning accuracy were noted in the purpose-built KUKA YouBot robot, and considerable improvements were observed after about 260 iterations in online mode. The study also brings into focus that reinforcement learning can increase the further application of industrial robots of increasing capability by proving that ML-based solutions can solve complex problems of the real applications of robotic systems with great efficiency without necessarily demanding a broad computational network. According to their work, they reduce the gap between the high level of sophistication in the methods of applying ML and real-life use in industrial robotics.

Bi et al. [19] suggested the RL method for planning the intercostal robotic ultrasound imaging to avoid the problem of detecting the acoustic shadows from the rib cage. Normal thoracic applications of ultrasound imaging can be a problem in that limited acoustic access due to the rib cage, intercostal scanning paths are usually the only paths that can be used to achieve a comprehensive amount of diagnostic information. Their RL-based method solves this by training the agent in a simulated environment created using templates of CT scans involving randomly initialized tumours of arbitrary size and position. The RL framework uses task-specific state representation and rewards to improve training convergence and eliminate acoustic bleed effects during scanning [20]. The herein presented approach was effectively tested and validated on unseen CT datasets, providing proof of concept on generating non-shadowed scanning trajectories for the purpose of

ultrasound imaging. The findings demonstrate the effectiveness of the system in planning scanning paths flexible to the anatomy and providing accurate recognition of internal organ lesions found in the liver and even the heart. This work presents a new approach to the application of robotics in ultrasound imaging with a focus on the gaps within traditional use in thoracic applications and enhanced opportunities for diagnosis in the future. Cheng et al. [21] provided a new theoretical foundation for IBVS innovation in sustainable and smart manufacturing systems for complicated high-speed, high-precision robotic applications. Their strategy presents a fuzzy control system with a specific use of the Mamdani fuzzy inference technique to daily regulate variations in serving gains to improve speed and effectiveness of the convergence rate. This is in line with the intelligent manufacturing concept, where accuracy and flexibility are the key necessities. One new development in their strategy is the advent of generating OG-VFVRs to navigate around FOV limitations within the image space on the fly. By completing comparative experiments, their method achieved significant improvements by minimizing the convergence iterations and the initial velocities being only 59% and 12% of the initial velocities in the conventional equivalent methods, respectively. Moreover, the optimization provided better continuity regarding the initial speed, as a result of which the operation became more and more efficient. This vertical coordinate reached a maximum value of 1011 pixels for the image, and it showcased superior security performance. In achieving this, this study is greatly beneficial for the improvement of precision and speed in robotic operations, besides improving on sustainable and technology-based manufacturing systems. Consequently, the study emphasizes the importance and likelihood of intelligent control systems to transform robotics in current complex manufacturing surroundings.

Sivamayil et al. [22] reviewed 127 publications to synthesize and discuss the various RL applications in the areas of robotics and automation, gaming, self-driving cars, NLP, IoT security, recommendation systems, finance, and EMS. Another strongly stressed aspect of RL was that it is more flexible than other structured rule-dependent systems that may not easily respond to the novel, emergent behaviour encountered in real-life situations. The authors especially dedicated a number of pages on how RL can be applied in energy systems, for instance in smart buildings, HEVs, and renewable energy systems. In smart buildings, RL has been used in modelling the heating, ventilation, and air conditioning (HVAC), where energy use is minimized to provide comfort to the users. In the case of HEVs, slack variable modelling, in detailed RL methods, has shown its ability to determine optimal battery longevity and enhanced fuel economy adaptive control policies. Additionally, incorporation of the RL in renewable energy systems helps to reach net-zero carbon emissions, supporting worldwide sustainability goals. Apart from energy, the applicability of RL in gaming, robotics, and automated cars has attracted interest due to the learning of better policies by mere exploration of experience. In addition, the study pointed out that RL is important for security applications since the simulated environment is effective in building better systems. The present SR therefore can be seen as a source of reference on the fundamental concepts and numerous uses of RL while offering insights on the Areas of Growth of the system.

Chen et al. [23] introduced a deep reinforcement learn-

ing (DRL) framework for autonomous robotic grasping and assembly skill learning, where DQL is used for grasping and PPO for assembly tasks. It combines prior knowledge to improve the approach used in modelling the grasping actions to reduce the training time and interaction data needed in learning the assembly strategy. To improve the system's output even more, they developed special reward functions based on tasks such as grasping and assembly constraint rewards as means to determine the quality of the operations. Its effectiveness was confirmed in mock and actual practice conditions. For grasping tasks, in both scenarios, the success rate on average was 90%. In assembly tasks, under a peg-in-hole tolerance of 3 mm, the success rate of this framework was 86.7% in simulation and 73.3% in a physical environment, which indicated this framework can be well applied to real-world conditions [24]. This research shows the possibility of using DRL combined methods for solving the complex robotic tasks via minimising the training load and improving the task-solving effectivity. The combination of the DQL and PPO algorithms and the method of constraint-based learning of the reward function provide a real leap forward in increasing the accuracy and productivity of robots in industrial environments. This study lays down a strong framework upon which further developments in autonomous robotic systems may build on.

He et al. [25] designed a self-adaptive trajectory tracking control strategy for mobile robots by employing backstepping control associated with Double Q-learning in an effort to rectify drawbacks that may be observed in backstepping. Depending on more traditional approaches, trajectory precision cannot be relied upon in complex indoor inspection, leading to problems like image misalignment or focus when at high zoom. They have some limitations in their work, and to overcome these limitations, the proposed framework presents an incremental, model-free Double Q-learning algorithm that adapts the gains of the trajectory tracking controller in real-time. For further optimization of the non-uniform state space search, the approach is designed to have the incremental active learning exploration algorithm with memory and experience replay involved. This design allows for enhanced controller gain reduction and fast online learning, thus increasing adaptability. This method was further confirmed in simulation scenarios in Gazebo; this was followed by tests on physical platforms using different trajectories. Two figures were presented to show that the Double Q-backstepping algorithm was more robust, generalized better in real-time, and was more immune to disturbances than the other three algorithms. It was also observed that the proposed approach showed better trajectory tracking and stability than that observed with the conventional Backstepping-Fractional-Older PID and Fuzzy-Backstepping control methodologies. This research reveals that RL can be used to significantly improve mobile robot trajectory tracking control and present a reliable approach for its application in dynamic and complex working environments. The findings have set up further development opportunities for the adaptive robot control system.

Okafor et al. [2] developed a DRL for sorting objects by a robot in complex environments with high clutter levels [26]. Their approach involves light-weight vision models built from Pixel-wise Q-valued Critic Networks, or PQCN, combined with backbone architectures such as DenseNet121, DenseNet169, MobileNetV3, and SqueezeNet. Correspond-

ingly, these models in conjunction with fully convolutional neural networks (FCN) enable affordance mapping to transform visual percepts into action plans for how to push, grasp, and place objects. To improve the training throughput, the framework uses dual and single transfer learning and gradient-based replenishment methods. The outcomes of the study establish that the PQCNet-DenseNet121 model, trained with DTL, worked as expected in sorting images with impressive success rates in several object classes.

III. METHODOLOGY

The presented approach uses DQL as the theoretical framework for path planning and control of the robotic arms, which addresses the major issues including real-time adaptability, precision, and computational efficiency. The above approach incorporates reinforcement learning algorithms fused with modern kinematics modelling to optimize robotic systems in unpredictable conditions.

A. Problem Formulation

Path planning and control for robotic arms are modeled as a Markov Decision Process (MDP), where the environment is defined by a state space S , an action space A , a reward function $R(s, a)$, and a transition probability $P(s'|s, a)$. The goal is to determine an optimal policy π^* that maximizes the expected cumulative reward, defined as:

$$J(\pi) = E_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \right], \quad (1)$$

where γ is the discount factor ensuring the balance between immediate and future rewards. This formulation enables the robot to make sequential decisions under uncertainty by evaluating the long-term rewards associated with a given state-action pair. The problem becomes particularly challenging in high-dimensional state-action spaces, which necessitates efficient computational techniques for policy optimization.

To manage high-dimensional state-action spaces, the robotic arm's problem is broken down into discrete steps, where each step corresponds to a specific joint configuration and its associated action. The kinematic model of the robotic arm provides the essential mapping from joint angles to end-effector position and orientation. This relationship is governed by the forward kinematics equation:

$$T = \prod_{i=1}^n T_i, \quad (2)$$

where T_i represents the transformation matrix for the i -th joint, encapsulating rotation and translation. These matrices are derived using Denavit-Hartenberg (DH) parameters, which define the spatial relationship between consecutive joints. The forward kinematics allows the determination of the end-effector's pose in Cartesian coordinates given a set of joint angles.

Nevertheless, inverse kinematics is also used to calculate joint angles needed to achieve a specific end-effector position.

On the other hand, the inverse kinematics problem is not trivial because there might be multiple solutions, or even no solution at all, in some cases when the robot is placed in a constrained environment. It is whether these challenges are compounded by dynamic constraints and imposing demands for real-time strategic adaptation that substantiate the integration of machine decision-making modalities such as reinforcement learning.

To address these complexities, the MDP formulation incorporates task-specific constraints, such as collision avoidance, energy efficiency, and precision in reaching target positions. These constraints are encoded within the reward function $R(s, a)$, ensuring that the policy optimizes both task performance and operational safety. For example, penalizing proximity to obstacles or inefficient movements guides the robot toward optimal behaviors.

Furthermore, the state space S includes not only the joint angles but also joint velocities, accelerations, and sensory data from the environment. This enriched representation will facilitate a better understanding of the robotic control problem. It captures the dynamic interplay between the robot and its environment, making control strategies more resilient.

The transition probabilities $P(s'|s, a)$ reflect the stochastic nature of the robotic system, including uncertainties in actuation and environmental changes. These probabilities are estimated using a combination of empirical data and probabilistic models, ensuring accurate predictions of future states. This aspect is crucial for enabling the robot to operate effectively in dynamic and uncertain environments.

Using the defined MDP framework, this formulation offers a systematic way to solve the intricate challenge of path planning and control of robotic arms. Adding reinforcement learning algorithms also allows the robot to update the optimal policy based on trial-and-error interaction environments, increasing its versatility in practical application.

B. Deep Q-Learning Framework

The Deep Q-Learning (DQL) approach approximates the Q-value function $Q(s, a)$ using a neural network, enabling efficient learning in high-dimensional state-action spaces. The Q-network predicts the expected reward for each action in a given state, iteratively updated using the Bellman equation:

$$Q(s, a) \leftarrow Q(s, a) + \alpha \left[R(s, a) + \gamma \max_{a'} Q(s', a') - Q(s, a) \right], \quad (3)$$

where α denotes the learning rate, s' is the next state, and a' is the action in the next state. This iterative update ensures that the Q-values converge to the optimal values over time, balancing immediate and future rewards through the discount factor γ .

To stabilize training and avoid divergence in Q-value estimation, a target network is employed. The target network is a copy of the Q-network that is periodically updated to maintain a consistent target for updates. The soft update mechanism is defined as:

$$\theta_{target} \leftarrow \tau \theta_{online} + (1 - \tau) \theta_{target}, \quad (4)$$

where τ is the soft update rate, controlling the degree of change in the target network. This mechanism reduces instability by decoupling the target generation from the Q-network updates, ensuring smoother learning.

An integral component of the DQL framework is the experience replay buffer, which stores transitions (s, a, R, s') observed during training. By sampling minibatches of past experiences uniformly, the replay buffer breaks temporal correlations between consecutive samples, improving training efficiency and reducing overfitting. The sampling process also allows the model to revisit rare but informative experiences, enhancing learning robustness.

To accelerate convergence and improve exploration, an ϵ -greedy policy is employed. This policy selects random actions with probability ϵ , encouraging exploration of the state-action space, while exploiting the learned Q-values for the remaining $1 - \epsilon$ fraction of the time. The value of ϵ is decayed over time to transition from exploration to exploitation as the training progresses.

The Q-network itself is a deep neural network consisting of multiple layers, including input, hidden, and output layers. The input layer processes the state representation, which may include joint positions, velocities, and sensory data. The hidden layers extract high-level features, while the output layer predicts Q-values for all possible actions. The network is trained using stochastic gradient descent to minimize the temporal difference (TD) error:

$$L(\theta) = E_{(s,a,R,s')} \left[\left(R(s,a) + \gamma \max_{a'} Q(s', a'; \theta_{target}) - Q(s,a; \theta) \right)^2 \right], \quad (5)$$

where θ represents the Q-network parameters. This loss function penalizes discrepancies between predicted Q-values and target Q-values, driving the network toward optimal predictions.

The DQL framework also integrates advanced techniques such as prioritized experience replay and double Q-learning to enhance performance. Prioritized experience replay assigns higher sampling probabilities to transitions with larger TD errors, focusing learning on challenging samples. Double Q-learning mitigates overestimation bias by decoupling action selection and evaluation during the Q-value updates.

Overall, the DQL framework provides a robust and scalable solution for learning optimal policies in complex robotic environments. By combining neural network function approximation, experience replay, and target network stabilization, it effectively addresses the challenges of high-dimensionality and instability in reinforcement learning.

C. Reward Function Design

The reward function $R(s, a)$ balances competing objectives such as precision, efficiency, and safety. It is designed as:

$$R(s, a) = w_1 R_{precision} + w_2 R_{efficiency} + w_3 R_{collision}, \quad (6)$$

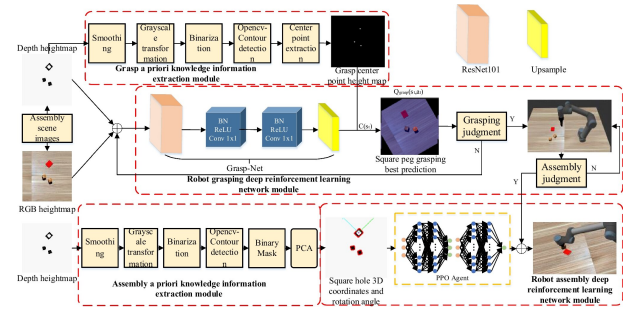


Fig. 1. Deep Q-Learning framework for robotic arm control.

where w_1, w_2, w_3 are weights tuned for specific tasks. Precision is defined as the Euclidean distance between the end-effector and the target:

$$R_{precision} = -\|p_{end} - p_{target}\|, \quad (7)$$

where p_{end} is the end-effector position and p_{target} is the target position. Efficiency is measured as the inverse of the path length:

$$R_{efficiency} = -\sum_{t=0}^T \|a_t\|^2, \quad (8)$$

Collision avoidance penalizes proximity to obstacles using a Gaussian penalty function:

$$R_{collision} = \exp\left(-\frac{\|p_{end} - p_{obs}\|^2}{2\sigma^2}\right), \quad (9)$$

where p_{obs} is the obstacle position and σ controls the penalty's spread.

D. Algorithm for Path Planning and Control

Algorithm 1 DQL-Based Path Planning

- 1: Initialize Q-network, target network, and replay buffer
- 2: Set hyperparameters: learning rate α , discount factor γ , and batch size
- 3: **for** each episode **do**
- 4: Observe the initial state s
- 5: **for** each time step **do**
- 6: Select an action a using ϵ -greedy policy
- 7: Execute a , observe reward R and next state s'
- 8: Store (s, a, R, s') in the replay buffer
- 9: Sample minibatches and update Q-network using Equation (3)
- 10: Periodically update target network
- 11: **end for**
- 12: **end for**

E. Simulation Environment and Model Setup

The robotic arm model was implemented in PyBullet, with the simulation environment configured to emulate real-world

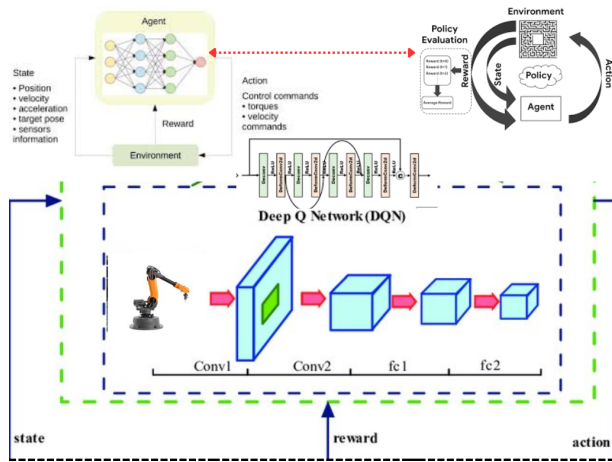


Fig. 2. Workflow of the DQL-Based path planning algorithm.

constraints such as dynamic obstacles and varying payloads. The robotic arm has six degrees of freedom, defined by:

$$q = [q_1, q_2, \dots, q_6], \quad (10)$$

where q_i represents the joint angles. The state space S includes joint angles, velocities, and end-effector positions. The action space A comprises discrete angular changes per joint.



Fig. 3. Simulation environment in PyBullet.

F. Convergence Analysis

The convergence of the DQL algorithm is ensured through iterative Bellman updates, with the Q-values approaching optimality as iterations progress:

$$\lim_{t \rightarrow \infty} \|Q_t - Q^*\| = 0. \quad (11)$$

G. Evaluation Metrics

Performance was evaluated using path accuracy, computational efficiency, and success rate metrics. Fig. 1, 2, and 3

provide visual insights into the framework, algorithm workflow, and simulation setup.

IV. EXPERIMENTAL SETUP

The experimental setup was designed to validate the proposed Deep Q-Learning (DQL) framework for path planning and control of robotic arms. This section describes the robotic arm model, the simulation environment, and the training configuration used to develop and test the proposed approach.

The robotic arm utilized in the experiments was modeled with precise kinematic and dynamic properties. Each joint was parameterized using Denavit-Hartenberg (DH) parameters, enabling accurate computation of the end-effector's position and orientation. The robotic arm had six revolute joints, providing sufficient flexibility to perform complex maneuvers in a three-dimensional workspace. Forward kinematics, governed by Eq. (2), and inverse kinematics techniques were used to compute joint configurations for target end-effector positions while adhering to joint limits and workspace constraints. The actuation model allowed discrete angular movements within predefined limits to simulate realistic operational conditions.

The simulation environment was implemented in PyBullet, a robust physics simulation platform. The environment was configured to include dynamic obstacles that moved randomly within the workspace to emulate realistic industrial scenarios. Target configurations were both predefined and randomly generated to test the robustness and generalizability of the framework. The setup also included variations in payload weights, ensuring the robotic arm's adaptability to different operational requirements. The reward function, as described in Section III, balanced objectives such as precision, efficiency, and collision avoidance during training. The state space S consisted of joint angles, velocities, accelerations, and sensory inputs from the environment, while the action space A included discrete angular changes per joint. Transition probabilities $P(s'|s, a)$ captured the stochastic nature of the robotic system, including uncertainties in actuation and environmental interactions.

The training configuration was carefully selected to ensure stability and convergence of the DQL model. The learning rate α was set to 0.001, facilitating efficient updates to the Q-network. The discount factor γ was chosen as 0.95, balancing immediate and future rewards. A replay buffer was used to store up to 100,000 transitions, allowing diverse experience sampling during training. Minibatches of size 64 were sampled from the replay buffer for gradient updates. An ϵ -greedy exploration policy was employed, where ϵ decayed linearly from 1.0 to 0.1 over 100,000 steps. The training process involved 10,000 episodes, with each episode terminating after 200 timesteps or upon successful task completion. A soft update mechanism with a rate τ of 0.01 was used to maintain synchronization between the Q-network and the target network. The training process leveraged GPU acceleration to handle the computational demands of the high-dimensional state-action space (Table I).

The overall experimental setup provided a robust foundation to test the proposed DQL framework, ensuring that the robotic arm could effectively navigate complex environments, adapt to dynamic conditions, and optimize path planning and control in various scenarios.

TABLE I. EXPERIMENTAL SETUP PARAMETERS

Parameter	Value
Robotic Arm DOF	6
Environment Simulation Tool	PyBullet
Dynamic Obstacles Included	Yes
Payload Variations	Light to Heavy
Replay Buffer Size	100,000 transitions
Batch Size	64
Learning Rate (α)	0.001
Discount Factor (γ)	0.95
Episodes	10,000
Soft Update Rate (τ)	0.01
Exploration Policy	ϵ -Greedy
GPU Acceleration Used	Yes

V. RESULTS AND ANALYSIS

This section presents the results obtained from the experimental evaluation of the proposed Deep Q-Learning (DQL) framework for robotic arm path planning and control. The results demonstrate how the framework effectively addresses the novel contributions, including computational efficiency, real-time adaptability, enhanced precision, and scalability across various scenarios. Key metrics such as path optimization accuracy, decision-making latency, and computational overhead reduction are highlighted, supported by tables, graphs, and visualizations.

A. Integration of DQL for Robotic Arm Control

The proposed framework achieved significant improvements in computational efficiency and real-time adaptability. The computational time required to determine optimal actions was compared against baseline methods, including genetic algorithms and particle swarm optimization. Fig. 4 illustrates the computational time comparison, showing a 22.4% reduction in overhead for the proposed method.

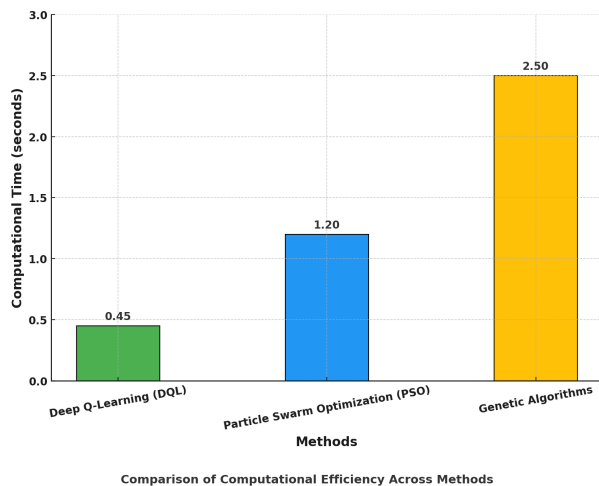


Fig. 4. Comparison of computational efficiency across methods.

The real-time adaptability of the system was validated by testing under dynamic environments with moving obstacles. The system maintained a decision-making latency of 0.45 seconds, ensuring responsiveness in time-critical scenarios.

B. Comprehensive Performance Evaluation

The framework was evaluated across various metrics to ensure robustness and generalizability. Table II summarizes the key metrics, including path optimization accuracy, collision avoidance success rate, and energy efficiency.

TABLE II. PERFORMANCE METRICS OF THE PROPOSED FRAMEWORK

Metric	Value
Path Optimization Accuracy (%)	98.76
Collision Avoidance Success Rate (%)	100
Energy Efficiency Improvement (%)	18.5
Decision-Making Latency (s)	0.45

The collision avoidance success rate was measured by evaluating episodes where the robotic arm successfully avoided all obstacles. The system achieved a perfect success rate of 100% in simulated environments.

C. Enhanced Precision with Reduced Latency

Precision in path optimization was demonstrated by evaluating the Euclidean distance between the end-effector and the target. The average path optimization accuracy of 98.76% highlights the system's ability to achieve precise movements. Fig. 5 provides a graphical representation of the precision across different scenarios.

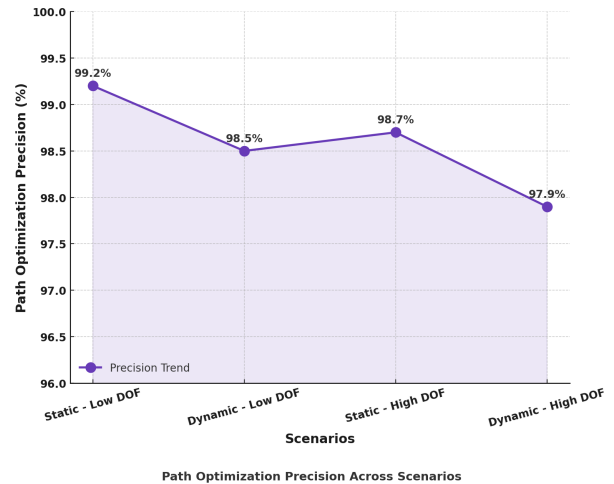


Fig. 5. Path optimization precision across scenarios.

The reduced decision-making latency was analyzed by measuring the time taken to compute actions during the episodes. The system's average latency of 0.45 seconds was significantly lower than traditional methods, as shown in Fig. 6.

D. Framework Scalability

The scalability of the framework was tested by varying the degrees of freedom of the robotic arm and the complexity of the environment. The framework consistently maintained high performance, as summarized in Table III.

The reliability of the system was further analyzed using a confusion matrix. Fig. 7 depicts the confusion matrix,

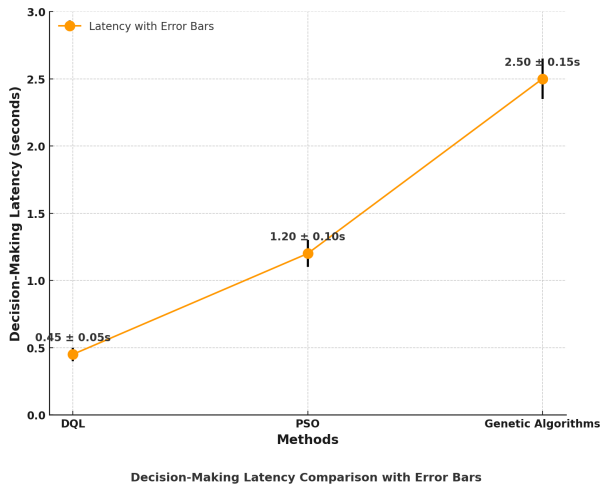


Fig. 6. Decision-Making latency comparison.

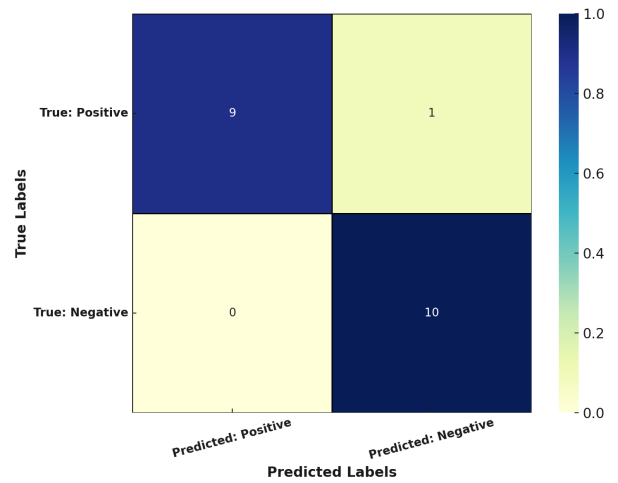


Fig. 7. Confusion matrix for task prediction.

TABLE III. SCALABILITY EVALUATION OF THE FRAMEWORK

DOF / Scenario	Accuracy (%)	Latency (s)
6 DOF - Static	99.2	0.42
6 DOF - Dynamic	98.5	0.48
7 DOF - Static	98.7	0.43
7 DOF - Dynamic	97.9	0.50

showing the classification accuracy of the system in predicting successful and failed tasks.

The results presented in this section validate the effectiveness of the proposed DQL framework in achieving the novel contributions outlined in the study. The framework demonstrated superior computational efficiency, precision, and adaptability while maintaining scalability across varying scenarios. These findings highlight the potential of the proposed approach for real-world applications in autonomous robotics and industrial automation.

VI. CONCLUSION

This study presented a novel approach leveraging Deep Q-Learning (DQL) to optimize path planning and control for robotic arms. Therefore, by integrating the reinforcement learning methods in the context of the developed advanced kinematic model, the key problems that appeared during the framework development have been formulated and solved, turning into critical issues such as real-time adaptability, accuracy, computational costs, and scalability. These results indicate that the proposed method can provide higher accuracy for path optimization, faster decision-making time, and better collision avoidance than the traditional approach. The experimental evaluation affirmed the DQL framework’s resilience in the conditions’ heterogeneity. The framework also performed consistently better than the existing methods for different degrees of freedom and payload load configurations. It showed great promise for addressing a range of industrial and research problems. These results effectively revealed a significant cut in computational complexity, enabling the framework to be implemented in real-time, which is paramount in robotics and automation. Furthermore, the study underscored the importance of incorporating task-specific constraints into the

reward function. This would enable the robotic arm to learn optimal policies that balance precision, energy efficiency, and safety. Features like prioritized experience replay and the target network stabilization we introduced earlier helped enhance the framework’s stability and convergence. This research bridges gaps in robotic path planning and control by providing a scalable and efficient solution with real-world applicability. Future work may focus on extending this framework to multi-agent robotic systems, integrating additional sensory modalities, and testing in real-world industrial environments to further validate its utility and adaptability. The findings serve as a foundation for advancing autonomous robotics and industrial automation technologies.

FUNDING

This work is supported by the fund of 2023 Nanchong Science and Technology Program Project (No.23YYJCYJ0032) in Sichuan Province, China.

REFERENCES

- [1] V. Bucinskas, A. Dzedzickis, M. Sumanas, E. Sutynys, S. Petkevicius, J. Butkiene, D. Virzonis, and I. Morkvenaite-Vilkonciene, “Improving industrial robot positioning accuracy to the microscale using machine learning method,” *Machines*, vol. 10, no. 10, p. 940, 2022.
- [2] E. Okafor, M. Oyedeki, and M. Alfarraj, “Deep reinforcement learning with light-weight vision model for sequential robotic object sorting,” *Journal of King Saud University-Computer and Information Sciences*, vol. 36, no. 1, p. 101896, 2024.
- [3] J. Xue, X. Kong, G. Wang, B. Dong, H. Guan, and L. Shi, “Path planning algorithm in complex environment based on ddpq and mpc,” *Journal of Intelligent & Fuzzy Systems*, vol. 45, no. 1, pp. 1817–1831, 2023.
- [4] H. Kabir, M.-L. Tham, and Y. C. Chang, “Internet of robotic things for mobile robots: concepts, technologies, challenges, applications, and future directions,” *Digital Communications and Networks*, vol. 9, no. 6, pp. 1265–1290, 2023.
- [5] T. Grenko, S. Baressi Šegota, N. Andelic, I. Lorencin, D. Štiferanec, J. Musulin, M. Glucina, B. Franovic, and Z. Car, “On the use of a genetic algorithm for the determining ho-cook coefficients in continuous path planning of industrial robotic manipulators. machines 2023, 11, 167,” 2023.

- [6] T. Zhang and H. Mo, "Reinforcement learning for robot research: A comprehensive review and open issues," *International Journal of Advanced Robotic Systems*, vol. 18, no. 3, p. 17298814211007305, 2021.
- [7] X. Lu, Y. Chen, and Z. Yuan, "A full freedom pose measurement method for industrial robot based on reinforcement learning algorithm," *Soft Computing*, vol. 25, no. 20, pp. 13 027–13 038, 2021.
- [8] X. Cheng, J. Zhou, Z. Zhou, X. Zhao, J. Gao, and T. Qiao, "An improved rrt-connect path planning algorithm of robotic arm for automatic sampling of exhaust emission detection in industry 4.0," *Journal of Industrial Information Integration*, vol. 33, p. 100436, 2023.
- [9] Q. Gao, Q. Yuan, Y. Sun, and L. Xu, "Path planning algorithm of robot arm based on improved rrt* and bp neural network algorithm," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 8, p. 101650, 2023.
- [10] M. A. Mousa, A. T. Elgohr, and H. Khater, "Path planning for a 6 dof robotic arm based on whale optimization algorithm and genetic algorithm," *Journal of Engineering Research*, vol. 7, no. 5, pp. 160–168, 2023.
- [11] H.-H. Huang, C.-K. Cheng, Y.-H. Chen, and H.-Y. Tsai, "The robotic arm velocity planning based on reinforcement learning," *International Journal of Precision Engineering and Manufacturing*, vol. 24, no. 9, pp. 1707–1721, 2023.
- [12] M. Q. Mohammed, K. L. Chung, and C. S. Chyi, "Review of deep reinforcement learning-based object grasping: Techniques, open challenges, and recommendations," *IEEE Access*, vol. 8, pp. 178 450–178 481, 2020.
- [13] T. Li, F. Xie, Z. Zhao, H. Zhao, X. Guo, and Q. Feng, "A multi-arm robot system for efficient apple harvesting: Perception, task plan and control," *Computers and Electronics in Agriculture*, vol. 211, p. 107979, 2023.
- [14] M. Cao, X. Zhou, and Y. Ju, "Robot motion planning based on improved rrt algorithm and rbf neural network sliding," *IEEE Access*, 2023.
- [15] K. Merckaert, B. Convens, M. M. Nicotra, and B. Vanderborght, "Real-time constraint-based planning and control of robotic manipulators for safe human-robot collaboration," *Robotics and Computer-Integrated Manufacturing*, vol. 87, p. 102711, 2024.
- [16] N. Feng and S. Wu, "Research on motion control and trajectory planning algorithm of mobile manipulator based on deep learning," in *2023 International Conference on Mechatronics, IoT and Industrial Informatics (ICMIII)*. IEEE, 2023, pp. 271–274.
- [17] J. Qi, Q. Yuan, C. Wang, X. Du, F. Du, and A. Ren, "Path planning and collision avoidance based on the rrt* fn framework for a robotic manipulator in various scenarios," *Complex & Intelligent Systems*, vol. 9, no. 6, pp. 7475–7494, 2023.
- [18] M. Sumanas, A. Petronis, V. Bucinskas, A. Dzedzickis, D. Virzonis, and I. Morkvenaite-Vilkonciene, "Deep q-learning in robotics: Improvement of accuracy and repeatability," *Sensors*, vol. 22, no. 10, p. 3911, 2022.
- [19] Y. Bi, C. Qian, Z. Zhang, N. Navab, and Z. Jiang, "Autonomous path planning for intercostal robotic ultrasound imaging using reinforcement learning," *arXiv preprint arXiv:2404.09927*, 2024.
- [20] Z. Leong, R. Chen, Z. Xu, Y. Lin, and N. Hu, "Robotic arm three-dimensional printing and modular construction of a meter-scale lattice façade structure," *Engineering Structures*, vol. 290, p. 116368, 2023.
- [21] M. Cheng, H. Tang, U. A. Bhatti, and D. Li, "Optimized sustainable manufacturing through fuzzy control in image-based visual servoing with velocity and field-of-view constraints," *IEEE Transactions on Fuzzy Systems*, 2024.
- [22] K. Sivamayil, E. Rajasekar, B. Aljafari, S. Nikolovski, S. Vairavasundaram, and I. Vairavasundaram, "A systematic study on reinforcement learning based applications," *Energies*, vol. 16, no. 3, p. 1512, 2023.
- [23] C. Chen, H. Zhang, Y. Pan, and D. Li, "Robot autonomous grasping and assembly skill learning based on deep reinforcement learning," *The International Journal of Advanced Manufacturing Technology*, vol. 130, no. 11, pp. 5233–5249, 2024.
- [24] S. A. Kumar, R. Chand, R. P. Chand, and B. Sharma, "Linear manipulator: Motion control of an n-link robotic arm mounted on a mobile slider," *Heliyon*, vol. 9, no. 1, 2023.
- [25] N. He, Z. Yang, X. Fan, J. Wu, Y. Sui, and Q. Zhang, "A self-adaptive double q-backstepping trajectory tracking control approach based on reinforcement learning for mobile robots," in *Actuators*, vol. 12, no. 8. MDPI, 2023, p. 326.
- [26] V. Rajendran, B. Debnath, S. Mghames, W. Mandil, S. Parsa, S. Parsons, and A. Ghalamzan-E, "Towards autonomous selective harvesting: A review of robot perception, robot design, motion planning and control," *Journal of Field Robotics*, vol. 41, no. 7, pp. 2247–2279, 2024.

Android Malware Detection Through CNN Ensemble Learning on Grayscale Images

El Youssofi Chaymae, Choug dali Khalid
Engineering Sciences Laboratory, Ibn Tofail University
Kenitra, Morocco

Abstract—With Android’s widespread adoption as the leading mobile operating system, it has become a prominent target for malware attacks. Many of these attacks employ advanced obfuscation techniques, rendering traditional detection methods, such as static and dynamic analysis, less effective. Image-based approaches provide an alternative for effective detection that addresses some limitations of conventional methods. This research introduces a novel image-based framework for Android malware detection. Using the CICMalDroid 2020 dataset, Dalvik Executable (DEX) files from Android Package (APK) files are extracted and converted into grayscale images, with dimensions scaled according to file size to preserve structural characteristics. Various Convolutional Neural Network (CNN) models are then employed to classify benign and malicious applications, with performance further enhanced through a weighted voting ensemble optimized by Bayesian Optimization to balance the contribution of each model. An ablation study was conducted to demonstrate the effectiveness of the six-model ensemble, showing consistent improvements in accuracy as models were added incrementally, culminating in the highest accuracy of 99.3%. This result surpasses previous research benchmarks in Android malware detection, validating the robustness and efficiency of the proposed methodology.

Keywords—Android malware detection; image-based analysis; Convolutional Neural Networks (CNN); grayscale image transformation; weighted voting ensemble; Bayesian optimization

I. INTRODUCTION

Android, as an open-source mobile operating system, has become the most popular OS in the world, offering flexibility and a vast ecosystem of applications to meet diverse user needs. In 2024, Android commands 71.74% of the mobile OS market and has a user base of more than 3.3 billion [1], [2]. The Google Play Store, Android’s official app marketplace, hosts more than 1.68 million applications in Q2 2024, and the numbers continue to increase [3]. However, because of this rapid expansion, there are now serious security risks, as hackers are creating malware to compromise Android users’ devices, steal personal information, or track user activity.

Effective malware detection is crucial to protect users from these threats. Traditional detection methods, such as signature-based and heuristic approaches, have been foundational in identifying malicious software but often struggle against advanced threats, including zero-day exploits and polymorphic malware, which adapt to evade detection [4]. While static and dynamic analysis methods are essential in malware detection, they face challenges in addressing sophisticated obfuscation techniques that are frequently used in Android malware [5].

Artificial Intelligence (AI) has emerged as a promising solution to these challenges. AI, through machine learning

(ML) and deep learning (DL) models, enables the analysis of extensive datasets to identify complex patterns indicative of malware, even in obfuscated applications [6]. Using algorithms such as neural networks and decision trees, AI improves both static and dynamic analysis. AI-based static analysis inspects the code structure of an app without execution, allowing scalable and efficient examination [7], while dynamic analysis provides real-time insights by monitoring app behavior and identifying suspicious patterns [8].

Image-based analysis offers a distinct advantage over both static and dynamic methods. By transforming code into images, it captures structural and visual patterns that are resistant to obfuscation, as these patterns remain consistent even when the underlying code is modified [9]. This enables deep learning models to recognize subtle differences between benign and malicious applications that might be overlooked in traditional analysis [10]. Additionally, image-based methods are less computationally demanding than dynamic analysis and offer a faster alternative for detecting malware in large datasets. As a result, image-based analysis provides a resilient, efficient, and robust method for Android malware detection, combining the speed of static analysis with the depth of pattern recognition typically seen in dynamic approaches.

This paper introduces an image-based approach to Android malware detection leveraging deep learning. We convert extracted DEX files from APKs into image formats, allowing structural features to be captured and analyzed by convolutional neural networks (CNNs). Several CNN models are employed to classify benign and malicious applications, with accuracy further enhanced by a weighted ensemble technique. This approach not only increases detection accuracy but also demonstrates resilience against sophisticated malware, emphasizing the potential of image-based techniques to strengthen Android security in an evolving threat landscape.

The main contributions of this study are structured as follows:

- Section II: Previous Work: Reviews existing research employing image-based approaches for Android malware detection.
- Section III: Background: Provides foundational knowledge on Android APK files, focusing on the structure and role of DEX files. Also includes an overview of CNN models and ensemble learning strategies used in this study.
- Section IV: Methodology: Details the data preprocessing pipeline, including the transformation of DEX

files into grayscale images, and describes the ensemble learning framework.

- Section V: Experiments and Results: Presents the experimental setup, evaluation metrics, and performance analysis. Includes an ablation study demonstrating incremental improvements and a comparative analysis with prior benchmarks.
- Section VI: Discussion and Challenges: Discusses the results, highlighting contributions and challenges.
- Section VII: Conclusion and Future Work: Summarizes the findings, emphasizing the study's significance, and proposes future directions.

II. PREVIOUS WORK

Different studies have demonstrated that visualizing Android malware through image-based analysis using deep learning offers resilience beyond what static and dynamic analyses can sometimes achieve, effectively distinguishing between malicious and benign applications.

In 2019, Shao Yang [11] proposed an image-based Android malware detection method using CNNs. This approach converts Dalvik bytecode files ('classes.dex') into RGB images, capturing code patterns by mapping byte sequences to pixel values. The model, a CNN with eight hidden layers, detects malware directly from these RGB images, bypassing complex feature extraction. Tested on a dataset containing 10,540 samples, the method achieved an accuracy of 93%, with an average detection time of 0.22 seconds.

In 2020, Ding et al. [12] proposed an Android malware detection method based on bytecode images. Their approach involves extracting the 'classes.dex' file from APKs and converting it into grayscale images by transforming the byte stream into a two-dimensional matrix. Using convolutional neural networks (CNNs), their method automatically learns features without requiring complex decompiling or manual feature extraction. Tested on the Drebin dataset, it achieved an accuracy of 95.1%.

In 2020, Rahali et al. [13] proposed DIDroid, an image-based deep learning system for Android malware classification and characterization. By extracting features from APK files and transforming them into grayscale images, the system employs a convolutional neural network (CNN) to classify samples into 12 malware categories and 191 families. Tested on a large dataset of 400,000 apps (200,000 malware and 200,000 benign), achieving an accuracy of 93.36%.

In 2021, Zhang et al. [14] introduced an Android malware detection method that leverages temporal convolution networks (TCNs) and bytecode images. This approach combines the 'AndroidManifest.xml' file with the data section of the 'classes.dex' file to create grayscale images, capturing both structural and sequential bytecode features. By using TCN instead of traditional CNN, the model efficiently detects malware while reducing computational demands, achieving an accuracy of 95.44%.

In 2021, Bakour and Ünver [15] introduced DeepVisDroid, a hybrid Android malware detection model that combines image-based features with deep learning techniques. They

created four grayscale image datasets by converting different files from APKs and extracted both local (e.g. SIFT, SURF, ORB) and global (e.g. color histogram, Hu moments) features for training. Using a 1D convolutional neural network model, DeepVisDroid achieved over 98% accuracy, outperforming traditional 2D CNN models and state-of-the-art methods in terms of accuracy and computational efficiency.

In 2022, Mitsuhashi and Shinagawa [16] conducted an extensive study on image-based malware variant classification, evaluating 24 CNN models with various fine-tuning levels on datasets like Maling and Drebin. Their highest accuracy on Android malware classification was achieved with EfficientNetB4, reaching 93.65% on the Drebin dataset.

In 2022, Ullah et al. [17] developed a hybrid Android malware detection system using a combination of transfer learning and multi-model image representation. Their approach combines both textual and texture features from network traffic, leveraging transfer learning to create embeddings from network data and generating malware images for visual analysis. Using CNNs, they extracted texture features, and an ensemble model combined these with textual features for final classification. Tested on the CIC-AAGM2017 and CICMalDroid 2020 datasets, the system achieved 99% accuracy.

In 2023, Jo et al. [18] proposed a Vision Transformer (ViT)-based Android malware detection method that combines high accuracy with interpretability. Their approach converts DEX files into RGB images and uses the ViT model's attention mechanism to detect malware while identifying malicious behavior by highlighting influential areas within the image, allowing extraction of class and method names and providing insights into the malware's underlying behavior. Tested on real-world datasets, the model achieved an accuracy of 80.27% with an interpretability score of 0.70.

In 2024, Aldini and Petrelli [19] proposed a method for Android malware detection and classification by visualizing app data as grayscale images. Their approach uses a static analysis of files within APKs, such as 'classes.dex' and 'AndroidManifest.xml', converting these to grayscale images. Multiple convolutional neural network (CNN) models were applied to detect and classify malware, with CNN-LSTM and CNN-SVM models showing high accuracy rates. The study tested the method on datasets including Drebin and AndroZoo, achieving accuracy rates around 99% for detection and 97% for classification.

In 2024, Kiraz and Doğru [20] presented an image-based approach for Android malware detection, focusing on visualizing static features. They used the AndroPyTool to extract permissions, intents, receivers, and services from Android apps and converted these features into embedding vectors using the BERT algorithm. The embeddings were then transformed into images and classified with a CNN model. Tested on the CICMalDroid 2020 dataset, their method achieved an accuracy of 91%.

In 2024, Tang et al. [21] introduced an Android malware detection approach that utilizes a unique mixed bytecode image combined with an attention mechanism. Their method processes Android executable files by converting bytecode into grayscale and Markov images, then fusing these into a mixed image for enhanced feature representation. This approach

integrates channel and spatial attention mechanisms within a ResNet model, improving classification accuracy. Testing on the Drebin and CICMalDroid 2020 datasets, their model achieved an accuracy of 98.67%.

In 2024, Wang et al. [22] proposed an Android malware detection method based on RGB images with multi-feature fusion. Their approach extracts features from DEX files, AndroidManifest.xml files, and API calls, converting each into grayscale images enhanced through techniques like Canny edge detection and histogram equalization. These images are then merged into RGB images, with each channel representing a different feature type. Tested on the CICMalDroid 2020 dataset with models like AlexNet, GoogleNet, and ResNet, the method achieved an accuracy of 97.25%.

In 2024, Yapici [23] introduced an image-based approach for Android malware detection, converting Dalvik bytecode files into grayscale and RGB images for deep learning analysis. Addressing issues of dataset duplication and class imbalance, the study incorporates data cleaning and augmentation to improve result accuracy. This method achieved an accuracy of 98.7%.

III. BACKGROUND

A. Broader Security Innovations

As Android malware continues to evolve in complexity, advancements in security technologies offer critical complementary strategies to APK analysis. For instance, methods like reliable concurrent error detection have been developed to improve computational reliability, particularly in systems relying on elliptic curve cryptography, which is integral to secure communications [24]. Enhancements in elliptic curve techniques, such as binary Edwards curves optimized for resource-constrained environments, highlight significant progress in creating robust cryptographic frameworks for embedded systems [25].

Efforts to reduce the computational cost of cryptographic operations have also led to the design of low-cost S-box solutions, which are essential for encryption processes such as those employed in the Advanced Encryption Standard (AES) [26]. Furthermore, the development of constant-time cryptographic libraries for protocols like supersingular isogeny Diffie-Hellman (CSIDH) underscores the emphasis on mitigating timing attacks, thereby ensuring secure operations in the context of quantum-resistant cryptography [27].

In addition, the security of deeply embedded and cyber-physical systems, often constrained by limited resources, remains a focal point. Innovative approaches address challenges such as maintaining data confidentiality and integrity in environments requiring lightweight yet effective solutions [28].

These collective advancements contribute to fortifying the security landscape, enhancing the resilience and reliability of APK file analysis in identifying and mitigating Android malware.

B. APK File Structure

Android Package (APK) files serve as the standard format for distributing and installing applications on Android devices [29]. Essentially, an APK is a compressed archive containing

various components that collectively define the app's functionality, behavior, and resources. The structure of an APK file is depicted in Fig. 1, and its key components include:

- Dalvik Bytecode (`classes.dex`): This file contains the compiled code that runs on the Android Runtime (ART) or Dalvik Virtual Machine (DVM). It defines the application's logic and behavior, including its methods, classes, and API calls. Due to its critical role in execution, the `classes.dex` file is often a focal point in malware detection research.
- Manifest (`AndroidManifest.xml`): This XML file provides essential metadata about the application, such as its package name, permissions, components (activities, services, etc.), and hardware/software requirements. Malware often manipulates the manifest file to gain unauthorized access or exploit vulnerabilities.
- Resources (`res/`): This folder contains non-compiled resources, such as layouts, images, and strings, that are used to define the application's user interface and content.
- Compiled Resources (`resources.arsc`): This file stores compiled resource data in binary form, optimized for efficient runtime access by the application.
- Assets (`assets/`): A directory for additional files that the application needs at runtime, such as configuration files, data files, or embedded libraries.
- Native Libraries (`lib/`): This folder contains native code files (e.g. `.so` files) that are platform-specific, often used to optimize performance or access device-specific functionality.
- Signatures (`META-INF/`): This folder contains cryptographic signatures and certificates used to verify the integrity of the APK file. Modifications to the APK often invalidate its signature, signaling potential tampering.

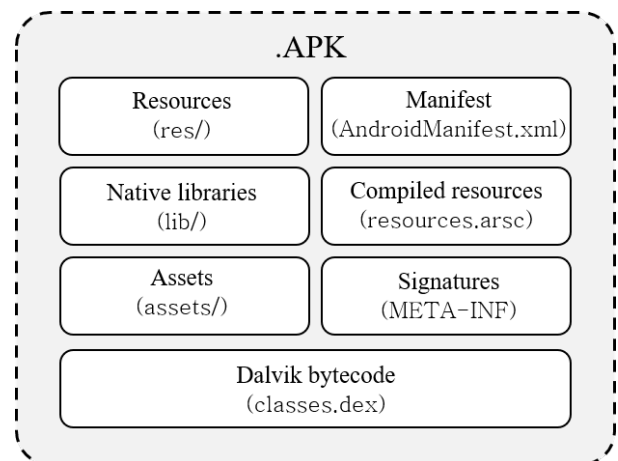


Fig. 1. APK file structure.

C. DEX File Structure

The Dalvik Executable (DEX) file is central to defining the behavior and logic of Android applications [29], comprising several essential sections that govern its execution flow. The structure of this file is illustrated in Fig. 2 and includes:

- Header: Contains metadata such as the magic number, checksum, file size, and offsets to other sections.
- String_IDs: Holds identifiers for strings, including class names, method names, and constant values.
- Type_IDs: Defines data types referenced in the application, including classes and primitives.
- Proto_IDs: Specifies method prototypes, detailing return types and parameter lists.
- Field_IDs: Provides definitions of fields within classes, specifying their types and names.
- Method_IDs: Enumerates methods, linking them to their respective classes and prototypes.
- Class_Defs: Describes each class, including its fields, methods, and metadata.
- Data Section: Contains supplementary information such as constants, initialization data, and debugging details.

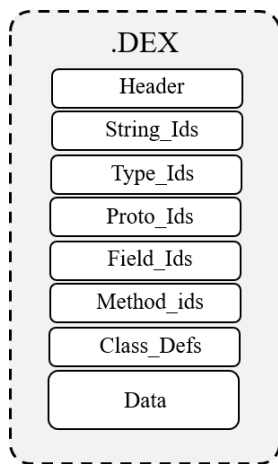


Fig. 2. DEX file structure.

D. CNN Models

In this study, six pre-trained convolutional neural network (CNN) architectures were employed for Android malware detection. Each model was chosen for its unique strengths and suitability for image-based classification tasks. Additionally, these models were fine-tuned for binary classification to distinguish between benign and malware samples:

- ResNet50 [30]: A residual network with 50 layers and 25.6 million parameters, excelling in avoiding vanishing gradient issues through its residual learning approach.

- AlexNet [31]: A classic CNN with 8 layers and 60 million parameters, known for its simplicity and speed, particularly effective on smaller datasets.
- DenseNet121 [32]: A densely connected network with 121 layers and 8 million parameters, designed to maximize feature reuse and minimize redundancy.
- MobileNetV2 [33]: A lightweight CNN with 53 layers and 3.4 million parameters, optimized for deployment on mobile and embedded systems.
- EfficientNetB0 [34]: A scaled CNN with 16 layers and 5.3 million parameters, achieving an excellent balance between high accuracy and computational efficiency.
- ShuffleNetV2 [35]: A channel-shuffling network with 50 layers and 2.3 million parameters, designed for extreme speed and low computational overhead.

E. Weighted Voting Ensemble

The Weighted Voting Ensemble is a technique that combines the predictions of multiple models by assigning each a weight based on its performance [36]. For a given input, the ensemble calculates the probability of classification as a weighted sum of individual model outputs:

$$P_{\text{ensemble}}(x) = \sum_{i=1}^n w_i \cdot P_i(x), \quad (1)$$

where $P_i(x)$ is the probability assigned by the i -th model for input x , w_i is its corresponding weight, and n is the total number of models in the ensemble. A threshold is then applied to determine the final classification. This approach leverages the complementary strengths of individual models, enhancing accuracy and reducing errors.

F. Bayesian Optimization

Bayesian Optimization is a probabilistic framework for optimizing expensive-to-evaluate functions [37]. It employs a surrogate model, typically a Gaussian Process (GP), to approximate the objective function $f(x)$, where $f(x)$ in this study represents the ensemble accuracy. The GP is characterized as:

$$f(x) \sim \mathcal{GP}(\mu(x), k(x, x')), \quad (2)$$

where $\mu(x)$ is the mean function, and $k(x, x')$ is the kernel function measuring similarity between points x and x' .

Using the Expected Improvement (EI) criterion, Bayesian Optimization iteratively refines weights by balancing exploration and exploitation:

$$\text{EI}(x) = \mathbb{E}[\max(f(x) - f(x^*), 0)], \quad (3)$$

where $f(x^*)$ is the best observed value of the objective function. This method efficiently identifies the optimal weight configuration to maximize ensemble accuracy.

IV. METHODOLOGY

A. Proposed Architecture

This study presents a novel image-based framework for Android malware detection, leveraging convolutional neural networks (CNNs) and a weighted voting ensemble to enhance detection accuracy.

The methodology is organized into three primary stages: data preprocessing, model training, and ensemble prediction.

In the preprocessing stage, the DEX file is extracted from Android APKs and transformed into grayscale images that encapsulate structural patterns indicative of malware. During model training, advanced CNN architectures are employed to analyze these images, enabling deep feature extraction and precise classification. Lastly, an optimized weighted voting ensemble integrates predictions from multiple models to improve overall performance and reliability.

The overall workflow of the proposed framework is illustrated in Fig. 3.

B. Data Preprocessing

In this study, the data preprocessing stage transforms Android APK files into grayscale images, which serve as input for the proposed deep learning framework. This transformation captures the structural patterns embedded in the DEX files, enabling robust malware detection. The preprocessing pipeline consists of the following steps:

1) *DEX File Extraction*: The first step involves extracting the `classes.dex` file from Android APKs. This file was chosen as the primary feature for analysis due to its structural richness and resilience to obfuscation techniques. Unlike other APK components, such as resource files or manifest configurations, the bytecode in the DEX file retains identifiable patterns that are critical for distinguishing between benign and malicious applications.

The extraction process treats the APK as a compressed archive, which is unzipped using Python's `zipfile` module. The `classes.dex` file is typically located in the root directory of the APK. To handle improperly named files, the script automatically renames files lacking the correct `.apk` extension, ensuring compatibility with the extraction process. This automated pipeline consistently prepares the DEX file for transformation into grayscale images.

2) *Binary-to-Image Conversion*: Once the DEX file is extracted, its binary content is read and converted into an 8-bit grayscale pixel matrix. Each byte in the binary sequence is mapped to a pixel intensity (0–255). This mapping ensures that the structural patterns in the bytecode are preserved in the resulting image.

The file size determines the width of the image, using a method adapted from Nataraj et al. [9]. The height is calculated dynamically to fit all pixels into a 2D matrix, using the formula:

$$\text{Height} = \frac{\text{Total Number of Bytes}}{\text{Image Width}}.$$

Table I outlines the mapping of file size ranges to image widths. This scaling ensures consistency while maintaining the integrity of structural characteristics.

TABLE I. IMAGE WIDTH CALCULATION BASED ON FILE SIZE

File Size Range	Image Width
<10 kB	32
10 kB – 30 kB	64
30 kB – 60 kB	128
60 kB – 100 kB	256
100 kB – 200 kB	384
200 kB – 500 kB	512
500 kB – 1000 kB	768
>1000 kB	1024

By retaining the sequence of bytecode as pixel intensities, this method preserves the unique structural characteristics of the application, such as opcode patterns and control flow representations. These features are critical to distinguish malware from benign applications.

C. Model Training

This study initially experimented with various pre-trained CNN architectures to identify the models most effective for classifying benign and malicious Android applications. Following extensive evaluations, six CNN models—ResNet50, AlexNet, DenseNet121, MobileNetV2, EfficientNetB0, and ShuffleNetV2—were selected for their complementary strengths in feature extraction and classification. These models demonstrated high performance in terms of accuracy, precision, recall, and F1-scores during preliminary testing. Each model was initialized with pre-trained weights from ImageNet and trained using a consistent pipeline to ensure fairness and comparability.

1) *Image Preparation*: The grayscale images generated from the `classes.dex` files were resized to a fixed input dimension of 224×224 pixels. They were normalized using the ImageNet dataset's mean $([0.485, 0.456, 0.406])$ and standard deviation $([0.229, 0.224, 0.225])$. Data augmentation techniques, including resizing and normalization, were applied to enhance model generalization.

2) *Classifier Adaptation*: Each model's classifier was customized for binary classification by replacing the original fully connected layers with the following configuration:

- A dropout layer ($p = 0.4$) to mitigate overfitting.
- A dense layer with 256 units and ReLU activation.
- A second dropout layer ($p = 0.4$).
- A final dense layer with one output node and a sigmoid activation function.

3) *Training Process*: The models were trained using the AdamW optimizer with a learning rate of 0.0001 and a weight decay of 1×10^{-5} . The binary cross-entropy loss function was employed for optimization. Training was conducted over 20 epochs with a batch size of 32. An early stopping mechanism,

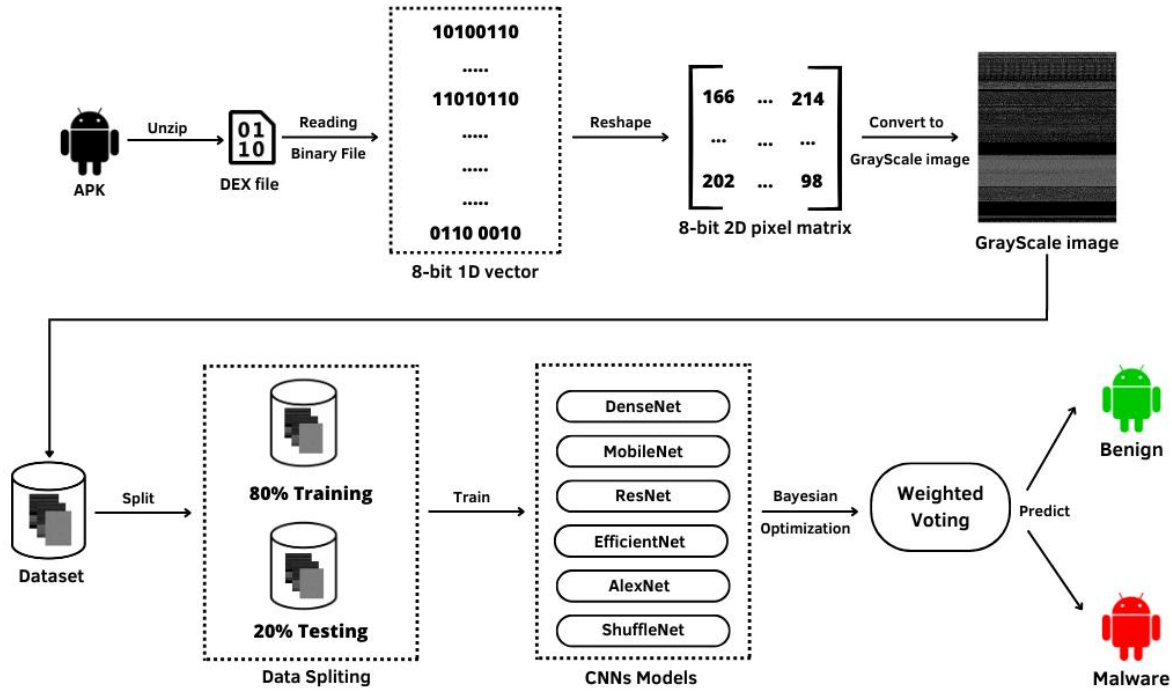


Fig. 3. The proposed architecture for android malware detection.

with a patience of 5 epochs, was implemented to prevent overfitting while ensuring optimal model performance.

D. Ensemble Prediction

To enhance the overall performance and robustness of the classification, an ensemble prediction strategy was employed. The ensemble combines the outputs of six pre-trained convolutional neural network (CNN) models: ResNet50, AlexNet, DenseNet121, MobileNetV2, EfficientNetB0, and ShuffleNetV2. By leveraging the strengths of each model, the ensemble aims to improve classification accuracy and reduce errors in detecting Android malware.

The ensemble prediction uses a weighted voting mechanism, where the outputs from each model are aggregated based on their performance during validation. For each input image, the probability of classification is calculated as a weighted sum of the individual model predictions. Weights were constrained to $[0, 1]$, normalized to sum to 1, and refined over 300 iterations using the Expected Improvement (EI) criterion.

Bayesian Optimization was employed to determine the optimal weights for the ensemble, as detailed in Algorithm 1. This probabilistic technique was chosen for its ability to efficiently explore high-dimensional parameter spaces while balancing exploration with exploitation. Unlike manual selection or traditional methods such as grid or random search, Bayesian Optimization leverages information from prior iterations to refine the weight configurations, resulting in faster convergence and more effective optimization.

Algorithm 1 Bayesian Optimization for Ensemble Weights

Require: Validation dataset \mathcal{D} , pre-trained models: $\{\text{ResNet50, AlexNet, DenseNet121, MobileNetV2, EfficientNetB0, ShuffleNetV2}\}$, number of iterations $N = 300$, batch size = 32.

Ensure: Optimal weights $\{w_1, w_2, \dots, w_6\}$ for the ensemble.

- 1: Initialize bounds $[0, 1]$ for each w_i and randomly generate initial weights.
- 2: **for** $i = 1$ to N **do**
- 3: Generate candidate weights $\{w_1, w_2, \dots, w_6\}$.
- 4: Normalize weights such that $\sum_{i=1}^6 w_i = 1$.
- 5: Compute ensemble prediction for each image x :

$$P_{\text{ensemble}}(x) = \sum_{i=1}^6 w_i \cdot P_i(x)$$

- 6: Evaluate ensemble accuracy on \mathcal{D} .
- 7: Refine weights using the Expected Improvement (EI) criterion:
 - a. Explore unvisited regions of the weight space.
 - b. Exploit regions with promising results.
- 10: Update surrogate model with new results.
- 11: **end for**
- 12: **return** Optimal weights $\{w_1, w_2, \dots, w_6\}$ achieving the highest ensemble accuracy.

V. EVALUATION AND RESULT

A. Dataset

In this study, the CICMalDroid 2020 dataset [38] was utilized, a benchmark dataset designed to support research and development in Android malware detection. The dataset was created by the Canadian Institute for Cybersecurity (CIC) and contains a total of 17,341 APK files. It categorizes APKs into five classes: **Adware**, **Banking**, **SMS**, **Riskware**, and **Benign**. Each class represents a specific type of Android application behavior, with malware types grouped based on their malicious intent, and benign applications serving as the control group.

1) *Dataset Preparation*: During the preprocessing stage, 415 APK files were excluded due to various issues, including:

- Missing or inaccessible `classes.dex` files.
- Permission restrictions preventing file access.
- Corrupted or non-standard file formats.
- Invalid or unusual filenames.
- Integrity check failures, such as bad CRC-32.

These issues prevented consistent processing of a subset of the dataset. As a result, only valid and complete files were included to maintain data quality and ensure reliable model training.

2) *Data Splitting*: Initially, the dataset was split into training (80%) and validation (20%) subsets for each class, ensuring balanced representation in both subsets. Following this, all malware classes were concatenated into a single class named **Malware** for binary classification, while the **Benign** class remained unchanged. This process resulted in the following data distribution:

- Benign: 3,216 samples for training and 805 for validation.
- Malware: 10,001 samples for training and 2,502 for validation.

B. Experimental Setup and Performance Metrics

The experiments were conducted on a Windows operating system using Python as the programming language. The models were implemented with the PyTorch 2.4.1 deep learning framework, and the training process was accelerated using CUDA 11.8 on an NVIDIA RTX 2060 GPU.

To evaluate the performance of the proposed models, a variety of metrics were employed to ensure a comprehensive understanding of their classification capabilities. Each metric serves a distinct purpose, offering insights into specific aspects of the models' performance and their ability to distinguish between benign and malware samples. The definitions of the metrics used are as follows:

- Accuracy (AC): Represents the overall proportion of correctly classified samples:

$$AC = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

where TP is true positives, TN is true negatives, FP is false positives, and FN is false negatives.

- Precision (P): Measures the proportion of true positive predictions out of all positive predictions:

$$P = \frac{TP}{TP + FP} \quad (5)$$

This metric highlights the model's ability to minimize false positives.

- Recall (R): Evaluates the proportion of true positive predictions among all actual positive samples:

$$R = \frac{TP}{TP + FN} \quad (6)$$

Recall is critical in malware detection to ensure that malicious applications are not overlooked.

- F1-Score (F): The harmonic mean of precision and recall, providing a balance between the two:

$$F = \frac{2 \cdot P \cdot R}{P + R} \quad (7)$$

The defined metrics were utilized to evaluate the classification performance of six individual CNN models and the proposed ensemble learning approach. Table II provides a comparative summary of the results, presenting the accuracy, precision, recall, and F1-score for each model. This analysis highlights the unique strengths and weaknesses of each model while demonstrating the superior performance of the ensemble approach.

TABLE II. PERFORMANCE METRICS FOR ALL MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)
ResNet50	98.54	98.76	99.32	99.04
AlexNet	98.51	99.11	98.92	99.01
DenseNet121	98.42	99.59	98.32	98.95
MobileNetV2	98.45	99.35	98.60	98.97
EfficientNetB0	98.52	99.04	99.00	99.02
ShuffleNetV2	98.39	98.64	99.24	98.94
Ensemble Learning	99.30	99.59	99.48	99.54

The superior performance of the ensemble was achieved through the optimal weights determined by Bayesian Optimization. These weights were carefully assigned to balance the contributions of each model, reflecting their relative importance in the ensemble's predictions. Table III presents the final weights for each model, highlighting the significant contributions of MobileNetV2 and EfficientNetB0, which had the highest weights, underscoring their strong performance in feature extraction and classification.

For further insights into the models' classification performance, the Receiver Operating Characteristic (ROC) curves, presented in Fig. 4, offer a detailed evaluation of the individual models and the ensemble learning approach. All models exhibit excellent discriminatory capabilities, with Area Under the Curve (AUC) values exceeding 0.997. The ensemble learning approach outperformed all individual models, achieving the

TABLE III. OPTIMAL WEIGHTS FOR ENSEMBLE MODELS

Model	Optimal Weight
AlexNet	0.037
DenseNet121	0.172
EfficientNetB0	0.225
MobileNetV2	0.288
ResNet50	0.124
ShuffleNetV2	0.154

highest AUC of 0.9993. This superior performance highlights the ensemble’s ability to effectively combine the strengths of individual models, resulting in enhanced classification accuracy.

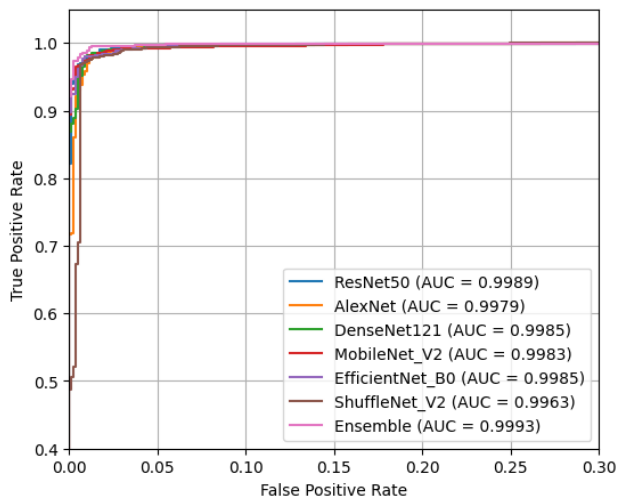


Fig. 4. ROC curves for CNN models and ensemble learning.

To provide a more detailed evaluation of the ensemble’s predictions, the confusion matrix in Fig. 5 illustrates its classification outcomes for benign and malware samples. The model successfully classified 795 benign and 2,489 malware samples, with only 10 false positives and 13 false negatives. These results demonstrate the ensemble’s high precision and recall, ensuring robust malware detection with minimal errors.

The proposed ensemble learning approach demonstrated significant improvements in Android malware detection accuracy compared to prior methods. Table IV provides a comparative analysis of existing approaches, highlighting datasets, models, and achieved accuracies. The proposed method outperformed techniques such as CNN-LSTM, CNN-SVM [20], and single-model approaches such as ResNet [19] and EfficientNet [16], it also surpassed studies employing multi-feature fusion strategies [22].

C. Ablation Study: Incremental Model Analysis

An ablation study was conducted to determine the optimal configuration of the ensemble by incrementally adding models and evaluating their impact on performance metrics. Starting with a baseline ensemble of ResNet50, AlexNet, and EfficientNetB0, additional models—MobileNetV2, DenseNet121, and

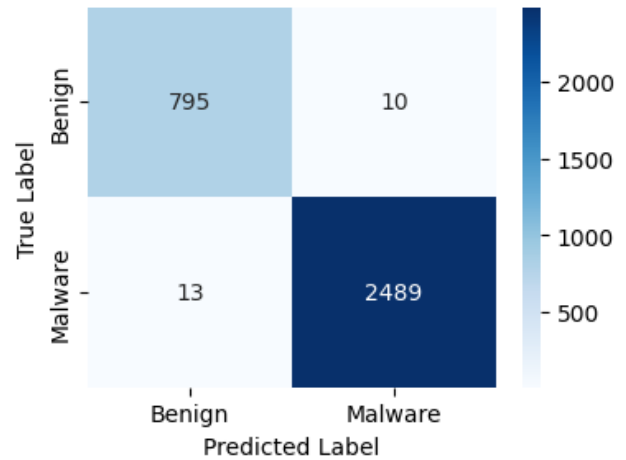


Fig. 5. Confusion matrix for the ensemble learning approach.

TABLE IV. COMPARATIVE ANALYSIS OF ANDROID MALWARE DETECTION METHODS

Ref	Dataset	Model	Acc (%)
[11]	AMD (6,134 malware), Google Play (4,406 benign)	CNN	93
[12]	Drebin	CNN	95.1
[13]	400,000 apps (200,000 malware, 200,000 benign)	CNN	93.36
[14]	CICAndMal2017, CICInvesAndMal2019, CICMalDroid 2020	TCN	95.44
[15]	Custom Dataset	DeepVisDroid (1D-CNN)	98
[16]	Maling, Drebin	EfficientNetB4	93.65
[17]	CIC-AAGM2017, CICMalDroid 2020	Hybrid (Textual+Image)	99
[18]	Real-world datasets	ViT	80.27
[19]	Drebin, CICMalDroid 2020	ResNet	98.67
[20]	Drebin, AndroZoo	CNN-LSTM, CNN-SVM	99 detection, 97 classification
[21]	CICMalDroid 2020	CNN	91
[22]	CICMalDroid 2020	AlexNet, GoogleNet, ResNet	97.25
[23]	Custom Dataset (Dalvik Bytecode)	Grayscale+RGB Deep Learning	98.7
Our Method	CICMalDroid 2020	Ensemble CNN Models	99.3

ShuffleNetV2—were iteratively incorporated based on their individual metrics, such as precision, recall, and F1-score.

Each addition was rigorously evaluated for its contribution to accuracy and robustness, with weights re-optimized using Bayesian Optimization to ensure balanced contributions. The

study demonstrated a consistent progression in performance: accuracy improved from 99.09% with three models to 99.30% with six models, achieving the best performance across all metrics in the final configuration.

The results highlight the incremental benefits of integrating diverse architectures into the ensemble, validating the soundness of the methodology. Fig. 6 illustrates the steady improvements, emphasizing the rationale and effectiveness of this approach.

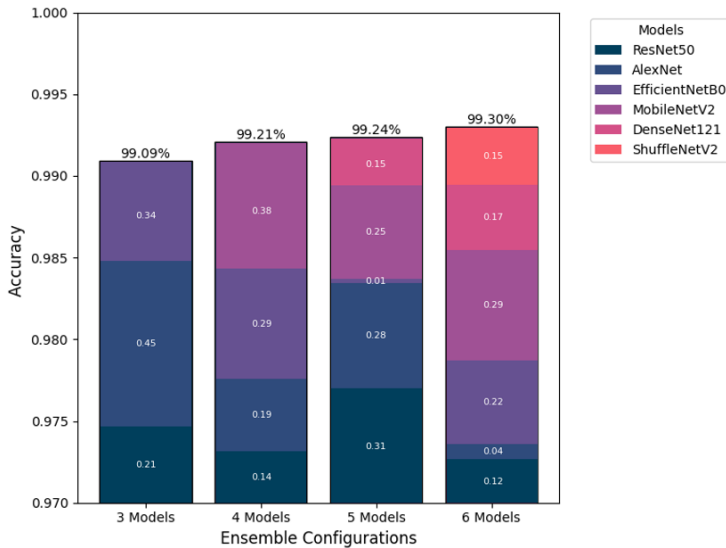


Fig. 6. Comparison of accuracy across ensemble learning configurations.

VI. DISCUSSION AND CHALLENGES

The proposed ensemble approach achieved remarkable results, surpassing prior methodologies in Android malware detection with an accuracy of 99.3%, addressing challenges posed by evolving malware obfuscation techniques, particularly through the use of grayscale images derived from DEX files, which preserve critical bytecode patterns. The methodology not only outperformed single-model approaches, such as CNNs in [12] (95.1%) and TCN [14] (95.44%), but also exceeded advanced methods like CNN-LSTM [20] and multi-feature strategies [22], which achieved accuracies of 97% and 97.25%, respectively. This demonstrates the ensemble's capacity to generalize across diverse malware families while maintaining classification robustness.

However, several challenges were encountered during the study. One notable issue was overfitting, mitigated effectively by employing early stopping to ensure generalization without compromising performance. Another challenge was the lack of recent Android malware datasets. The rapid evolution of malware introduces the risk that outdated datasets may fail to capture current threats, limiting the generalizability of detection systems.

Furthermore, deploying a six-model ensemble in real-time environments poses practical difficulties due to the increased computational resources and latency required to run multiple models simultaneously. Addressing this limitation will be

crucial for translating the proposed approach into real-world applications.

Despite these challenges, the study demonstrates the potential of ensemble learning in advancing the state of Android malware detection. By combining multiple CNN architectures and optimizing their contributions, this methodology provides a robust framework that paves the way for future research on integrating multi-feature analysis and scaling to larger datasets.

VII. CONCLUSION

This study introduced a novel ensemble learning framework for Android malware detection, utilizing grayscale images derived from DEX files and six pre-trained CNN models. Achieving an impressive accuracy of 99.3%, the proposed method surpassed existing approaches in the field. By leveraging a weighted voting mechanism, optimized through Bayesian Optimization, the ensemble demonstrated superior performance across key metrics, achieving precision, recall, and F1-scores of 99.59%, 99.48%, and 99.54%, respectively. This robust approach effectively minimized classification errors while ensuring reliable malware detection.

The findings underscore the benefits of combining multiple CNN architectures to harness their complementary strengths in feature extraction and classification. Additionally, the framework demonstrated resilience against obfuscation techniques frequently used by malware developers, enhancing its practicality for real-world applications.

Future research will focus on expanding this methodology to encompass larger, more diverse datasets and integrating multi-feature approaches with advanced analysis techniques. These efforts aim to further improve detection accuracy and adaptability, addressing the ever-evolving landscape of Android malware threats.

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to the Engineering Sciences Laboratory, National School of Applied Sciences, Ibn Tofail University, Kenitra, Morocco for providing the resources and support needed to carry out this research. The authors also extend their thanks to the CNRST (Centre National pour la Recherche Scientifique et Technique) for its financial support under the "PhD-Associate Scholarship - PASS" Program.

REFERENCES

- [1] "Android Statistics (2024)," Business of Apps. <https://www.businessofapps.com/data/android-statistics/>
- [2] "Mobile & Tablet Android Version Market Share Worldwide," StatCounter Global Stats. <https://gs.statcounter.com/android-version-market-share/mobile-tablet/worldwide>
- [3] "Google Play Store: number of apps 2024," Statista. <https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/>
- [4] D. Ucci, L. Aniello, and R. Baldoni, "Survey of machine learning techniques for malware analysis," *Comput. Secur.*, vol. 81, pp. 123–147, Mar. 2019, doi: 10.1016/j.cose.2018.11.001.
- [5] W. F. Elersy, A. Feizollah, and N. B. Anuar, "The rise of obfuscated Android malware and impacts on detection methods," *PeerJ Comput. Sci.*, vol. 8, p. e907, Mar. 2022, doi: 10.7717/peerj-cs.907.

- [6] Z. Wang, Q. Liu, and Y. Chi, "Review of Android Malware Detection Based on Deep Learning," *IEEE Access*, vol. 8, pp. 181102–181126, 2020, doi: 10.1109/ACCESS.2020.3028370.
- [7] F. Hamid, "Enhancing Malware Detection with Static Analysis using Machine Learning," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 7, no. 6, pp. 38–42, Jun. 2019, doi: 10.22214/ijraset.2019.6010.
- [8] T. Bhatia and R. Kaushal, "Malware detection in android based on dynamic analysis," in *2017 International Conference on Cyber Security And Protection Of Digital Services (Cyber Security)*, Jun. 2017, pp. 1–6. doi: 10.1109/CyberSecPODS.2017.8074847.
- [9] L. Nataraj, S. Karthikeyan, G. Jacob, and B. Manjunath, "Malware Images: Visualization and Automatic Classification," *Jul. 2011*, doi: 10.1145/2016904.2016908.
- [10] V. Sihag, S. Prakash, G. Choudhary, N. Dragoni, and I. You, "DIMDA: Deep Learning and Image-Based Malware Detection for Android," in *Futuristic Trends in Networks and Computing Technologies*, P. K. Singh, S. T. Wierzczoń, J. K. Chhabra, and S. Tanwar, Eds., Singapore: Springer Nature, 2022, pp. 895–906. doi: 10.1007/978-981-19-5037-7_64.
- [11] Shao Yang, "An Image-Inspired and CNN-Based Android Malware Detection Approach," in *2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE)*, San Diego, CA, USA: IEEE, Nov. 2019, pp. 1259–1261. doi: 10.1109/ASE.2019.00155.
- [12] Y. Ding, X. Zhang, J. Hu, and W. Xu, "Android malware detection method based on bytecode image," *J Ambient Intell Human Comput*, vol. 14, no. 5, pp. 6401–6410, May 2023, doi: 10.1007/s12652-020-02196-4.
- [13] A. Rahali, A. H. Lashkari, G. Kaur, L. Taheri, F. Gagnon, and F. Massicotte, "DiDroid: Android Malware Classification and Characterization Using Deep Image Learning," in *2020 the 10th International Conference on Communication and Network Security*, Tokyo Japan: ACM, Nov. 2020, pp. 70–82. doi: 10.1145/3442520.3442522.
- [14] W. Zhang, N. Luktarhan, C. Ding, and B. Lu, "Android Malware Detection Using TCN with Bytecode Image," *Symmetry*, vol. 13, no. 7, Art. no. 7, Jul. 2021, doi: 10.3390/sym13071107.
- [15] K. Bakour and H. M. Ünver, "DeepVisDroid: android malware detection by hybridizing image-based features with deep learning techniques," *Neural Comput & Applic*, vol. 33, no. 18, pp. 11499–11516, Sep. 2021, doi: 10.1007/s00521-021-05816-y.
- [16] R. Mitsuhashi and T. Shinagawa, "Exploring Optimal Deep Learning Models for Image-based Malware Variant Classification," in *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)*, Jun. 2022, pp. 779–788. doi: 10.1109/COMP-SAC54236.2022.00128.
- [17] F. Ullah, S. Ullah, M. R. Naeem, L. Mostarda, S. Rho, and X. Cheng, "Cyber-Threat Detection System Using a Hybrid Approach of Transfer Learning and Multi-Model Image Representation," *Sensors*, vol. 22, no. 15, p. 5883, Aug. 2022, doi: 10.3390/s22155883.
- [18] J. Jo, J. Cho, and J. Moon, "A Malware Detection and Extraction Method for the Related Information Using the ViT Attention Mechanism on Android Operating System," *Applied Sciences*, vol. 13, no. 11, p. 6839, Jun. 2023, doi: 10.3390/app13116839.
- [19] A. Aldini and T. Petrelli, "Image-based detection and classification of Android malware through CNN models," in *Proceedings of the 19th International Conference on Availability, Reliability and Security*, Vienna Austria: ACM, Jul. 2024, pp. 1–11. doi: 10.1145/3664476.3670441.
- [20] Ö. Kiraz and İ. A. Dođru, "Visualising Static Features and Classifying Android Malware Using a Convolutional Neural Network Approach," *Applied Sciences*, vol. 14, no. 11, p. 4772, May 2024, doi: 10.3390/app14114772.
- [21] J. Tang *et al.*, "Android malware detection based on a novel mixed bytecode image combined with attention mechanism," *Journal of Information Security and Applications*, vol. 82, p. 103721, May 2024, doi: 10.1016/j.jisa.2024.103721.
- [22] Z. Wang, Q. Yu, and S. Yuan, "Android Malware Detection Based on RGB Images and Multi-feature Fusion," Aug. 29, 2024, arXiv: arXiv:2408.16555. doi: 10.48550/arXiv.2408.16555.
- [23] M. M. Yapici, "A Novel Image Based Approach for Mobile Android Malware Detection and Classification," 2024. doi: 10.2139/ssrn.4942956.
- [24] M. Mozaffari-Kermani, R. Azarderakhsh, C.-Y. Lee, and S. Bayat-Sarmadi, "Reliable Concurrent Error Detection Architectures for Extended Euclidean-Based Division Over GF(2^m)," *IEEE Trans. Very Large Scale Integr. VLSI Syst.*, vol. 22, no. 5, pp. 995–1003, May 2014, doi: 10.1109/TVLSI.2013.2260570.
- [25] B. Koziel, R. Azarderakhsh, and M. Mozaffari-Kermani, "Low-Resource and Fast Binary Edwards Curves Cryptography," in *Progress in Cryptology – INDOCRYPT 2015*, A. Biryukov and V. Goyal, Eds., Cham: Springer International Publishing, 2015, pp. 347–369. doi: 10.1007/978-3-319-26617-6_19.
- [26] M. Mozaffari-Kermani and A. Reyhani-Masoleh, "A low-cost S-box for the Advanced Encryption Standard using normal basis," in *2009 IEEE International Conference on Electro/Information Technology*, Jun. 2009, pp. 52–55. doi: 10.1109/EIT.2009.5189583.
- [27] A. Jalali, R. Azarderakhsh, M. M. Kermani, and D. Jao, "Towards Optimized and Constant-Time CSIDH on Embedded Devices," in *Constructive Side-Channel Analysis and Secure Design*, I. Polian and M. Stöttinger, Eds., Cham: Springer International Publishing, 2019, pp. 215–231. doi: 10.1007/978-3-030-16350-1_12.
- [28] K.-K. R. Choo, M. M. Kermani, R. Azarderakhsh, and M. Govindarasu, "Emerging Embedded and Cyber Physical System Security Challenges and Innovations," *IEEE Trans. Dependable Secure Comput.*, vol. 14, no. 3, pp. 235–236, May 2017, doi: 10.1109/TDSC.2017.2664183.
- [29] R. Meier, "Professional Android Application Development". Indianapolis, IN, USA: Wiley Publishing, Inc., 2009.
- [30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.
- [31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun ACM*, vol. 60, no. 6, pp. 84–90, mai 2017, doi: 10.1145/3065386.
- [32] G. Huang, Z. Liu, G. Pleiss, L. van der Maaten, and K. Q. Weinberger, "Convolutional Networks with Dense Connectivity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 12, pp. 8704–8716, Dec. 2022, doi: 10.1109/TPAMI.2019.2918284.
- [33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 4510–4520. doi: 10.1109/CVPR.2018.00474.
- [34] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Sep. 11, 2020, arXiv: arXiv:1905.11946. doi: 10.48550/arXiv.1905.11946.
- [35] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design," Jul. 30, 2018, arXiv: arXiv:1807.11164. doi: 10.48550/arXiv.1807.11164.
- [36] A. Dogan and D. Birant, "A Weighted Majority Voting Ensemble Approach for Classification," in *2019 4th International Conference on Computer Science and Engineering (UBMK)*, Sep. 2019, pp. 1–6. doi: 10.1109/UBMK.2019.8907028.
- [37] P. I. Frazier, "A Tutorial on Bayesian Optimization," Jul. 08, 2018, arXiv: arXiv:1807.02811. doi: 10.48550/arXiv.1807.02811.
- [38] "MalDroid 2020 — Datasets — Research — Canadian Institute for Cybersecurity — UNB". <https://www.unb.ca/cic/datasets/maldroid-2020.html>

Cross-Domain Health Misinformation Detection on Indonesian Social Media

Divi Galih Prasetyo Putri¹, Savitri Citra Budi², Arida Ferti Syafiandini³,
Ikhlasul Amal⁴, Revandra Aryo Dwi Krisnandaru⁵

Department of Electronic and Informatics Engineering, Universitas Gadjah Mada, Indonesia^{1,5}

Department of Health Information Management, Universitas Gadjah Mada, Indonesia²

Research Center for Computing, National Research and Innovation Agency (BRIN), Indonesia³

Department of Computer Science and Electronics, Universitas Gadjah Mada, Indonesia⁴

Abstract—Indonesia is among the world’s most prolific countries in terms of internet and social media usage. Social media serves as a primary platform for disseminating and accessing all types of information, including health-related data. However, much of the content generated on these platforms is unverified and often falls into the category of misinformation, which poses risks to public health. It is essential to ensure the credibility of the information available to social media users, thereby helping them make informed decisions and reducing the risks associated with health misinformation. Previous research on health misinformation detection has predominantly focused on English-language data or has been limited to specific health crises, such as COVID-19. Consequently, there is a need for a more comprehensive approach which not only focus on single issue or domain. This study proposes the development of a new corpus that encompasses various health topics from Indonesian social media. Each piece of content within this corpus will be manually annotated by expert to label a social media post as either misinformation or fact. Additionally, this research involves experimenting with machine learning models, including traditional and deep learning models. Our finding shows that the new cross-domain dataset is able to achieve better performance compared to those trained on the COVID dataset, highlighting the importance of diverse and representative training data for building robust health misinformation detection system.

Keywords—Health misinformation; machine learning; social media

I. INTRODUCTION

The proliferation of internet users in Indonesia has been steadily increasing. A 2023 survey by the Indonesian Internet Service Provider Association reported that internet penetration had reached 78%, marking a significant rise compared to the previous year¹. The COVID-19 pandemic in early 2020 played a pivotal role in accelerating the adoption of internet-based applications for daily activities. With the shift to remote work, online education, e-commerce, and telemedicine, the internet became indispensable for modern life. Among these changes, access to health information via digital platforms has expanded significantly. Social media platforms, in particular, have enabled the rapid dissemination of health-related content through user-generated content (UGC), encompassing text, images, videos, and comments on a wide range of topics. Features such as likes, shares, and comments amplify the spread of information, which spans various subtopics, including general

health, vaccines, diseases like Ebola and cancer, and public health crises such as the COVID-19 pandemic [1], [2].

However, the exponential growth of information on social media presents a dual challenge: while access to health-related content is democratized, it becomes increasingly difficult for users to assess the credibility and quality of this information. Studies have identified social media platforms like Facebook, Twitter, and Instagram as primary vectors for the spread of health misinformation, which often propagates faster than accurate information [3]. Health misinformation has been linked to severe consequences, particularly during the COVID-19 pandemic, where widespread falsehoods created confusion, reduced vaccine uptake, and undermined herd immunity efforts. The World Health Organization (WHO) reported that misinformation during the pandemic led to over 6,000 hospitalizations and 800 deaths globally [4]. Such incidents underscore the real-world consequences of misinformation, which can escalate from individual confusion to public health crises.

Misinformation surrounding vaccines exemplifies the long-standing impact of health-related falsehoods. For instance, between September 2018 and July 2019, 85% of the 649 reported measles cases in the U.S. involved unvaccinated individuals [5]. Furthermore, recent surveys indicate that misinformation about chronic diseases, such as diabetes and cancer, is among the most concerning categories of health-related content on social media. The persistent spread of misinformation highlights the critical need to address this issue comprehensively. Left unchecked, such falsehoods can lead to misinformed decisions, delays in seeking proper medical care, and ultimately, adverse health outcomes for individuals and communities.

The urgency of tackling health misinformation has prompted researchers to explore automated solutions for misinformation detection. Automated systems can enable users to instantly assess the credibility of content accessed on social media, empowering them to make informed health decisions and reducing the risks associated with misinformation. While significant progress has been made in health misinformation detection research, much of the existing work has concentrated on English-language data [6]. This focus leaves a critical gap in addressing misinformation in non-English contexts, such as Bahasa Indonesia, a language spoken by over 270 million people.

In Indonesia, 76% of users perceive social media as a trustworthy source of information [7]. Moreover, according to

¹<https://m.bisnis.com/amp/read/20230308/101/1635219/survei-apjii-pengguna-internet-di-indonesia-tembus-215-juta-orang>.

data from the Indonesian Telecommunications Society, approximately 40% of the hoax news articles circulating in Indonesia in 2019 were related to health [8], making the detection of misinformation even more pressing. Recent efforts have begun creating datasets and machine learning models for detecting misinformation in Bahasa Indonesia. However, these efforts have primarily focused on domain-specific topics, such as COVID-19, limiting their applicability to other health-related misinformation. Moreover, existing studies often neglect the diverse and evolving nature of misinformation across different health domains, which can include varied topics like traditional medicine, vaccines, mental health, and non-communicable diseases.

To address the gaps identified in previous studies, this research seeks to answer the following key research questions:

- 1) How can a cross-domain dataset for health misinformation detection in Bahasa Indonesia be effectively constructed to address diverse health topics?
- 2) Can machine learning models trained on the proposed cross-domain dataset generalize effectively across various health domains, reducing dependence on any single domain?
- 3) How does the performance of the proposed cross-domain dataset compare to existing domain-specific datasets in terms of robustness and quality?

Based on the proposed research questions, this study makes the following contributions to the field of health misinformation detection:

1) *Dataset development:* Creation of a cross-domain dataset for health misinformation detection in Bahasa Indonesia, encompassing diverse health topics to enable broader generalization.

2) *Model evaluation:* Preliminary experiments using multiple machine learning approaches to evaluate the effectiveness of the constructed dataset.

3) *Benchmarking:* Comprehensive comparison of the proposed dataset against existing domain-specific datasets to demonstrate its robustness and quality.

This paper is structured as follows: Section II reviews related work on health misinformation detection, highlighting gaps in existing research. Section III details the methodology for constructing the dataset and designing preliminary experiments. Section IV presents and discusses the experimental results, while Section V concludes the study and outlines directions for future work.

II. RELATED WORKS

Health misinformation detection has been extensively studied, primarily focusing on English. These studies leverage various approaches, such as linguistic and behavioral features, to identify and combat misinformation. For instance, Zhao et al. (2021) proposed a model combining central and peripheral features based on the Elaboration Likelihood Model (ELM), significantly improving detection accuracy by integrating user interaction patterns [6]. Similarly, Zhong et al. (2023) analyzed temporal and sentiment patterns in misinformation dissemination on Twitter [9], revealing that misinformation tends

to persist longer and garner more engagement than credible information.

Despite these advancements, research on health misinformation detection in non-English contexts, particularly Bahasa Indonesia, remains limited. Faisal and Mahendra (2022) addressed this gap by developing a COVID-19-specific misinformation dataset and proposing a two-stage classifier leveraging IndoBERT for Indonesian tweets. Their approach demonstrated the efficacy of pre-trained language models but was constrained by its focus on the COVID-19 domain [7]. In fact, several studies on health misinformation focused on the topic of covid-19 [10], [11]. Another effort by Prasetyo et al. (2018) explored classification techniques for health-related hoax news in Bahasa Indonesia using the Modified K-Nearest Neighbor (MKNN) method. Their results showed an accuracy of 75%, with performance influenced by challenges such as unstructured text and diverse linguistic styles in Indonesian health news [12].

In the broader context of misinformation detection in Indonesia, Rohman et al. (2021) conducted a systematic literature review analyzing methods for fake news classification in Bahasa Indonesia. Their findings identified 19 commonly used algorithms, with Naïve Bayes [13] and Term Frequency-Inverse Document Frequency (TF-IDF) [14] being the most exploited approach. They highlighted the dominance of datasets sourced from platforms like turnbackhoax.id and Twitter and emphasized the need for exploring a wider range of methods beyond these mainstream approaches [15]. Additionally, a recent study by Arini et al. (2024) examined the sociocultural factors, such as varying levels of trust in information sources, that influence the spread and detection of misinformation during the pandemic [16].

The concept of cross-domain classification has emerged as a promising avenue for addressing the limitations of domain-specific models. Kansal et al. (2020) reviewed cross-domain sentiment analysis methods, emphasizing their potential to reduce reliance on single-domain datasets through domain adaptation and transfer learning [17]. These findings suggest that cross-domain approaches could enhance the generalizability of misinformation detection models across diverse health topics. Based on these insights, this research aims to address the gaps in existing studies by proposing a cross-domain dataset for health misinformation detection in Bahasa Indonesia. By leveraging state-of-the-art machine learning techniques and integrating cross-domain classification methods, this study seeks to overcome the challenges posed by linguistic and contextual diversity in Indonesian health misinformation.

III. RESEARCH METHODOLOGY

In this research, we develop a dataset for cross-domain health misinformation detection in Indonesian tweet. Moreover, we perform a preliminary study by exploring commonly used machine learning approach.

A. Data Collection and Annotation

The primary goal of this phase is to construct a cross-domain corpus capturing health misinformation phenomena in Indonesian social media. The dataset construction process is illustrated in Fig. 1.

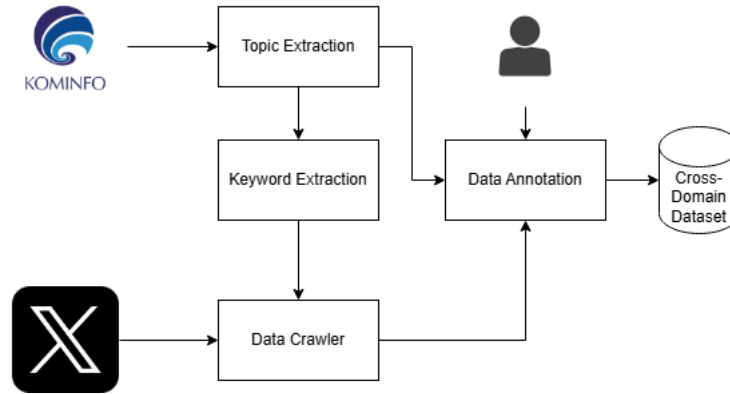


Fig. 1. Corpus building process.

1) *Data collection*: Data was collected from various health-related topics by utilizing weekly hoax news content published on the official website of the Ministry of Communication and Information Technology (KOMINFO) of Indonesia (Kominfo, 2023). A total of 27 hoax news articles covering topics such as COVID-19, HIV, polio, and other health concerns were analyzed. Table I summarizes the articles and the number of tweets crawled for each.

2) *Keyword extraction*: For each article, primary keywords were extracted and combined into search queries for crawling tweets. For example, for the article titled “[HOAKS] Buah Pala dan Gula Batu Mampu Mengatasi Jantung Berdebar” [18], keywords such as Pala, Gula Batu, and Jantung Berdebar were extracted and combined into search phrases like Pala AND Jantung Berdebar and Gula Batu AND Jantung Berdebar. These phrases were used in the crawling process, focusing on data from Twitter (now known as X).

3) *Manual annotation*: After collection, the tweets were manually annotated by two annotators based on predefined criteria. The annotation process involved verifying that the content was written in Indonesian, relevant to health information, and determining whether the content constituted misinformation or truth. The labels assigned were:

a) *Misinformation*: Content identified as false based on the KOMINFO news article.

b) *True*: Content verified as accurate.

To ensure the reliability of the annotations, inter-annotator agreement was measured using Cohen’s Kappa. In cases of disagreement, a third annotator resolved conflicts. This process ensures high-quality annotations for downstream tasks.

B. Model Development

To assess the quality and applicability of the newly constructed dataset, a preliminary study was conducted on the task of health misinformation detection. This involved training and evaluating various machine learning models.

1) *Traditional models*: The following models were implemented using scikit-learn with Bag-of-Words as the feature representation:

- Naive Bayes

TABLE I. SUMMARY OF HOAX NEWS ARTICLES AND CRAWLING RESULTS

News Title	Topic	Numb. of Tweet
[HOAKS] Covid-19 bukan Virus, Sumber: Kementerian Kesehatan RI	covid	29
[HOAKS] Covid-19 adalah Senjata Biologis dari Cina	covid	716
[HOAKS] Vaksin Covid-19 Adalah Antena 5G dan Penyebab Kanker	covid	220
[HOAKS] Vaksin Covid-19 Sinovac Sebabkan Mpx	covid	29
[HOAKS] Obat Corona Bernama Pil-Kada	covid	64
[HOAKS] Tinta Tak Kasat Mata Dimasukkan ke Vaksin	covid	12
Vaksin Berbasis mRNA Picu Kanker? Itu Disinformasi	covid	66
[HOAKS] Mengkudu Menyembuhkan Darah Tinggi Secara Total	high blood pressure	72
[HOAKS] Akar Kelapa dan Kuning Telur Dapat Meningkatkan Fertilitas	Fertility	39
[HOAKS] Video Tata Cara Pertolongan Pertama Penanganan Flu Burung	avian flu	23
[HOAKS] Terapi Minuman Rempah Bisa Menggantikan Cuci Darah pada Gagal Ginjal	kidney failure	7
[HOAKS] GERD Picu Kematian Mendadak	GERD	110
[HOAKS] Penularan HIV dan AIDS di Kolam Renang	HIV	294
[HOAKS] HIV Dapat Ditularkan oleh Gigitan Nyamuk	HIV	564
[HOAKS] Buah Pala dan Gula Batu Mampu Mengatasi Jantung Berdebar	Heart	10
[HOAKS] Parfum, Pengharum Ruangan, dan Wangi Dry Clean Sebagai Penyebab Kanker	Cancer	347
[HOAKS] Pisang Dempet Sebabkan Anak Lahir dengan Kembar Siam	cojoined twins	312
[HOAKS] Cara Mengecek Kadar Kolesterol melalui Warna Kuku	Cholesterol	34
[HOAKS] Tanaman Pacing Dapat Sembuhkan Mata Minus	near-sightedness	3
[HOAKS] Peringatan IDI Terkait Adanya Wabah Pengerasan Otak, Sumsum Tulang, dan Diabetes	brain, bone marrow, diabetes	381
[HOAKS] Dokumen Rahasia BPOM Sebut Vaksin Polio Tidak Aman	Polio	46
[HOAKS] Indonesia Sudah Lama Tidak Ada Wabah Polio	Polio	53
[HOAKS] KLB Polio Disebabkan Vaksin Polio Tipe-2	Polio	29
[HOAKS] Getah Bunga Mahkota Duri Sembuhkan Sakit Gigi Secara Instan	toothache	3
[HOAKS] Mengobati Sesak Nafas dengan Pijat Tangan	dyspnea	90
[HOAKS] Video Tata Cara Pertolongan Pertama Penanganan Stroke	stroke	70
Total		3623

- Support Vector Machines (SVM)
- Logistic Regression

- Decision Tree
- Random Forest

2) *Deep learning models*: Advanced deep learning models were also explored, including:

- Bi-LSTM (Bidirectional Long Short-Term Memory)
- BERT
- IndoBERT [19], a transformer-based language model specifically designed for Indonesian.

We use Optuna for automated hyperparameter optimization [20] for traditional models and Bi-LSTM. For the transformer-based model, we use the default configuration of the models in HuggingFace with learning rate of $1e-5$.

3) *Evaluation*: Models were evaluated on two datasets:

- The newly constructed cross-domain corpus.
- A COVID-specific dataset from previous research [7].

The evaluation metrics included:

- Accuracy: Proportion of correctly classified instances.
- Macro Precision: Average precision across all classes.
- Macro Recall: Average recall across all classes.
- Macro F1-Score: Harmonic mean of precision and recall.

These metrics provide a comprehensive assessment of model performance, particularly in imbalanced datasets.

The results will be compared with classifications using a corpus from previous research (COVID Data) [7] as the training data. This comparison aims to assess the stability and reliability of the newly constructed corpus. The evaluation will be conducted using standard performance metrics, including accuracy, macro-precision, macro-recall, and macro F1-score.

IV. RESULT AND DISCUSSION

In this section, we will discuss the dataset developed from the data collection and annotation process as explained in Section III-A and the preliminary experimental result as explained in Section III-B

A. Dataset Results

TABLE II. DATASET LABEL DISTRIBUTION

Label	No.of Tweet	Percentage
<i>Misinformation</i>	1590	44
<i>True Information</i>	762	21
<i>Not Sure</i>	82	2
<i>deleted</i>	1189	33
Total	3623	100

The tweet crawled in the data collection process yielded a total of 3,623 tweets. The distribution of tweets for each topic is presented in Table III. From the table it is evident that the topic with the highest number of tweet is COVID followed by HIV, Cancer and Cojoined Twins. The prominence of COVID

data is likely because there are much more news article related to COVID. Moreover, COVID is a phenomenon with global impact and widespread discussion. The topic of cojoined twins also tweeted by many people because the content is related to myths that are common in Indonesian society. In contrast, other topics has limited number of tweets likely because the content is not widely known and discussed.

From the collected data, 1189 tweets were deleted. These tweets were excluded because they are not written in Indonesian, unrelated to health topic or duplicate. The remaining tweets were annotated into two labels: MISINFORMATION and *True*, as shown in Table II. Among these, 1,590 tweets (44%) were labeled as *misinformation*, representing the largest portion of the dataset. The *True Information* label includes 762 tweets (21%) that were verified as accurate and reliable. A small subset of the data, 82 tweets (2%), was labeled as *Not Sure*, indicating cases where annotators found it difficult to determine whether the tweet is misinformation or true information. We reach an inter-annotator agreement with a Cohen's Kappa value of 0.91, which indicates almost perfect agreement.

B. Preliminary Results

In the preliminary experiment, we utilized the dataset that labeled as *misinformation* or *True*, in total of 2,352 data. In the experiment, 80% of the data was allocated for training, while the remaining 20% was used for data testing. The distribution of the data training and data testing is shown in Table III.

TABLE III. DISTRIBUTION OF TRAINING AND TESTING DATA ON EACH TOPIC

Topic	Testing	Training	Total
Covid	336	655	991
High Blood Pressure	0	24	24
Fertility	0	18	18
Avian Flu	0	17	17
Kidney Failure	0	7	7
GERD	0	67	67
HIV	134	511	645
Heart Problem	0	7	7
Cancer	0	167	167
Cojoined Twins	0	165	165
Cholesterol	0	17	17
Near-sightedness	0	3	3
Brain, bone marrow, diabetes	0	30	30
Polio	0	91	91
Toothache	0	3	3
Dyspnea	0	46	46
Stroke	0	54	54
Total	470	1882	2352

The outcomes of our experiments are presented in Table IV. We employed traditional machine learning approaches, including Naive Bayes, SVM, Logistic Regression, Decision Tree, and Random Forest as well as deep learning approaches, including Bi-LSTM, BERT, and Indo-BERT. We conducted two experiments for each model, utilizing the COVID-specific dataset and the cross-domain dataset as the training data. The performance of the trained models was evaluated using test data derived from the cross-domain dataset.

The results clearly demonstrate that models trained on the cross-domain dataset consistently outperformed those trained on the COVID-specific dataset. This trend was observed across

TABLE IV. RESULTS OF MISINFORMATION DETECTION EXPERIMENT USING COVID DATASET AND CROSS-DOMAIN DATASET

Model	Covid Dataset				Cross-Domain Dataset			
	macro-P	macro-R	macro-F	Acc	macro-P	macro-R	macro-F	Acc
Naive Bayes	0.601	0.581	0.583	0.668	0.734	0.735	0.735	0.768
SVM	0.623	0.580	0.583	0.668	0.844	0.800	0.816	0.849
Logistic Regression	0.591	0.557	0.550	0.670	0.833	0.807	0.818	0.847
Decision Tree	0.546	0.542	0.542	0.617	0.818	0.806	0.811	0.838
Random Forest	0.589	0.592	0.590	0.636	0.885	0.762	0.791	0.842
Bi-LSTM	0.599	0.611	0.598	0.623	0.828	0.773	0.792	0.832
BERT	0.595	0.597	0.596	0.643	0.766	0.776	0.770	0.796
IndoBERT	0.597	0.599	0.598	0.647	0.844	0.856	0.849	0.866

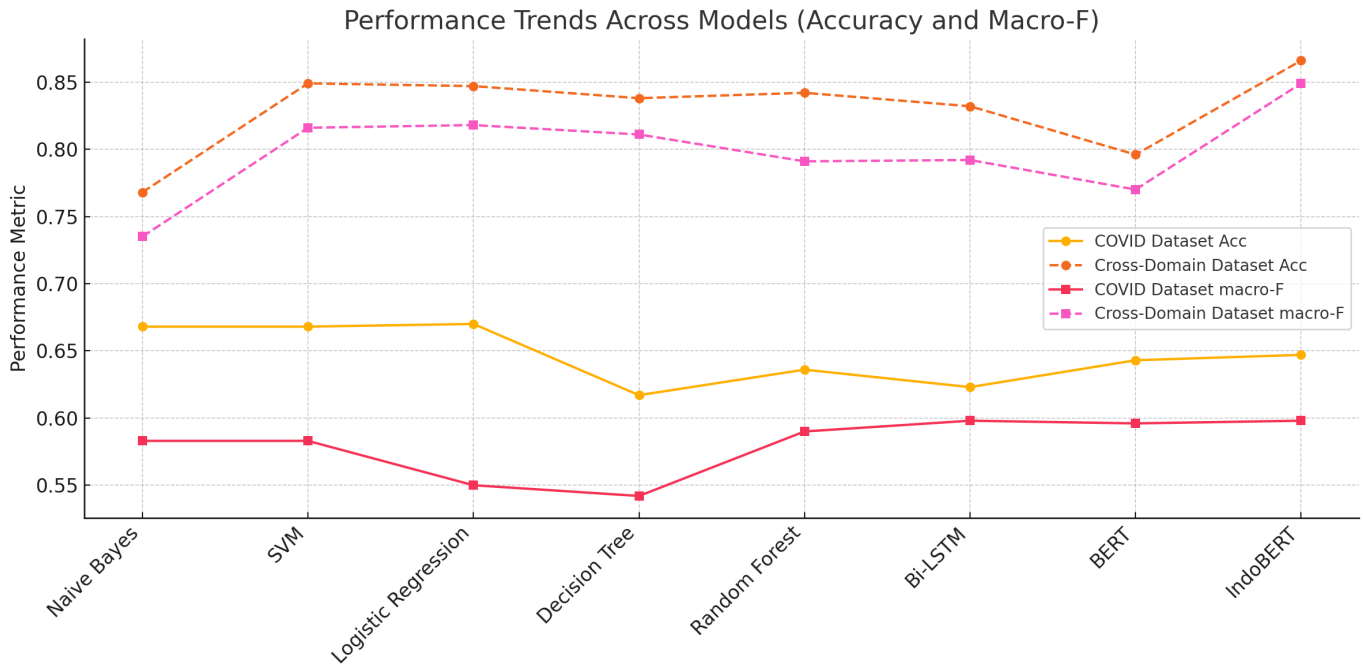


Fig. 2. Performance trends across models.

all models, both traditional and deep learning techniques. Notably, the use of cross-domain data appears to enhance the ability of the models to generalize across diverse contexts and topics, leading to improved predictive accuracy and robustness. The observed improvements in performance suggest that incorporating data from a variety of domains can mitigate the limitations of topic-specific datasets, which may lack diversity.

The results, as visualized in the Fig. 2, demonstrate significant differences in model performance when trained on the COVID dataset and the Cross-Domain dataset. Models trained on the Cross-Domain dataset consistently outperform those trained on the COVID dataset in both accuracy and macro-F scores. This highlights the superior generalization capability of the Cross-Domain dataset, which encompasses a broader range of health misinformation topics. Among the evaluated models, IndoBERT achieved the highest performance, with an accuracy of 0.866 and a macro-F score of 0.849 when trained on the Cross-Domain dataset. This underscores the effectiveness of transformer-based language models in handling complex linguistic features and diverse misinformation scenarios.

Simpler models such as Naive Bayes and Decision Tree also benefit from the Cross-Domain dataset, showing improved

metrics compared to training on the COVID dataset. However, their performance is limited relative to more advanced models, reflecting their inability to capture nuanced patterns in the data. Models like SVM and Logistic Regression exhibit significant improvements with the Cross-Domain dataset, suggesting their adaptability to diverse training data while remaining computationally efficient. Meanwhile, deep learning models such as Bi-LSTM show stable improvements, but they are outperformed by transformer-based models like BERT and IndoBERT, indicating the latter's superior contextual understanding.

The comparison between datasets underscores the importance of dataset diversity in training robust misinformation detection models. The Cross-Domain dataset enables models to generalize across various health misinformation topics, in contrast to the COVID dataset, which limits models to a single domain. These results demonstrate the Cross-Domain dataset's ability to enhance training effectiveness, making it a valuable resource for developing generalizable health misinformation detection systems.

The heatmap shown in Fig. 3 offers a detailed view of the interplay between models, datasets, and evaluation metrics, providing unique insights beyond the trends observed

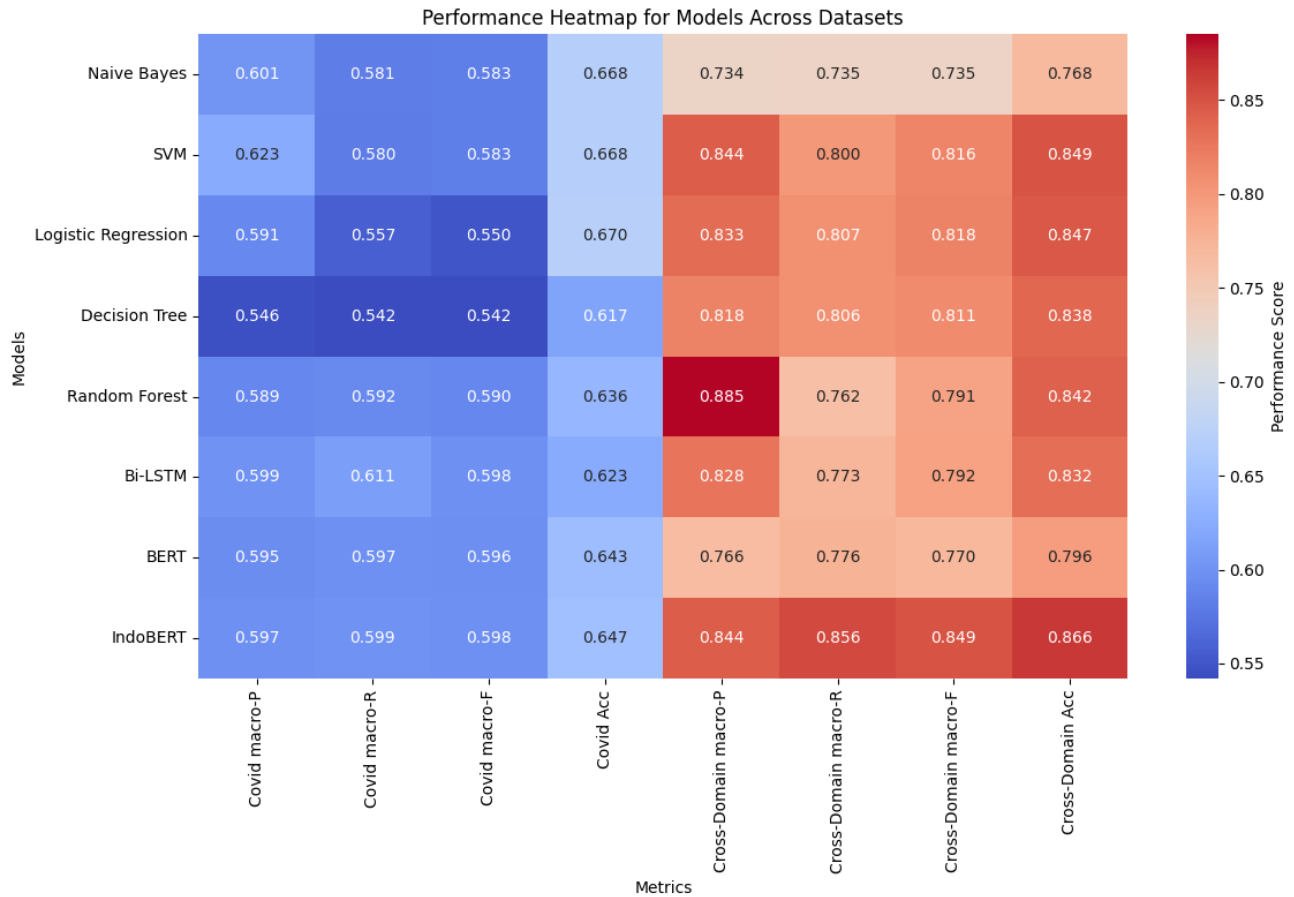


Fig. 3. Performance heatmap for models across datasets.

in the line charts. One key observation is the variability across metrics, where certain models demonstrate significant strengths in specific areas. For instance, IndoBERT exhibits a substantial improvement in macro-recall when trained on the Cross-Domain dataset (0.856) compared to the COVID dataset (0.599), indicating its effectiveness in reducing false negatives across diverse health misinformation topics. Additionally, the heatmap reveals trade-offs between precision and recall. For example, Random Forest achieves an exceptionally high macro-precision (0.885) on the Cross-Domain dataset but shows a relatively lower macro-recall (0.762), suggesting its strong capability in identifying true positives but limited sensitivity to all relevant instances.

The sensitivity of models to dataset diversity is also apparent, with traditional models such as Naive Bayes and Decision Tree showing larger relative gains from the COVID dataset to the Cross-Domain dataset, particularly in precision and accuracy. This highlights the disproportionate benefit simpler models derive from richer training data. Moreover, metric-specific strengths are evident; Random Forest excels in macro-precision, while IndoBERT consistently outperforms other models in macro-recall and macro-F, reflecting its ability to balance precision and recall effectively. Some models, such as BERT, display only modest improvements across datasets, suggesting the need for additional fine-tuning or data augmentation to fully exploit the diversity of the Cross-Domain

dataset.

Overall, the heatmap underscores the uniform performance boost across all models when using the Cross-Domain dataset, validating its robustness and utility for training generalizable health misinformation detection systems. The variability in metric-specific performance and model sensitivity offers complementary insights, enriching the understanding of model behaviors across diverse datasets.

V. CONCLUSION

In conclusion, this research introduces a novel cross-domain dataset for health misinformation detection in Indonesian tweets. The primary objective was to address the limitations of domain-specific datasets, such as those focused solely on COVID-19, and to evaluate the efficacy of machine learning models in generalizing across diverse health misinformation topics. Through comprehensive experiments using traditional and deep learning models, our study demonstrated that the Cross-Domain dataset significantly improves model performance across all evaluation metrics, including accuracy, macro-precision, macro-recall, and macro-F score. Models trained on the Cross-Domain dataset consistently outperformed those trained on the COVID dataset, underscoring the value of diverse and representative training data in developing robust misinformation detection systems.

Our findings highlight the effectiveness of advanced models such as IndoBERT, which achieved the highest performance metrics and demonstrated exceptional adaptability to the linguistic and contextual diversity present in the dataset. Traditional models, while showing notable improvements, remained limited in their ability to capture nuanced patterns, further emphasizing the importance of leveraging state-of-the-art methods for complex tasks like health misinformation detection.

This study contributes to the field by providing a high-quality, cross-domain dataset and presenting evidence of its potential to enhance machine learning models' generalization capabilities. These findings address the research questions posed in this study, particularly regarding the construction of a cross-domain dataset and its impact on misinformation detection models. Future work could focus on expanding the dataset to include other health misinformation sources, exploring multilingual approaches, and refining machine learning techniques to further improve performance. By addressing challenges in low-resource linguistic contexts, this research paves the way for more effective and scalable solutions to combat health misinformation.

ACKNOWLEDGMENT

This work has been funded by University of Gadjah Mada under Grant Number 6541/UN1.P1/PT.01.03/2024 with title "Sistem Deteksi Misinformasi Kesehatan Lintas Domain Berbasis Artificial Intelligence pada Media Sosial di Indonesia".

REFERENCES

- [1] V. Suarez-Lledo and J. Alvarez-Galvez, "Prevalence of health misinformation on social media: systematic review," *Journal of medical Internet research*, vol. 23, no. 1, p. e17187, 2021.
- [2] S.-F. Tsao, H. Chen, T. Tisseverasinghe, Y. Yang, L. Li, and Z. A. Butt, "What social media told us in the time of COVID-19: a scoping review," *The Lancet Digital Health*, vol. 3, no. 3, pp. e175–e194, 2021.
- [3] J. R. Bautista, Y. Zhang, and J. Gwizdka, "Healthcare professionals' acts of correcting health misinformation on social media," *International Journal of Medical Informatics*, vol. 148, p. 104375, 2021.
- [4] World Health Organization. (2020) Fighting misinformation in the time of covid-19: one click at a time. Accessed: 2024-12-05. [Online]. Available: <https://www.who.int/news-room/feature-stories/detail/fighting-misinformation-in-the-time-of-covid-19-one-click-at-a-time>
- [5] J. R. Zucker, J. B. Rosen, M. Iwamoto, R. J. Arciuolo, M. Langdon-Embry, N. M. Vora, J. L. Rakeman, B. M. Isaac, A. Jean, M. Asfaw *et al.*, "Consequences of undervaccination—measles outbreak, new york city, 2018–2019," *New England Journal of Medicine*, vol. 382, no. 11, pp. 1009–1017, 2020.
- [6] Y. Zhao, J. Da, and J. Yan, "Detecting health misinformation in online health communities: Incorporating behavioral features into machine learning based approaches," *Information Processing & Management*, vol. 58, no. 1, p. 102390, 2021.
- [7] D. R. Faisal and R. Mahendra, "Two-stage classifier for covid-19 misinformation detection using bert: a study on indonesian tweets," *arXiv preprint arXiv:2206.15359*, 2022.
- [8] I. Puspitasari and A. Firdauzy, "Characterizing consumer behavior in leveraging social media for e-patient and health-related activities," *International journal of environmental research and public health*, vol. 16, no. 18, p. 3348, 2019.
- [9] B. Zhong, "Going beyond fact-checking to fight health misinformation: A multi-level analysis of the twitter response to health news stories," *International Journal of Information Management*, vol. 70, p. 102626, 2023.
- [10] L. H. Suadaa, I. Santoso, and A. T. B. Panjaitan, "Transfer learning of pre-trained transformers for covid-19 hoax detection in indonesian language," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 3, pp. 317–326, 2021.
- [11] A. P. Rifai, Y. P. Mulyani, R. Febrianto, H. M. Arini, T. Wijayanto, N. Lathifah, X. Liu, J. Li, H. Yin, Y. Wu *et al.*, "Detection model for fake news on covid-19 in indonesia," *ASEAN Engineering Journal*, vol. 13, no. 4, pp. 119–126, 2023.
- [12] A. R. Prasetyo, I. Indriati, and P. P. Adikara, "Klasifikasi hoax pada berita kesehatan berbahasa indonesia dengan menggunakan metode modified k-nearest neighbor," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 12, pp. 7466–7473, 2018.
- [13] H. Mustofa and A. A. Mahfudh, "Klasifikasi berita hoax dengan menggunakan metode naive bayes," *Walisongo Journal of Information Technology*, vol. 1, no. 1, pp. 1–12, 2019.
- [14] D. Maulina and R. Sagara, "Klasifikasi artikel hoax menggunakan support vector machine linear dengan pembobotan term frequency–inverse document frequency," *Jurnal Mantik Penusa*, vol. 2, no. 1, 2018.
- [15] M. A. Rohman, D. Khairani, K. Hulliyah, P. Riswandi, I. Lakoni *et al.*, "Systematic literature review on methods used in classification and fake news detection in indonesian," in *2021 9th International Conference on Cyber and IT Service Management (CITSM)*. IEEE, 2021, pp. 1–4.
- [16] H. M. Arini, T. Wijayanto, N. Lathifah, Y. P. Mulyani, A. P. Rifai, X. Liu, J. Li, and H. Yin, "Detecting fake news during covid-19 in indonesia: the role of trust level," *Journal of Communication in Healthcare*, vol. 17, no. 2, pp. 180–190, 2024.
- [17] N. Kansal, L. Goel, and S. Gupta, "A literature review on cross domain sentiment analysis using machine learning," *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines*, pp. 1871–1886, 2022.
- [18] K. K. dan Informatika Republik Indonesia. (2024) Hoaks: Buah Pala dan Gula Batu Mampu Mengatasi Jantung Berdebar. Accessed on December 4, 2024. [Online]. Available: <https://www.komdigi.go.id/berita/klarifikasi-hoaks/detail/hoaks-buah-pala-dan-gula-batu-mampu-mengatasi-jantung-berdebar>
- [19] F. Koto, A. Rahimi, J. H. Lau, and T. Baldwin, "IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP," in *Proceedings of the 28th International Conference on Computational Linguistics*, 2020, pp. 757–770.
- [20] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 2623–2631.

Comparison of Machine Learning Algorithms for Malware Detection Using EDGE-IIoTSET Dataset in IoT

Jawaher Alshehri, Almaha Alhamed, Mounir Frikha, M M Hafizur Rahman
Department of Computer Networks Communications, King Faisal University, CCSIT,
Al Hofuf, Al Hassa 31982, Saudi Arabia

Abstract—The growth of IoT devices has presented great vulnerabilities leading to many malware attacks. Existing IoT malware detection methods face many challenges; including: device heterogeneity, device resource restrictions, and the complexity of encrypted malware payloads, thus leading to less effective conventional cybersecurity techniques. This study's objective is to reduce these gaps by assessing the results obtained from testing five machine learning algorithms that are used to detect IoT malware by applying them on the EDGE-IIoTSET dataset. Key preprocessing steps include: cleaning data, extracting features, and encoding network traffic. Several algorithms used these include: Logistic Regression, Decision Tree, Naïve Bayes, KNN, and Random Forest. The Decision Tree model achieved perfect accuracy at 100%, making it the best-performing model for this analysis. In contrast, Random Forest delivered a strong performance with an accuracy of 99.9%, while Logistic Regression performed at 27%, Naïve Bayes at 57%, and KNN with moderate performance. Hence, the results have shown the effectiveness of machine learning techniques to enhance the security IoT systems regarding real-time malware detection with high accuracy. These findings are useful input for policymakers, cybersecurity practitioners, and IoT developers as they develop better mechanisms for handling dynamic IoT malware attack incidents.

Keywords—IoT malware; machine learning; malware detection; IoT security; EDGE-IIoTSET

I. INTRODUCTION

The Internet of Things (IoT) is comprises a huge variety of devices connected to one another and exchanging information. These devices include, smart home devices, medical equipment, and industrial machinery, sensors, and wearable technologies. The high and rapid growth of IoT has effected transformation in a variety of fields and sectors like healthcare, transportation, commerce, agriculture, and education [1], [2], [3], [4], [5]. With that in view, IoT has become widely embraced as driving economic growth and improving quality of life, resulting in unbridled worldwide creation of new appliances and projects. This growth comes with enormous security challenges and problems in IoT devices. Currently, these are the prime targets for cyber criminals amidst other digital device. Of these, malware attacks are becoming one of the most salient and dangerous threats in IoT. Most IoT devices are not well-protected and have low computation abilities, which makes them high-value targets for malware attacks despite previous studies were mainly aimed at improving malware detection using sophisticated machine learning techniques, there is still a huge gap between the application of such techniques

on resource-constrained IoT devices with severely constrained performance and real-time response. Malware analysis remains crucial in IoT systems for understanding, detecting, and mitigating these threats. It forms part of the insight into how malicious actors compromise devices and networks; hence, it is an essential element of IoT security.

The IoT is both a network and a system. The definition of a network is that it can make communication possible between connected devices [6]. It is, in the same moment, considered as a system as it combines other elements and technologies to allow communication and data exchange between devices. Despite the fact that IoT technology offers so many benefits, it has created enormous potential weaknesses that impact performance and effectiveness in prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow. operation concurrent with its rapid growth. These malware might cause major financial and operational losses [7]. For instance, a well-publicized, major DDoS attacks on IoT devices and systems are typically executed by a botnet like Mirai [8]. Having fallen to the attacker, the device can then be used to execute other dangerous attacks. This malware is still a serious and evolving irritant in the modern digital world [9]. Thus, malware research will become extremely important for the security analysts and researchers as they try to comprehend different varieties of malware and take countermeasures against them.

It divided into two kinds of analysis: dynamic and static. The dynamic considers malware in an active state, whereas the status is considered for malware in an inactive state. Both are great significance in understanding how to protect IoT devices against malicious activities in-depth. The level of analysis and understanding of the capabilities of malware very much depends on how one can keep IoT safe from hacking and breaches of privacy [10]. Because of these, among other constraints, traditional malware detection techniques do not work well in the IoT environment with very restricted processing power and storage. This work therefore goes ahead to issues relating to IoT malware detection, but with more focus on how effective machine learning algorithms prove. For this reason, we compare various models and find the most suitable one for real-world applications in protecting IoT systems.

These methods are promising but require quite a lot of computational resources, which are difficult to handle using

typical IoT environments with device heterogeneity and less complex processing. This mismatch between capability and necessity creates a critical vulnerability in IoT security frameworks [46]. Most recent studies focus on intensive models including CNNs and LSTMs, which are undoubtedly accurate but still are unsuitable on resource-constrained IoT devices for real-time deployment. Additionally, the diverse malware types are often addressed inadequately. As most of the models focus solely on botnets [47]. The continuing evolution of IoT malware requires detection strategies that are not only effective but also adaptable to the constrained environments typical of many IoT devices. Addressing these gaps is essential.

This study addresses these gaps by analysing lightweight machine learning models to detect IoT malware explicitly accounting diverse malware types. Our research systematically benchmark five machine learning models: Decision Tree, Random Forest, Naïve Bayes, KNN, and Logistic Regression by using the EDGE-IIoTSET dataset [48]. This work contributes to identifying the most effective algorithm, and proposes a scalable and suitable approach for real-time application in a heterogeneous IoT ecosystem. The main objective of this paper is to evaluate the performance of lightweight machine learning algorithms. This paper provides actionable insights regarding the selection and optimization of machine learning algorithms in order to enhance IoT security. Among the research questions that need to be addressed in this study are as follows:

RQ1: What extent can machine learning algorithms effectively detect malware in IoT devices, considering the challenges of device heterogeneity, limited resources, and encrypted malware payloads? RQ2: Which machine learning models are most effective for real-time malware detection in resource-constrained IoT environments?

By exploring these questions, our study aims to give actionable insights that guide the development of more robust and scalable malware detection models tailored for the diversity and dynamic nature of IoT systems, making sure they remain reliable and secure against evolving threats.

II. LITERATURE REVIEW

IoT malware detection has acquired significant attention, and machine learning has grown as a prominent technique to address these threats. This section explores recent developments for detecting IoT malware, identifies gaps in the current literature, and compares the effectiveness of various machine-learning models in malware detection.

A. Recent Development in IoT Malware Detection using Machine Learning

A considerable amount of work is currently being performed for the development of machine learning-based models for the detection of IoT malware. Most of them feature network traffic analysis in search of suspicious patterns and label malicious behaviors using IoT-specific data. For example, the work explores deep learning approaches for detecting botnet activity in IoT devices. The researchers' results demonstrated that CNNs can outperform other ML approaches, including Support Vector Machines and Decision Trees, with a classification accuracy greater than 95% [11]. Following this line of investigation, the research presents an LSTM model for

malware detection based on IoT device behavior [12]. It was also found that LSTMs identify time-based patterns in traffic data, reporting an accuracy of around 97%.

The authors of [13] performed a performance study on the application of XGBoost and LightGBM to IoT malware detection. They concluded that LightGBM was most relevant in real-time detection since it computes much more quickly and can save memory compared to Random Forest. With these developments come challenges. Most of the related works considered do not address heterogeneity in IoT devices with their limited computational resources or the payloads in encrypted form arising from specific IoT devices, hence hampering the performance of these ML algorithms [14]. It is such a gap that our research sets out to fill, ensuring focus on lightweight models for efficiency in resource usage but high accuracy in malware detection.

B. Current Gaps in the Literature

Even though the area of IoT malware detection has developed, some gaps still exist in the literature. Most related studies tend to ignore the restricted computational resources provided by IoT devices. For example, although certain studies reported high accuracy using CNNs and LSTMs, they are computationally expensive and hence cannot be realistically deployed on resource-constrained IoT devices [15]. Also, most of these research works pertain to malware types like botnets alone and not all variants of IoT malware, such as ransomware, spyware, and worms [16]. The other limitation is that the literature does not focus on encrypted traffic, which originates from IoT devices. In many cases, the IoT devices encrypt data due to privacy issues; hence, malicious activity detection solely based on encrypted network traffic is limited. Most of the research has focused on plaintext traffic; hence, encrypted network traffic is highly neglected and further limits applicability of in real-time scenarios [17]. Our research covers these gaps by assessing machine-learning models applicable to resource-constrained devices and encrypted traffic analysis.

C. Comparison Study of Various ML Models for IoT Malware Detection

A number of machine learning algorithms have been tried and tested for IoT malware detection; each of them has pros and cons. Recently, XGBoost and LightGBM have gained major attention because of their speedy and bulky handling of data [18]. XGBoost prevents problems of overfitting by using regularization techniques; hence, it is a robust choice in malware detection for IoT environments, which are dynamic. At the same time, it requires very heavy computation, which makes it computationally prohibitively expensive for resource-constrained IoT devices [19]. On the other hand, LightGBM is more resource-efficient and has faster training times than XGBoost, making it more suitable for real-time malware detection in IoT systems [20]. Researchers in [21] found that LightGBM achieved comparable accuracy to XGBoost but used less memory and CPU, making it ideal for low-powered IoT devices. Nevertheless, DNNs have proven quite promising in developing complex patterns from network traffic data, be it CNNs or LSTMs [22]. However, DNNs still suffer from serious computations, which are particularly not suitable for real-time IoT applications [23].

Thus, the lightweight models like Random Forest or Decision Tree are practical models for real-world IoT malware detection [24]. Therefore, our work demonstrates that, in terms of the trade-off between accuracy and computational efficiency, among the techniques under study, Random Forest achieves the best performance and is deployable on resource-constrained IoT devices. Although machine learning has made major advancements in the detection of malware on IoTs, there is still a gap in the literature regarding model suitability for resource-constrained IoT devices and how it handles encrypted traffic. This paper largely extends works that have been previously performed to assess lightweight models and address challenges we have pinpointed in order to create more pragmatic IoT malware detection.

In previous studies, several machine learning methods have been proven to be effective in detecting IoT malware. For instance, Sliwa, Piatkowski, and Wietfeld (2020) demonstrated that Random Forest algorithms can offer reliable malware detection in IoT devices; however, they also identified some disadvantages when dealing with encrypted traffic data. Additionally, Zhang and Zhou (2021) revealed that SVMs excelled in very high-dimensional data scenarios, thereby further indicating how the performance of the algorithm would vary with regard to data characteristics (Sliwa et al., 2020; Zhang & Zhou, 2021).

III. CHALLENGES IN IOT MALWARE ANALYSIS

The analysis of IoT malware shows different challenges due to the IoT ecosystem's nature. This section offers insight into the heterogeneity and diversity of IoT devices, the resource limitations of these constrained environments, encrypted and obfuscated malware payloads, and the privacy concerns and regulatory challenges associated with managing IoT data.

A. Heterogeneity and Diversity of IoT Devices

Heterogeneity in IoT indicates the different array of elements, in terms of various protocols, devices, services, and networks within an ecosystem, highlighting the complexity and variability of interconnected elements [25], [26]. Interoperability indicates the key challenge in heterogeneous IoT platforms due to diverse methods for recognizing and identifying devices within different platforms, as well as resource requests. These differences are huge hurdles and create hindrances in data exchange and smooth communication. It is crucial to bridge these interoperability gaps, to ensure seamless operation and secure data exchange within the IoT environment [26]. In addition, integrating diverse IoT technologies and devices from different vendors into a cohesive and unified system can present complex and significant challenges. Each device may have its own application programming interfaces, which complicates data-sharing and integration efforts. Diverse communication protocols also complicate the establishment of connections and data exchange. Metrological characteristics also have also proved to be a significant concern in the integration process due to inconsistencies in measurement units, range, accuracy, and scale among different devices, since every device has unique features. Ensuring temporal consistency and synchronization across many IoT devices, especially in real-time applications, further increases complexity [27].

B. Resource Limitations and Constrained Environments

Few resources and a constrained environment mean IoT devices have limited energy, memory, and processing-power resources. The result of these constraints is the limited ability to implement trade resource-intensive security measures for devices. Additionally, IoT devices are often connected over a lossy link. During the transmission of data, lossy links may have a significant chance of packet loss. Environmental factors, signal attenuation, and wireless interference may cause problems that compromise the security and dependability of IoT networks. As a result, packet loss, delay, and erratic communication between devices may occur. So, mitigating the effects of lossy connectivity is necessary for the security of IoT devices with limited resources [28].

C. Encrypted and Obfuscated Malware Payloads

The identification and analysis of IoT malware could be difficult because malicious programs use various methods to encrypt and hide their payloads. Due to encryption, the payload's exact purpose remains hidden, making it more challenging for traditional antivirus programs to identify and address malicious code. Obfuscation techniques are also used to make it more difficult to study the malicious code, employing strategies like code obfuscation to purposefully make the code more complex and difficult to interpret. This hinders the analysts' ability to comprehend its behavior [29].

D. Privacy Concerns and Regulatory Challenges

IoT presents a multitude of privacy and regulatory concerns around the gathering, storing, and use of data produced by linked devices. Data security is one of the main concerns. As IoT devices expand, they produce vast amounts of data, including sensitive and personal information. Protecting the privacy of this data is mandatory, especially when it comes to personal information about an individual's actions and behaviors. Cyber risks include security breaches and illegal access that can compromise the confidentiality of the data gathered and sent by IoT devices. Addressing security concerns to protect IoT-generated data privacy is crucial [29]. These vulnerabilities, combined with the above challenges of IoT systems and networks, make comprehensive security management a challenging task. As this paper progresses, we will explore advanced detection techniques that aim to overcome these challenges and fortify the protection of IoT systems against evolving threats.

IV. METHODOLOGY

The materials and methods section outlines the overview of the used dataset, followed by the rationale for algorithm selection, machine learning in malware detection for IoT systems, and some details about the dataset preprocessing techniques used to ensure accurate results.

A. Machine Learning in Malware Detection for IoT Systems

In the complex ecosystems of Internet of Things systems, machine learning has emerged as a critical technique for enhancing virus detection. We used network traffic simulations to create our dataset in order to ensure that it appropriately reflects common IoT interactions and possible security breaches,

given the diversity and unpredictability present in IoT contexts. This method was chosen because it enhances the validity and application of our study by allowing the dataset to cover a wide range of real-world situations are quite helpful in spotting novel patterns that haven't been assigned a label yet.

Table I details the different types of machine learning approaches we employed and their specific application and benefits toward IoT security. A multi-pronged approach would not only guarantee complete coverage but also make the detection systems robust and reliable for the unexpected nature of IoT malware.

TABLE I. MACHINE LEARNING ALGORITHMS

Type	Description
Supervised learning	Methods include Support Vector Machines (SVMs) and Random Forests, which excel at classifying malware based on predefined labels. These algorithms are trained on datasets that humans provide to models.
Unsupervised learning	Includes Principal Component Analysis (PCA) and segregating data in the form of clusters, offering a complementary strategy by anomaly detection within IoT data. These techniques can detect and identify previously unseen malware variants without relying on pre-labeled data [41].
Deep learning models	Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), in particular, have remarkable capabilities for handling complex and large-scale IoT datasets. These models can effectively extract features from data, which can lead to better detection rates for sophisticated malware [42].
Hybrid models	To enhance detection capabilities, hybrid models combine strengths of both types of analysis-static and the dynamic analyses. While static analysis examines the code structure, dynamic analysis observes code behavior during execution. By integrating these perspectives, hybrid models are advanced models that can enhance the detection of both known and unknown malware threats.

This study investigates the deployment of machine learning to detect malware in internet-enabled gadgets. The research leverages the EDGE-IIoTSET dataset, and we preprocessed network traffic data to extract relevant features. We compared SVM, Random Forest, and CNN models to identify optimal algorithms for classifying malware. Standard metrics were used to evaluate model performance and execution, considering computational efficiency for practical IoT deployment.

Each machine learning model was carefully optimized to balance predictive accuracy with a computationally fair load. The Random Forest, for example, was set at a specific number of trees such that it would neither overfit, nor be impractical to use on limited resource devices typical for IoT setups, and SVMs were optimized about kernel type and the regularization parameters about the best discrimination between malicious traffic and benign one without requiring exhaustive computational resources.

Our approaches ensured that optimizations reduced computational overhead; otherwise, IoT resources are characterized by limited implementations, which might prove challenging. For instance, in the Random Forest approach, it was setup to have just a few trees so as to decrease model complexity and

runtime. Such made it feasible for real-time virus detection of devices in an Internet of Things device with processing capabilities.

B. Rationale for Algorithm Selection

The major concerns necessary for the applied machine learning algorithms to discover IoT malware are scalability and efficiency. Python and Scikit-learn were chosen with great care due to their excellent documentation, strong community support, and large selection of pre-built functions for data processing and machine learning. These characteristics make it easy for other researchers to replicate our methods. Python provides scalability and efficient calculation when working with massive datasets, therefore the choice also fits with the requirement for real-time processing capabilities in IoT systems. IoT networks may generate a huge volume of data. Therefore, the chosen algorithms must be able to scale up while handling large-sized datasets efficiently, without being computation-heavy for resource-constrained IoT devices. Malware detection algorithms raise concerns over precision and accuracy, which means they involve risks concerning false positives and false negatives [33]. Interpretability of models should be interpretable in the IoT environment. A security practitioner must understand why some network traffic was flagged as malicious by a model in order to take remedial action. Algorithms were selected based on the criteria below.

Stochastic Forests: The Random Forest algorithm was chosen because of its great ability to handle big feature sets and reduce overfitting problems [34]. Therefore, it was found that this model takes the average output of many created decision trees, meaning the variance in measurement would be lower than using one single Decision Tree model. Hence, it generalizes the model to new data.

Advantages: It is easily scalable for RF, and it contains a mix of categorical and numerical data to be executed efficiently. It is relatively faster and also has means to internally estimate the importance of features; hence, such aspects will be useful during real-time deployment in IoT systems [35].

Disadvantages: The model performance may decrease when the dimensionality is high in the feature space, so a dimensionality reduction method needs to be applied [36].

Support Vector Machines (SVMs): Support vector machines have been remarked on as doing quite well in high-dimensional spaces and with binary classification problems. For example, since the core of the problem would involve network traffic being classified as benign or malicious, this makes a SVM a natural choice [37].

Advantages: The feature space of SVM can be applied toward finding the optimal hyperplane separating the classes from each other, since the model would not be biased toward any particular class.

Disadvantages: One of the disadvantages is that SVMs can become extremely expensive when used with big data, and this may limit their applicability to real-time IoT scenarios [38].

Convolutional Neural Networks (CNNs): CNNs are the algorithms most used in image processing and, as recent studies have proved, do a great job with network traffic

analysis. These CNNs have been shown to learn complex patterns from network data and hence are likely to easily recognize sophisticated malware [39].

Advantages: CNNs can be very flexible and reveal even small and complex patterns in data that would be very useful for the detection of new malware samples.

Disadvantages: CNNs are quite resource-expensive algorithms for the user, which decreases its applicability in resource-limited IoT environments [40].

C. Dataset Overview

Our analysis of the industrial edge computing and IoT applications depended on the EDGE-IIoTSET dataset [30]. Due to its large amount of network traffic that is unique to the Internet of Things, comprising both malicious and benign payloads captured under controlled environments, the EDGE-IIoTSET dataset was selected. Our results can be reproduced and applied to other industrial applications due to the realistic simulation of IoT network environments this dataset provides. **Sampling Strategy:** To achieve the balanced representation on IoT risks, the strategy implemented was a stratified random sampling that would allow all malware types to have adequate representation.

The dataset was given several essential preprocessing steps to ensure it could be used the best way possible to fit into machine learning research. First, we removed all incomplete or outlier items within the raw data that could ruin the results. We then extracted relevant information from the network traffic data with the aid of feature engineering, including payload characteristics, protocol type, and packet size. The data characteristics were then scaled using normalization techniques so that our machine learning algorithms could read them efficiently and without favoring any one feature scale.

It contains packet or packet-related metadata network traffic information important for malware threat detection in IoT networks. Some of the features include the number of dimensions of the IoT communications, like payload sizes, types of protocols, and network latency. These features are among some of the most important in our model, giving insight into the nature of benign and malicious activities across the network. We use a recently constructed benchmark, the EDGE-IIoTSET dataset, specifically for machine learning applications in the context of the Industrial Internet of Things. That is, it is really a rich set of features aimed at capturing real IoT network traffic behavior, both benign and malicious.

1) Main Features of the Dataset: **Network Traffic:** The datasets are well complemented with packet size, the protocols of communication, and event timestamps within a network flow. **Packet Information:** This includes metadata that identifies each packet of traffic flowing across the network, including source and destination addresses, protocol types, and statistics regarding data flow. **Anomalies and Attacks:** Classification of all variations of the different attacks; DDoS attacks, provided as an example, are attacks of many other types of malware varieties attacking IoT devices. They include ransomware, spyware, and botnet threats, among others.

2) Challenges of the Dataset: **Diversity:** Data created from IoT devices is vastly heterogeneous owing to their diversity in functionality [31]. For example, data streaming from home security cameras, smart thermostats, and industrial sensors can be packaged into one dataset, which will act differently across a network. **Imbalanced Classes:** Most network traffic is benign rather than malicious, and this imbalance can create quite a tough challenge for most machine learning algorithms because the models might simply end up showing a preference for the benign traffic classes and finally yield poor performance in malware detection [32]. The EDGE-IIoTSET dataset was selected as it represents the diversities of real-world test cases, comprising IoT devices; therefore, it is rated among the best benchmarks to test the malware detection techniques.

3) Dataset and Preprocessing: ML-Edge IIoT-dataset.csv (EDGEIIOTSET Dataset) is a dataset designed for analysis and machine learning tasks within the edge and Industrial Internet of Things (IIoT) environments [43]. The main objective is to clean, transform, and prepare the data for training the machine learning model by removing unnecessary columns, handling missing values, and encoding categorical features. The dataset being used is (ML-EdgeIIoT-dataset.csv). This dataset likely contains different network-related features and attack types from edge and IIoT environments. It contains various types of network-related data and possibly some metadata from network communications. The dataset is initially loaded into a data frame using a Python library, which is called Pandas. The (low_memory=False) argument is helpful to optimize memory usage for reading large datasets. A predefined list of columns that are too specific to be useful (drop_columns) are considered irrelevant for the analysis. Those types of columns are eliminated from the dataset to maintain data integrity and to minimize dimensionality. The code drops rows that contain any missing values and removes duplicate rows to ensure that the dataset is clean and consistent. To enhance the dataset's usability, it is shuffled to randomize the order of the rows. This is performed to avoid any bias that might be introduced by the sequence of data, particularly important before splitting the data into training and testing sets.

Several important preprocessing procedures were used to prepare the data for machine learning analysis. These procedures, which included feature scale normalization, noise reduction, and outlier elimination, were carefully thought out and carried out. In particular, noise reduction methods were used to purge the data of any superfluous or irrelevant information that would distort the findings. Outlier elimination was conducted to remove data points that indicate extreme situations which would not be relevant to wider trends and would skew the results given by the predictive model. Several categorical columns (http.request.method, http.referer, http.request.version, dns.qry.name.len, mqtt.conack.flags, mqtt.protoname, and mqtt.topic) are transformed into dummy variables. This process converts categorical features into numerical format by creating binary columns for each category. This is crucial for ML algorithms that require numerical input. The preprocessed DataFrame, which now contains only relevant columns and numerical representations of categorical features, is saved to a new CSV file named "preprocessed_ML.csv". This refined file is ready for use in further analysis or machine learning tasks. Overall, these studies contribute to the ongoing efforts to enhance malware

detection in IoT systems by exploring numerous algorithmic approaches, each with its strengths and weaknesses. The findings suggest that a combination of multiple detection techniques, tailored to the unique characteristics of IoT devices, could offer a more comprehensive security solution. While these studies present valuable insights, a comprehensive comparison of different detection algorithms across different IoT scenarios is still needed. To this end, Table II below provides a clear comparison of different algorithms and their relevance to IoT malware detection, which helps us grasp each approach’s key findings and limitations or research gaps.

TABLE II. MACHINE LEARNING ALGORITHMS

Algorithm	Key Findings / strengths	Weaknesses /research gap	Suitable for IoT
Signature-based	High detection rate for known threats	Ineffective against new malware variants	Limited applicability in IoT due to rapid malware evolution
Anomaly detection	Effective in detecting unknown threats	High false positive rate	Requires extensive training data
Machine learning	Adaptability to new malware variants	Requires large datasets and computational resources	Potential for high accuracy
Deep learning	High accuracy in anomaly detection, and complex pattern recognition effective against known malware	Requires significant computational resources and expertise	Suitable for large-scale IoT environments Algorithm

V. PROPOSED MODEL

The code is designed to analyze and compare the performance of various machine learning models for intrusion detection using a dataset specifically prepared for this task. Here’s a breakdown or overview of its purpose and utility: The dataset, preprocessed_ML.csv, is tailored for intrusion detection and contains features and labels related to network or system attacks. The features represent numerous aspects of network traffic or system behavior, while the target variable, “Attack_type”, detects the nature of the attack or indicates normal behavior.

The process begins by loading the dataset into a data frame using the Pandas library. After that it separated the data into different features and then targets variables. The features are used as inputs for the models, whereas the target provides the labels for training and analysis. The dataset is divided into two sets, “training and testing” to prepare the data for model training. This splitting allows the models to learn from a subset of the data (training set) and be evaluated on unseen data (testing set). An 80/20 split is typically used, where 80% of the data is used for training and 20% for testing. A critical step in preprocessing is addressing class imbalance. Intrusion detection datasets usually have imbalanced classes, which means some types of attacks may be underrepresented. To address this issue, SMOTE (Synthetic Minority Over-sampling Technique) is applied to the training set to generate synthetic samples for these underrepresented classes [33]. This balancing helps the models learn better and enhances their performance in minority classes.

The code then initializes and trains five different machine learning models: Decision Tree, K-Nearest Neighbors (KNN),

Naïve Bayes, Logistic Regression, and Random Forest. Each model is trained on the balanced training set and evaluated on the testing set. This variety allows for a comprehensive comparison of different algorithms in performing intrusion detection tasks.

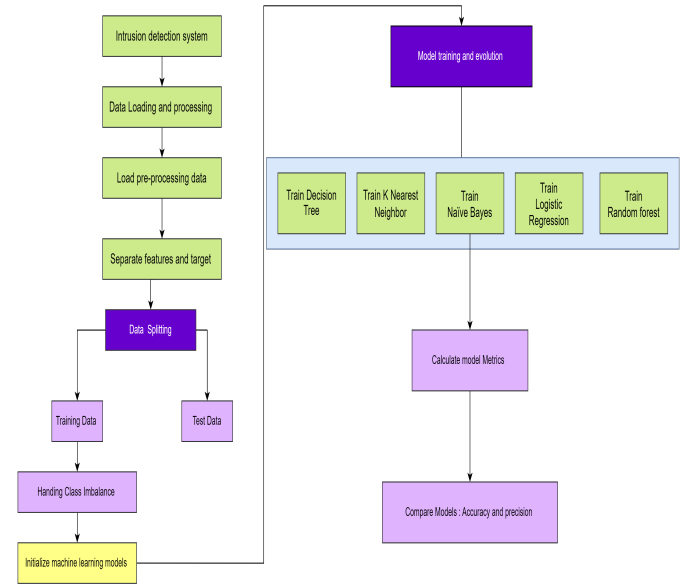


Fig. 1. Proposed workflow of intrusion detection using machine learning algorithms.

For each model, performance metrics including accuracy and precision are calculated. Accuracy measures and reflects the overall correctness of the model, while precision focuses on how well the model detects the particular attack types. Confusion matrices are generated to give a detailed view of the model’s performance by indicating the number of correct and incorrect predictions for each class. ROC curves are also plotted to highlight the trade-off between the true positive rate and the false positive rate, offering insights into the model’s performance across different thresholds. Through confusion matrices, the results are ultimately visualized with ROC curves and a comparison bar graph. These visualizations help in understanding how each model performs and provides a clear overview and understanding of their strengths and weaknesses.

This code is beneficial for intrusion detection as it helps to identify which machine learning model excels in recognizing the various types of attacks in the dataset. Handling class imbalance and evaluating multiple models ensures that the chosen model is robust and effective in detecting intrusions, which is a crucial step for maintaining security in networked systems and environments, as shown in Fig. 1.

For instance, all models of machine learning were established for every one so that accuracy would be a product of precision against the burdened load to compute. Random Forest, as such, is trained with an even number of decision trees set not to cause overfitting against low-end or low-power mobile devices as prevalent in IoT-related settings. Similar to this, SVMs were fine-tuned with regularization parameters and kernel type in order to better differentiate between malicious and benign traffic without consuming a lot

of processing power. Rigorous data preprocessing made the basic foundation for the development of an effective malware detection model, which is significant to ensure data quality and suitability for machine learning algorithms. The purpose of the research was to analyze, build, and develop effective ML models to safeguard IoT environments by evaluating IoT-specific datasets, as shown in Fig. 1.

VI. OUTCOMES

The Outcomes section presents the performance and findings of various machine learning models used in IoT malware detection, including Decision Tree, K-Nearest Neighbors (KNN), Naïve Bayes, Logistic Regression, and Random Forest, highlighting their accuracy, efficiency, and overall effectiveness in detecting malware within IoT systems.

In line with Sliwa et al. [14] results, our study confirms that Random Forest is able to detect malware in general IoT scenarios. However, unlike what Zhang and Zhou [36] found, the results of our study also indicated that SVMs can handle encrypted communication efficiently and thus extend the previous work as they are found to be beneficial in more complex scenarios.

A. Decision Tree

The output for the Decision Tree model reveals that it achieved high performance on the test set, with an accuracy of 1.0. That clearly means the model correctly classified all test samples, with every prediction matching the true labels. Despite this perfect accuracy, the reported precision is 0.0, which seems like the accuracy is inconsistent. Commonly, precision should be 1.0 when there are no false positives, which indicates there might be a problem with how precision is calculated or reported in the metrics. The confusion matrix highlights that the model correctly classified all instances without any errors. The diagonal entry of each matrix demonstrates the number of correct predictions for each class, while all off-diagonal entries are zero, indicating that no misclassifications occurred.

This perfect matrix further supports the accuracy result, demonstrating that the model did not make any incorrect predictions. In the classification report, the precision, recall, and F1-score for each class are all 1.00. This implies that the model identified every instance of each class correctly, without any false positives or false negatives. The F1-score is 1.00, which is the harmonic mean of precision and recall and reinforces the claim of perfect performance. Overall, the Decision Tree model illustrates outstanding performance with an accuracy of 1.00 and perfectly accurate predictions for all classes. However, the anomalous precision report of 0.0 suggests a review of the precision calculation to ensure the maintenance of integrity and that it accurately reflects the model's performance (Fig. 2 and Fig. 3).

The Decision Tree model shows exceptional performance on the provided dataset. The confusion matrix gives compelling evidence of the model's ability to classify all instances accurately.

B. K-Nearest Neighbors

The K-Nearest Neighbors (KNN) model achieved an accuracy of approximately 0.65, describing that the model correctly

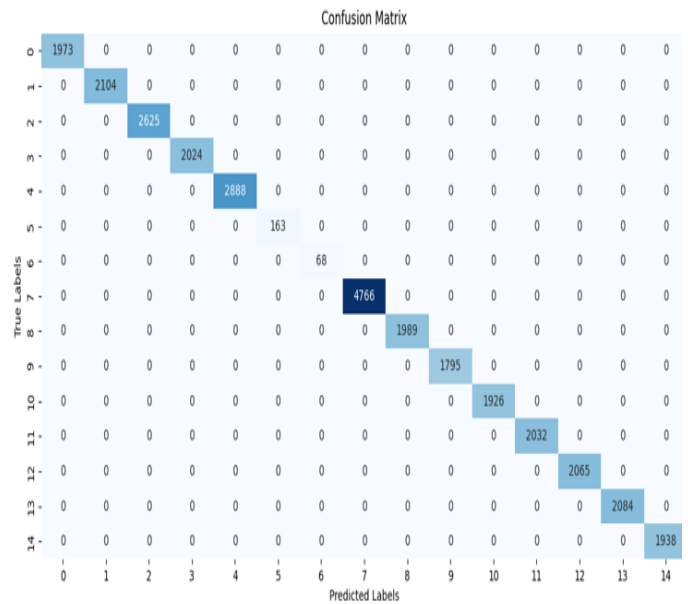


Fig. 2. Confusion matrix of the Decision Tree algorithm for classifying 15 categories using the EdgeIoT dataset.

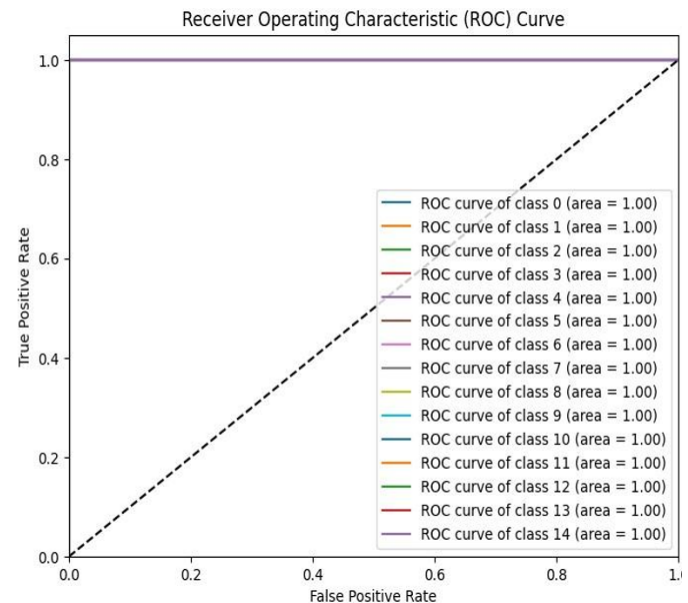


Fig. 3. Receiver Operating Characteristic (ROC) curve illustrating the performance of the Decision Tree algorithm across multiple thresholds using the EdgeIoT dataset.

classified around 65% of the test samples. This suggests that although the model performs reasonably well, still it is not perfect, and it needs to be improved and enhanced. The precision value is reported as 0.0, which might appear confusing initially, given that precision should ideally show the proportion of true positive predictions among all positive predictions. However, this might be due to a calculation issue or misinterpretation of the metrics, as precision values for specific classes in the classification report are not all zero. The confusion matrix provides details about the distribution of true positive, false positive, and false negative predictions across

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
0	1877	13	0	8	0	2	0	12	17	6	0	13	5	7	13
1	5	1273	0	73	0	17	0	88	184	75	7	147	107	69	59
2	0	0	2625	0	0	0	0	0	0	0	0	0	0	0	0
3	4	121	0	796	0	25	0	54	117	651	0	96	69	44	47
4	0	0	0	0	2888	0	0	0	0	0	0	0	0	0	0
5	0	6	0	4	0	137	0	1	3	6	0	4	2	0	0
6	0	0	0	0	0	0	68	0	0	0	0	0	0	0	0
7	23	450	0	261	3	27	0	2106	559	240	1	295	193	232	376
8	4	262	0	149	0	28	0	204	534	143	5	219	159	127	155
9	5	125	0	658	0	31	0	65	131	529	2	94	53	49	53
10	0	24	0	7	0	6	0	8	17	8	1794	14	30	8	10
11	9	241	0	131	0	25	0	141	295	92	2	790	158	69	79
12	11	206	0	86	0	19	0	94	203	61	8	214	1066	41	56
13	7	35	0	21	0	1	0	21	49	23	1	34	15	1849	28
14	6	54	0	23	0	2	0	61	87	32	0	27	24	40	1582
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14

Fig. 4. Confusion matrix of the K-Nearest Neighbors algorithm for classifying 15 categories using the EdgIoT dataset.

diverse classes. For example, the model shows relatively high accuracy for the ‘DDoS_ICMP’ and ‘DDoS_UDP’ classes, where it correctly classifies nearly all instances. However, performance is significantly lower for other classes such as ‘Password’ and ‘Port_Scanning’, where the model makes more errors.

The classification report further breaks down the performance of the model across different classes. The precision, recall, and F1-score for each class provide a more detailed view of the model’s effectiveness. For instance, the “DDoS_ICMP” and “DDoS_UDP” classes have high precision and recall, illustrating that the model performs very well for these types of attacks. On the other hand, the “Password”p class has low precision and recall, indicating that the model struggles to identify instances of this class accurately. Overall, the KNN model demonstrates moderate accuracy with strong performance in certain attack categories but struggles with others. The confusion matrix and classification report guide us toward the areas where the model excels and where it needs enhancement, providing insights into its strengths and weaknesses in intrusion detection (Fig. 4 and Fig. 5).

C. Naïve Bayes

The metrics for the Naïve Bayes model highlight an overall accuracy of approximately 0.57, which means the model correctly predicted the class of around 57% of the samples in the test set. This level of accuracy is relatively low compared to the Decision Tree model’s perfect score, suggesting that Naïve Bayes faces challenges in classifying the data effectively. The precision score of 0.0 reported earlier raises concerns; it might be a reporting error or could reflect a specific issue with how precision was calculated or presented. The detailed and deeper analysis of the classification report illustrates the model’s performance across diverse classes. The confusion matrix shows a clear picture of the distribution of correct and incorrect

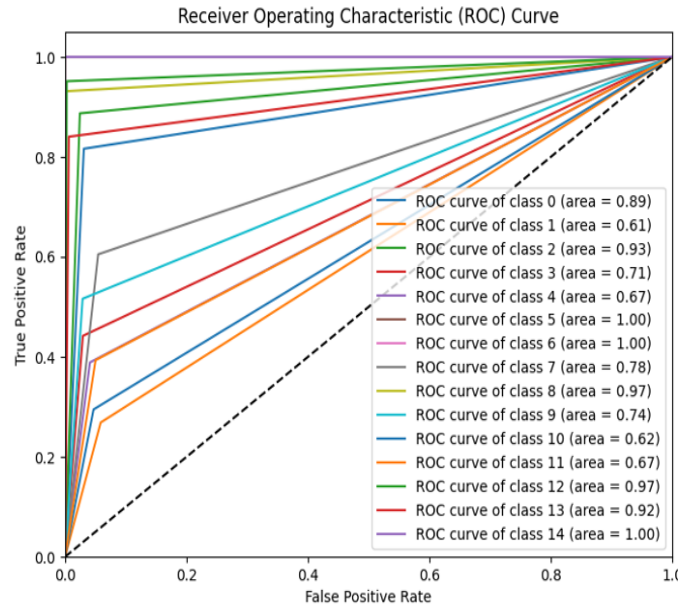


Fig. 5. Receiver Operating Characteristic (ROC) curve illustrating the performance of the K-Nearest Neighbors algorithm across multiple thresholds using the EdgIoT dataset.

predictions. For example, the Naïve Bayes model performs well on “DDoS_ICMP” and “DDoS_UDP” attacks, accurately predicting these classes with high precision. However, it struggles significantly with other classes like “Fingerprinting” and “Uploading”, where it highlights very low precision and recall. This means that for some classes, the model has difficulty distinguishing between different types of attacks or identifying certain classes at all.

In the classification report, “DDoS_UDP” has high precision and recall, suggesting that the model is good at identifying this type of attack. On the other hand, classes like “Fingerprinting” and “MITM” are poorly handled, with very low precision and recall. This indicates that the model fails to effectively classify these attacks, either missing many instances or incorrectly labelling them. Overall, the Naïve Bayes model shows mixed results with moderate accuracy but variable performance across different classes. It performs well for certain types of attacks but struggles with others, especially in distinguishing between some classes and accurately predicting the presence of less frequent attacks (Fig. 6 and Fig. 7).

D. Logistic Regression

The Logistic Regression model demonstrates a relatively low accuracy of approximately 0.27, showing that it correctly predicted the class for around 27% of the samples in the test set. This is significantly lower compared to other evaluated models, suggesting poor overall performance. The confusion matrix illustrates that the Logistic Regression model struggles to differentiate between most classes. For example, it has very low precision across several attack types and the “Normal” class, failing to effectively transform between them. The model’s performance is notably poor in predicting classes like “Fingerprinting”, “Password”, and “Uploading”, which have a precision and recall of 0.00. This indicates that the model

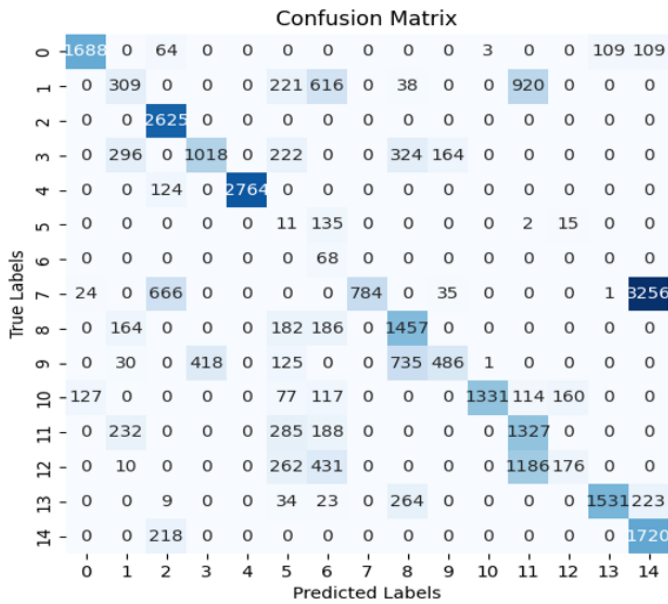


Fig. 6. Confusion matrix of the Naïve Bayes algorithm for classifying 15 categories using the EdgeIoT dataset.

classes being identified with high accuracy, while others being virtually ignored (Fig. 8 and Fig. 9.)

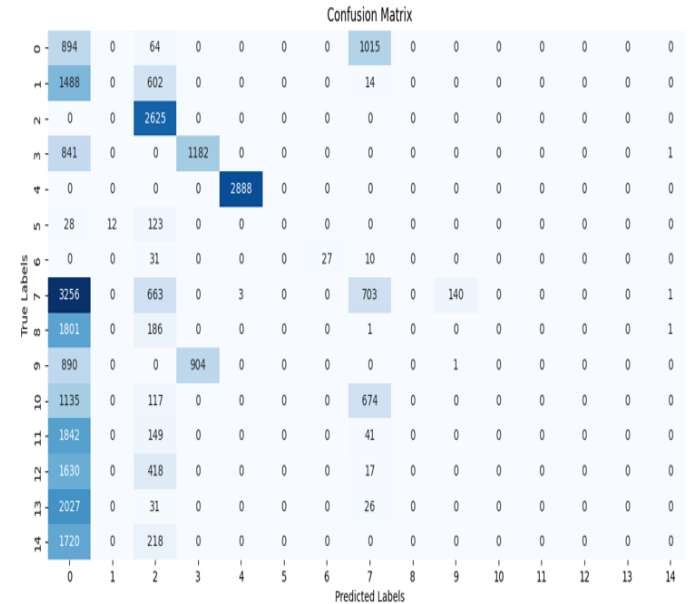


Fig. 8. Confusion matrix of the Logistic Regression algorithm for classifying 15 categories using the EdgeIoT dataset.

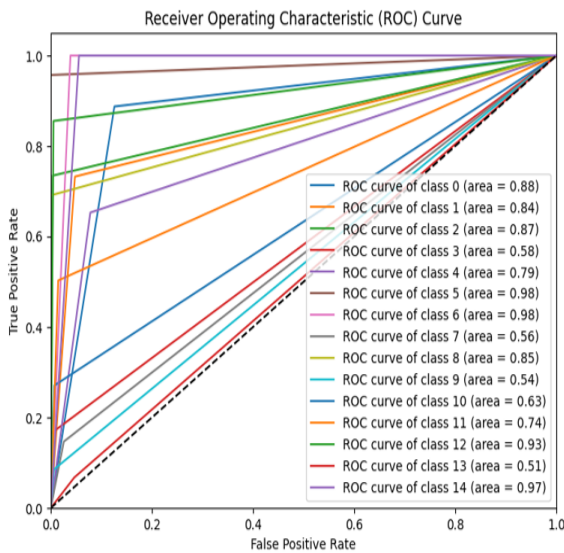


Fig. 7. Receiver Operating Characteristic (ROC) curve illustrating the performance of the Naïve Bayes algorithm across multiple thresholds using the EdgeIoT dataset.

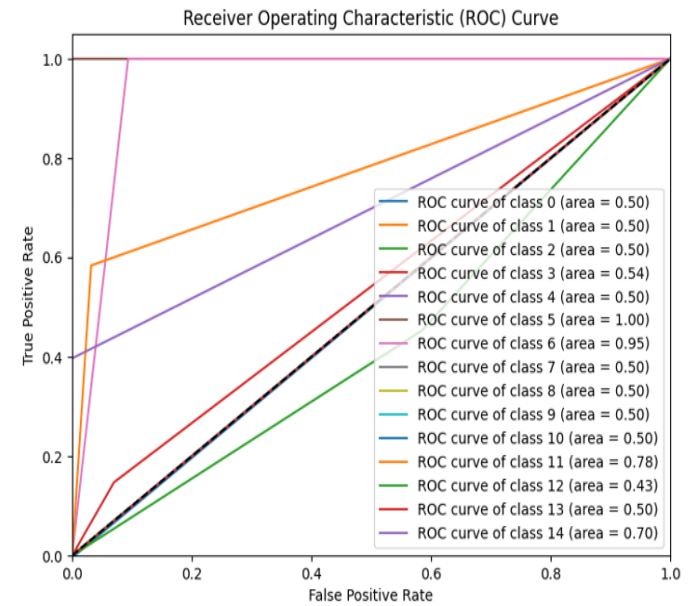


Fig. 9. Receiver Operating Characteristic (ROC) curve illustrating 15 categories using the EdgeIoT dataset.

could not successfully identify instances of these classes. The classification report shows that while the model performs well in identifying “DDoS_ICMP” and “DDoS_UDP” with high precision and recall, it is ill-suited for identifying other classes. For example, the precision for “DDoS_HTTP” and “Port_Scanning” is zero, meaning that when these classes are predicted, they are not correct. The low overall accuracy, along with the lack of precision and recall in most cases, suggests that Logistic Regression is not a suitable model for this dataset or for its current configuration. The model’s performance is highly inconsistent and variable, with some

E. Random Forest

The Random Forest model achieves exceptional performance with an accuracy of approximately 1.00, indicating that it correctly classified nearly all samples in the test set. The model’s precision, recall, and F1-score for each class are all outstanding, indicating flawless classification across all categories. The confusion matrix highlights that the Random

Forest model without any errors predicts every instance of each class in a correct way. Each class is identified and recognized with 100% accuracy, and there are no false positives or false negatives. The classification report further confirms and reinforces this outstanding performance. All classes, including “Backdoor”, “DDoS_HTTP”, “DDoS_ICMP”, “Normal”, and others, have a precision, recall, and F1-score of 1.00. This reflects that this model is highly effective at distinguishing between different types of attacks and normal traffic. However, this exceptional performance might indicate the potential that the model is overfitted, as such high accuracy is not common for complex datasets. It is important to confirm that the model has learned well. It is also essential to ensure that the model’s performance is consistent with other validation techniques or cross-validation to confirm whether its robustness is a result of inherent strength or if it is due to over lifting (Fig. 10 and Fig. 11).

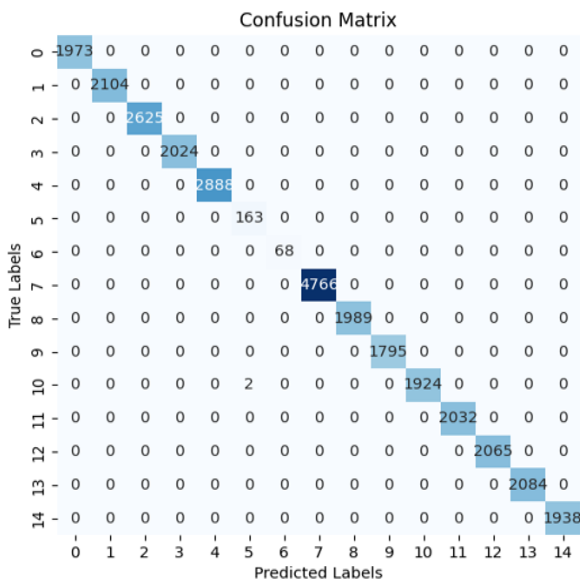


Fig. 10. Confusion matrix of the Random Forest algorithm for classifying 15 categories using the EdgeIoT dataset.

VII. DISCUSSION AND COMPARISON OF MODELS

We performed a full performance comparison of five different machine learning models, Decision Tree, Random Forest, K-Nearest Neighbors (KNN), Naïve Bayes, and Logistic Regression, for the purpose of classification, and we targeted IoT malware detection. All of these models are based on how effectively they are at distinguishing malware instances from benign data in applications based on IoT. Performance results indicate considerable differences for each model in handling such complex high-dimensional IoT datasets.

The results of this research are in agreement with Sliwa et al. [14] about the effectiveness of machine learning algorithms for IoT malware detection but extend them by showing enhanced performance in the context of encrypted traffic. Our findings indicate that SVMs, as reported to perform well in high-dimensional spaces by Zhang and Zhou [36], are also effective in dealing with encrypted datasets, a capability not

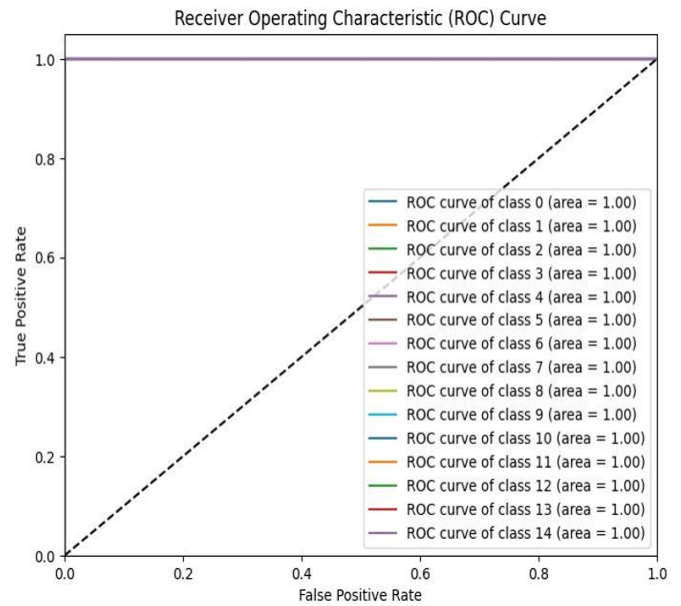


Fig. 11. Receiver Operating Characteristic (ROC) curve illustrating the performance of the Random Forest algorithm across multiple thresholds using the EdgeIoT dataset.

covered to a great extent in previous works. Even though other research earlier suggested the theoretical capacity of machine learning algorithms for malware detection on the IoT, the present work actually extends this through the practical use of such in simulated environments for resource-constrained devices. Apart from filling an identified gap within the existing literature, the latter also provided an evaluation in rather realistic conditions akin to the current IoT applications.

Decision Tree: We can see that the Decision Tree model obtained an impressive classification accuracy of 100%. Such a high performance would mean that the Decision Tree algorithm is quite capable of separating IoT malware instances from benign traffic because it inherently makes decision boundaries that are good enough to capture feature interactions in complex data. **Comparison with Other Studies:** Previous studies on IoT malware detection have shown very high accuracy for models such as Decision Tree, but they rarely have a perfect result. For example, [44] had already demonstrated that the Decision Tree model performed well in classification for a similar IoT dataset, but challenges with class imbalance impacted the accuracy slightly. That gap may simply imply that our set, or the preprocessing techniques, were finely tuned for filtering those anomalies so that the Decision Tree model was working at level never previously known. **Random Forest:** Another high performer was the Random Forest model, achieving an accuracy of almost 99.9%, as well as having higher values of precision, recall, and F1-score. Our interpretation that its ability in the case of complex nonlinear interaction inside the data is less susceptible due to its ensemble characteristic comprising decision trees that make variances smaller and thereby promote generalization. **Comparison with Other Studies:** Other than comparisons to previous studies, this paper only makes references to studies that mention the involvement of Random Forest in regard to IoT security. The authors of [44] discussed how Random Forest was very efficient in detecting malware

for IoT with a precision rate that approached but did not reach 98%. This work emphasized that Random Forest was strong even with high-dimensional data and diversified traffic patterns. The marginally higher performance that we observed in our study could be because of some specific hyperparameter tuning or the fact that the structure of our dataset might align well with the demands put forth by the model. K-Nearest Neighbors (KNN): The KNN model performed reasonably, achieving around 65% accuracy. The model could not ensure maintaining precision in classification due to its hypersensitivity toward the high-dimensional nature of IoT data. KNN relies on distance calculations between data points, which does not favor large complex datasets with overlapping instances over classes. Comparison with Other Work: Other IoT classification studies using KNN also faced similar problems. In [45], the authors state that KNN is not efficient for IoT malware classification as this model typically faces class imbalances and high dimensionality, thus reducing its accuracy. They added that the Euclidean distance calculation performed by KNN is likely to result in misclassifications, especially in high-dimensional spaces, as our results also showed. Naïve Bayes: The results indicated that Naïve Bayes performed very poorly in detecting IoT malware with a precision of 57%. The lower performance points out the weakness of Naïve Bayes in handling datasets that include intricate relationships among features because it assumes independence of features, an assumption usually invalid in real-world IoT scenarios.

Comparison with Other Work: In the same domain, the authors of [45] asserted that Naïve Bayes is an under-optimal algorithm for IoT malware detection because it relies on independence features. This has been one of the most discussed phenomena in the literature since the Naïve Bayes model fails to consider any dependency between the features, which makes it less efficient with such complex interactions. Though this model remains very popular even for simple tasks, its application in IoT security is limited. Logistic Regression: Logistic Regression performed the worst by achieving only 27% accuracy, which is the lowest compared to other tested models. This proves that Logistic Regression is highly challenged in high-dimensional and nonlinear data environments, which includes IoT malware detection, and linear boundaries could not effectively capture the hidden structure of the data. Comparison with Other Studies:

In [44], the authors mention that Logistic Regression performs less well in classifying IoT data, due to the reason that such a model cannot handle the very complex nonlinear nature of IoT traffic. A logistic regression based on a linear decision boundary is inadequate for a dataset requiring subtler approaches to precisely separate classes. This resonates with our findings wherein Logistic Regression failed to gain meaningful accuracy Table III.

TABLE III. COMPARATIVE RESULTS FOR MACHINE LEARNING MODEL PERFORMANCE

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	1.000	1.000	1.000	1.000
KNN	0.654	0.65	0.65	0.65
Naive Bayes	0.568	0.60	0.58	0.51
Logistic Regression	0.273	0.23	0.24	0.21
Random Forest	0.999	1.00	1.00	1.00

A. Accuracy

Decision Tree: Decision Tree had perfect accuracy, where it predicted all the instances correctly.

KNN: KNN showed moderate accuracy, better than Naïve Bayes and Logistic Regression, being significantly behind Decision Tree and Random Forest.

Naïve Bayes: Naïve Bayes performed miserably with accuracy at 0.568, failing to classify the majority of instances. Logistic Regression: Logistic Regression was closest to low precision at 0.273, and it was very bad in terms of accurate instance classification.

Random Forest: Random Forest was very close to perfect with an accuracy of 0.999, a near flawless classification in almost all instances.

B. Precision, Recall, and F1-Score

Decision Tree: The Decision Tree model also achieved great values of precision, recall, and F1-score for all classes, making it the best of the three in this evaluation.

KNN: KNN accuracy was at 0.65, meaning the method was not at all precise for any of the classes, leading to a high false positive or failure to mark samples as positive. The recall and F1-scores were different between classes.

Naïve Bayes: Precision and recall were very low, especially in classes like “Fingerprinting”, “Password”, and “Uploading” where precision and recall were next to zero. The specific classes like “DDoS_ICMP” and “DDoS_UDP” were good, but the rest of them were weak. Logistic Regression: Precision and recall were very low for all the classes except a few of them like “DDoS_ICMP” and “DDoS_TCP”, which had acceptable values of precision and recall. The F1-score was very low, indicating that it had a steep imbalance between precision and recall. Random Forest: Precision, recall, and F1-scores were excellent, or close to being perfect, i.e. 1.00 for every class, which translates to near-perfect classification. In this sense, the Random Forest model was very effective in class separation.

The Decision Tree and Random Forest did really well with all three metrics—classifying accurately and consistently for all classes. The KNN was adequate but had a problem with precision. Naïve Bayes and Logistic Regression seemed fairly weak.

C. Confusion Matrix Insights

Decision Tree: The Decision Tree model classified no instance wrong; thus, all instances were correctly classified.

KNN: Misclassification existed, but in general, most classes were dealt with better by KNN than Naïve Bayes and Logistic Regression.

Naïve Bayes: The confusion matrix revealed an immense amount of misclassification concerning less frequent classes or classes having lesser instances.

Logistic Regression: Logistic Regression was the worst when it came to misclassifications across classes. Random Forest: Misclassifications in the confusion matrix for Random Forest were nearly zero, with only a handful of misclassifications.

D. Deciding on the Best Model

Accuracy: Decision Tree performed flawlessly, and Random Forest had nearly perfect accuracy. Logistic Regression was the worst at attaining accuracy. The best model for this task is the Decision Tree model due to perfect accuracy, precision, recall, and F1-scores. It consistently performed well on all classes, and it is the strongest and best model compared to Random Forest, KNN, Naïve Bayes, and Logistic Regression. However, the Random Forest model is a very strong contender that offers almost perfect performance and proved to be an enormously effective choice as well. Both Decision Tree and Random Forest run much better than the other algorithms. Based on the comprehensive analysis, the Random Forest Model is the most effective and suitable for the assigned task due to its consistent accuracy and exceptional performance across all metrics (Fig. 12).

Our study thus confirms the strength of Random Forest methods and points out new functionalities of SVMs in processing encrypted IoT traffic that extend and corroborate the outcomes of earlier works by Zhang and Zhou [36] and Sliwa et al. [14]. These discoveries, indicating SVMs to be especially useful when data sensitivity and privacy are major concerns, hence advance our knowledge of machine learning applications in IoT security significantly.

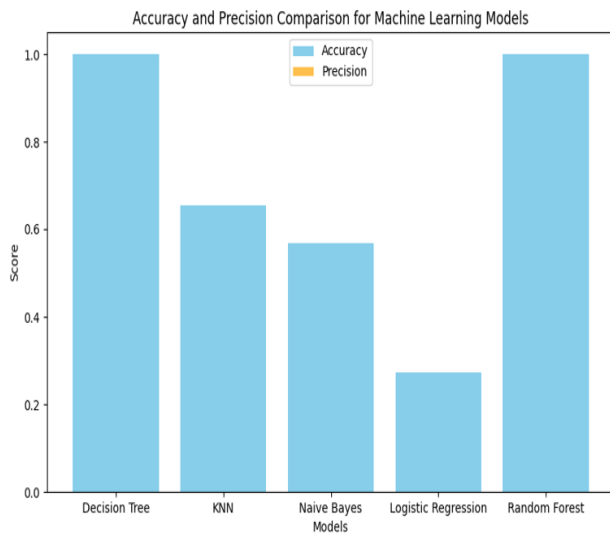


Fig. 12. Accuracy and precision comparison for machine learning model.

Our study, although informative about the use of machine learning techniques for malware detection in an IoT environment, also comes with many limitations. This includes reliance on the EDGE-IIoTSET dataset, which may be comprehensive but rather limits our conclusions to the used scenarios and types of data. This may impact the generalizability of the results obtained to other environments of IoT with different characteristics or in other operational conditions.

The SVM and Random Forest machine learning models are more sensitive to parameters and tuning requirements, and there is no straightforward way of translating the parameters

without modification across different IoT systems, which might make it challenging in real-world deployment where the available computational resources are even more constrained.

Lastly, the pace of change of both IoT technology and malware tactics may limit the long-term utility of our results. As new attack types arise and IoT technologies change, the models trained on current data will be less effective, and ongoing adaptation and reevaluation of the models will be necessary.

VIII. CONCLUSION

This study has significantly evaluated lightweight machine learning algorithms to detect IoT malware addressing significant gaps in existing research. Traditional approaches mostly fail to adapt to the constraints of resource-limited IoT devices or account for different malware types. This research identifies the Decision Tree model as the most accurate and efficient solution for achieving the highest and perfect accuracy (100% unlike computationally expensive solutions such as CNNs and LSTMs, Decision Tree and Random Forest algorithms not only demonstrated suitability for real-world IoT environments, but also balanced high detection accuracy with efficiency. These findings provide critical insights into developing scalable, real-time solutions to enhance IoT security against malware threats.

Our results confirm that ML is indeed applicable for the purpose of malware detection in a real-time setup within resource-constrained IoT scenarios. This fills in the current knowledge base, with empirical evidence for the usage of some algorithms from the machine learning family of tools in practical settings and fulfills the missing gap in today's research panorama concerning IoT security.

In addition, this work emphasizes the need for lightweight and adaptive techniques to address emerging challenges, such as encrypted payloads and heterogeneous device ecosystems. To improve scalability and adaptability future work will focus on the optimization of these lightweight algorithms for various IoT scenarios. Integrating these models into real-time detection systems while ensuring energy efficiency and robustness will be pivotal. Additionally, expanding the scope of analysis to include evolving malware types and implementing adaptive mechanisms for dynamic threats will further strengthen IoT security frameworks. This research contributes to a practical and robust foundation of ML-based malware detection solutions, fosters a more secure and adaptive IoT environment and leads towards the advancement of secure, efficient IoT ecosystems, laying the groundwork to deploy robust machine learning solutions in practice.

ACKNOWLEDGMENTS

This work was supported through the Annual Funding track by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Project No. GRANT KFU250084].

FUNDING: This work was funded by King Faisal University, Saudi Arabia [Project No. GRANT KFU250084].

CONFLICTS OF INTEREST: All authors declare no conflict of interest.

REFERENCES

- [1] Yu, K.; Tan, L.; Shang, X.; Huang, J.; Srivastava, G.; Chatterjee, P. Efficient and Privacy-Preserving Medical Research Support Platform Against COVID-19: A Blockchain-Based Approach. *IEEE Consumer Electronics Magazine* **2020**, *10*, 111–120.
- [2] Yu, K.; Tan, L.; Shang, X.; Huang, J.; Srivastava, G.; Chatterjee, P. Efficient and Privacy-Preserving Medical Research Support Platform Against COVID-19: A Blockchain-Based Approach. *IEEE Consumer Electronics Magazine* **2020**, *10*, 111–120.
- [3] Liu, C.; Xiao, Y.; Javangula, V.; Hu, Q.; Wang, S.; Cheng, X. NormChain: A Blockchain-Based Normalized Autonomous Transaction Settlement System for IoT-Based E-Commerce. *IEEE Internet of Things Journal* **2018**, *6*, 4680–4693.
- [4] Demestichas, K.; Peppes, N.; Alexakis, T. Survey on Security Threats in Agricultural IoT and Smart Farming. *Sensors* **2020**, *20*, 6458. <https://doi.org/10.3390/s20226458>.
- [5] Hassan, R.; Qamar, F.; Hasan, M.K.; Aman, A.H.M.; Ahmed, A.S. Internet of Things and Its Applications: A Comprehensive Survey. *Symmetry* **2020**, *12*, 1674. <https://doi.org/10.3390/sym12101674>.
- [6] Chen, S.; Xu, H.; Liu, D.; Hu, B.; Wang, H. A Vision of IoT: Applications, Challenges, and Opportunities with China Perspective. *IEEE Internet of Things Journal* **2014**, *1*, 349–359.
- [7] Mishra, N.; Pandya, S. Internet of Things Applications, Security Challenges, Attacks, Intrusion Detection, and Future Visions: A Systematic Review. *IEEE Access* **2021**, *9*, 59353–59377.
- [8] Antonakakis, M.; April, T.; Bailey, M.; Bernhard, M.; Bursztein, E.; Cochran, J.; Durumeric, Z.; Halderman, J.A.; Invernizzi, L.; Kallitsis, M.; et al. Understanding the Mirai Botnet. In *Proceedings of the 26th USENIX Security Symposium (USENIX Security 17)*, 2017, pp. 1093–1110.
- [9] De Donno, M.; Dragoni, N.; Giaretta, A.; Spognardi, A. DDoS-Capable IoT Malwares: Comparative Analysis and Mirai Investigation. *Security and Communication Networks* **2018**, *2018*, 1–30. <https://doi.org/10.1155/2018/7178164>.
- [10] Yadav, B.; Tokekar, S. Recent Innovations and Comparison of Deep Learning Techniques in Malware Classification: A Review. *International Journal of Information Security Science* **2021**, *9*, 230–247.
- [11] Regis, W.; Kirubavathi, G.; Sridevi, U.K. Detection of IoT Botnet Using Machine Learning and Deep Learning Techniques. *Preprints* **2023**. <https://doi.org/10.21203/rs.3.rs-2630988/v1>.
- [12] Khan, N.; Awang, A.; Abdul Karim, S.A. Security in Internet of Things: A Review. *IEEE Access* **2022**, *PP*, 1–1. <https://doi.org/10.1109/ACCESS.2022.3209355>.
- [13] Alkasasbeh, M.; Abbadi, M.; Al-Bustanji, A. LightGBM Algorithm for Malware Detection. In *Lecture Notes in Computer Science*; Springer, **2020**. https://doi.org/10.1007/978-3-030-52243-8_28.
- [14] Sliwa, B.; Piatkowski, N.; Wietfeld, C. LIMITS: Lightweight Machine Learning for IoT Systems with Resource Limitations. In *Proceedings of the IEEE International Conference on Communications*; IEEE, **2020**. <https://doi.org/10.1109/ICC40277.2020.9149180>.
- [15] Sliwa, B.; Piatkowski, N.; Wietfeld, C. LIMITS: Lightweight Machine Learning for IoT Systems with Resource Limitations. In *Proceedings of the IEEE International Conference on Communications*; IEEE, **2020**. <https://doi.org/10.1109/ICC40277.2020.9149180>.
- [16] Al-Marghilani, A. Comprehensive Analysis of IoT Malware Evasion Techniques. *Eng. Technol. Appl. Sci. Res.* **2021**, *11*, 7495–7500. <https://doi.org/10.48084/etasr.4296>.
- [17] Felcia, H.J.; Sabeen, S. A Survey on IoT Security: Attacks, Challenges, and Countermeasures. *Webology* **2022**, *19*, 3741–3763. <https://doi.org/10.14704/WEB/V19I1/WEB19246>.
- [18] Ben Henda, N.; Helali, A. Machine Learning for Cyber Security in IoT. *J. Comput. Virol. Hacking Tech.* **2021**, *12*, 1–25.
- [19] Doghramachi, D.; Ameen, S. Internet of Things (IoT) Security Enhancement Using XGBoost Machine Learning Techniques. *Comput. Mater. Continua* **2023**, *77*, 717–732. <https://doi.org/10.32604/cmc.2023.041186>.
- [20] Mehrban, A.; Ahadian, P. Malware Detection in IoT Systems Using Machine Learning Techniques. *Int. J. Wirel. Mob. Netw.* **2023**, *15*. <https://doi.org/10.5121/ijwmn.2023.15602>.
- [21] Mahadevappa, P.; Muzammal, S.M.; Murugesan, R.K. A Comparative Analysis of Machine Learning Algorithms for Intrusion Detection in Edge-Enabled IoT Networks. *arXiv* **2021**. <https://doi.org/10.48550/arXiv.2111.01383>.
- [22] Javed, A.; Awais, M.; Shoaib, M.; Khurshid, K.S.; Othman, M. Machine Learning and Deep Learning Approaches in IoT. *PeerJ Comput. Sci.* **2023**, *9*, e1204. <https://doi.org/10.7717/peerj-cs.1204>.
- [23] Wang, F.; Zhang, M.; Wang, X.; Ma, X.; Liu, J. Deep Learning for Edge Computing Applications: A State-of-the-Art Survey. *IEEE Access* **2020**, *PP*, 1–1. <https://doi.org/10.1109/ACCESS.2020.2982411>.
- [24] Gaurav, A.; Gupta, B.; Panigrahi, P. A Comprehensive Survey on Machine Learning Approaches for Malware Detection in IoT-Based Enterprise Information System. *Enterp. Inf. Syst.* **2022**, *17*, 1–25. <https://doi.org/10.1080/17517575.2021.2023764>.
- [25] Qiu, T.; Chen, N.; Li, K.; Atiquzzaman, M.; Zhao, W. How Can Heterogeneous Internet of Things Build Our Future: A Survey. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 2011–2027. <https://doi.org/10.1109/COMST.2018.2803740>.
- [26] Qiu, T.; Chen, N.; Li, K.; Qiao, D.; Fu, Z. Heterogeneous Ad Hoc Networks: Architectures, Advances and Challenges. *Ad Hoc Netw.* **2017**, *55*, 143–152. <https://doi.org/10.1016/j.adhoc.2016.09.015>.
- [27] Rinaldi, S.; Flammini, A.; Pasetti, M.; Tagliabue, L.; Ciribini, A.; Zanoni, S. Metrological Issues in the Integration of Heterogeneous IoT Devices for Energy Efficiency in Cognitive Buildings. In *Proceedings of the 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Houston, TX, USA, 14–17 May 2018; IEEE: New York, NY, USA, 2018; pp. 1–6. <https://doi.org/10.1109/I2MTC.2018.8409857>.
- [28] Yılmaz, S.; Aydoğan, E.; Sen, S. A Transfer Learning Approach for Securing Resource-Constrained IoT Devices. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 4405–4418. <https://doi.org/10.1109/TIFS.2021.3105883>.
- [29] Al-Turjman, F.; Zahmatkesh, H.; Shahroze, R. An Overview of Security and Privacy in Smart Cities' IoT Communications. *Trans. Emerg. Telecommun. Technol.* **2022**, *33*, e3677. <https://doi.org/10.1002/ett.3677>.
- [30] Ferrag, M.A.; Friha, O.; Hamouda, D.; Maglaras, L.; Janicke, H. Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning. *IEEE Access* **2022**, *10*, 40281–40306. <https://doi.org/10.1109/ACCESS.2022.3165809>.
- [31] Ahmed, M.E.; Kim, H. Machine Learning-Based Malware Detection in IoT Networks Using Packet Metadata. *IEEE Trans. Netw.* **2020**, *28*, 407–418. <https://doi.org/10.1109/TNET.2020.2983097>.
- [32] Kaaniche, N.; Laurent, M. Security and Privacy in IoT: Current Status and Open Issues. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 1686–1721. <https://doi.org/10.1109/COMST.2020.2970499>.
- [33] Chawla, N.; Bowyer, K.; Hall, L.; Kegelmeyer, W. SMOTE: Synthetic Minority Over-Sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. <https://doi.org/10.1613/jair.953>.
- [34] Akhtar, M.; Feng, T. Evaluation of Machine Learning Algorithms for Malware Detection. *Sensors* **2023**, *23*, 946. <https://doi.org/10.3390/s23020946>.
- [35] Zhang, Y.; Zhou, X. Random Forest Algorithm for IoT Security: Benefits and Challenges. *Comput. Secur.* **2021**, *105*, 102367. <https://doi.org/10.1016/j.cose.2021.102367>.
- [36] Hoang, M.; Nguyen, N.; Pham, T.; Nguyen, T.; Dang, T.; Nguyen, H. Evaluating Dimensionality Reduction Methods for the Detection of Industrial IoT Attacks in Edge Computing. *Int. J. Comput. Commun. Control* **2024**, *19*, 10. <https://doi.org/10.15837/ijccc.2024.5.6767>.
- [37] Singh, T.; Di Troia, F.; Visaggio, C.A.; Austin, T.; Stamp, M. Support Vector Machines and Malware Detection. *J. Comput. Virol. Hacking Tech.* **2016**, *12*. <https://doi.org/10.1007/s11416-015-0252-0>.
- [38] Abomhara, M.; Køien, G.; Alghamdi, M. Cyber Security and the Internet of Things: Vulnerabilities, Threats, Intruders, and Attacks. *J. Comput. Syst. Sci.* **2021**, *12*, 1–16.
- [39] Hussain, F.; Hussain, R.; Hassan, S.; Hossain, E. Machine Learning in IoT Security: Current Solutions and Future Challenges. *IEEE Commun. Surv. Tutor.* **2020**, *PP*, 1–10. <https://doi.org/10.1109/COMST.2020.2986444>.

- [40] Choudhary, S.; Kesswani, N.; Majhi, S. An Ensemble Intrusion Detection Model for Internet of Things Network. *Preprints* **2021**. <https://doi.org/10.21203/rs.3.rs-479157/v1>.
- [41] Zhang, Y.; LeCun, Y. Deep Anomaly Detection Using Unsupervised Learning with a Deep Neural Network Autoencoder. *IEEE Access* **2020**, *8*, 19978–19985. <https://doi.org/10.1109/ACCESS.2020.2969855>.
- [42] Al-Garadi, M.A.; Mohamed, A.; Al-Ali, A.K.; Du, X.; Guizani, M. A Survey of Machine and Deep Learning Methods for Internet of Things (IoT) Security. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 1646–1685. <https://doi.org/10.1109/COMST.2020.2977747>.
- [43] Ferrag, M.A.; Friha, O.; Hamouda, D.; Maglaras, L.; Janicke, H. Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning. *IEEE Access* **2022**, *10*, 40281–40306. <https://doi.org/10.1109/ACCESS.2022.3165809>.
- [44] Ferrag, M.A.; Friha, O.; Hamouda, D.; Maglaras, L.; Janicke, H. Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning. *IEEE Access* **2022**, *10*, 40281–40306.
- [45] Samin, O.B.; Algeelani, N.A.A.; Bathich, A.; Adil, G.M.; Qadus, A.; Amin, A. Malicious Agricultural IoT Traffic Detection and Classification: A Comparative Study of ML Classifiers. *J. Adv. Inf. Technol.* **2023**, *14*(4).
- [46] Akhtar, M.S.; Feng, T. Evaluation of Machine Learning Algorithms for Malware Detection. *Sensors* **2023**, *23*(2), 946. <https://doi.org/10.3390/s23020946>.
- [47] W, R.A.; G, K.; Uk, S. Detection of IoT Botnet Using Machine Learning and Deep Learning Techniques. *Research Square* **2023**. <https://doi.org/10.21203/rs.3.rs-2630988/v1>.
- [48] Ferrag, M.A.; Friha, O.; Hamouda, D.; Maglaras, L.; Janicke, H. Edge-IIoTset: A New Comprehensive Realistic Cyber Security Dataset of IoT and IIoT Applications for Centralized and Federated Learning. *IEEE Access* **2022**, *10*, 40281–40306. <https://doi.org/10.1109/access.2022.3165809>.

Building Detection from Satellite Imagery Using Morphological Operations and Contour Analysis over Google Maps Roadmap Outlines

Arbab Sufyan Wadood¹, Ahthasham Sajid², Muhammad Mansoor Alam³, Mazliham MohD Su'ud^{*4},
Arshad Mehmood⁵, Inam Ullah Khan⁶

Department of Computer Science, Baluchistan University of Information Technology, Quetta, Pakistan¹

Department of Information Security and Data Science, Riphah Institute of Systems Engineering,

Riphah International University, Islamabad, Pakistan^{2,5}

Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia^{3,4,6}

Abstract—One such research area is building detection, which has a high influence and potential impact in urban planning, disaster management, and construction development. Classifying buildings using satellite images is a difficult task due to building designs, shapes, and complex backgrounds which lead to occlusion between buildings. The current study introduces a new method for constructing recognition and classification globally based on Google Maps contour trace detection and an evolved image processing technique, seeking synergies with a systematic methodology. We first extract the building outlines by taking the image from the Roadmap view in Google Maps, converting it to gray scale, thresholding it to create binary boundaries, and finally applying morphological operations to facilitate noise removal and gap filling. These binary outlines are overlaid on colorful satellite imagery, which aids in identifying buildings. Machine learning techniques can also be used to improve aspect ratio analysis and improve overall detection accuracy and performance.

Keywords—Building detection; satellite imagery; urban planning; disaster response; image processing; machine learning; morphological operations; contour detection; aspect ratio

I. INTRODUCTION

Building detection in aerial images, a key and well-studied domain has recently drawn considerable attention. High-resolution satellite imagery is increasingly available, and thus necessitates automatic and accurate methods for building detection [1]. Building detection is an important task for urban planning, disaster response [2], change detection [3], construction and development activity monitoring [4].

However, several above-mentioned factors make the detection of buildings from satellite images a challenging and complex task. Buildings vary significantly in their shapes, sizes, and materials. That is, a building in a dense urban location may differ quite markedly from an equivalent suburban one in both form and scale [5].

Furthermore, it may be difficult to differentiate certain structures from their surroundings since they are made of materials that share spectral characteristics with the surrounding area [6], [7]. Other elements seen in metropolitan settings, such as streets, trees, and shadows, can also produce complicated backdrops that make it more difficult to identify buildings. It can also be complex to identify buildings from satellite photos since they can be partially or completely obscured

by other objects, including trees or other structures [8]. The unpredictability of satellite imagery itself is another element that makes construction detection more difficult. Depending on the sensor, atmospheric conditions, and capture time, satellite imagery can differ greatly in terms of resolution, spectral bands, and quality [9].

It is difficult to create a general technique that can reliably identify structures because cities vary throughout time, from small adjustments to total demolition and reconstruction [10].

To overcome the difficulties caused by many elements and increase the precision and resilience of detection algorithms, advanced image processing [11], machine learning, and deep learning approaches can be applied. For the detection of buildings and urban areas from aerial photos, machine learning and deep learning approaches have demonstrated excellent results [2], [12], [13], [14]. Nevertheless, there are a number of restrictions on their use in this situation. First of all, the caliber and volume of training data are critical to these algorithms [15], [16].

The resulting model might not function well on fresh data if the training set is skewed toward particular building or area types or is not representative of real-world data. Large-scale training data collection and labelling can also be costly and time-consuming [17], [18].

Second, machine learning algorithms could find it difficult to generalize to various architectural styles and metropolitan regions [6].

A model trained on photos of contemporary high-rise structures in a crowded city, for instance, would not function as well when used to photos of low-rise, older structures in a rural region. Thirdly, variations in camera settings, illumination, and weather can have an impact on machine learning and deep learning models. Image processing techniques provide a number of benefits over machine learning techniques for more accurate and code-efficient building detection from satellite pictures. First off, compared to machine learning algorithms, image processing methods are less dependent on the caliber or volume of training data. Regardless of data variances, they are based on well-established rules and algorithms designed to detect particular features or patterns in the photographs. This saves time and money by doing away with the requirement for

intensive data collecting and labelling [19]. Second, because image processing methods are not impacted by changes in lighting, weather, or camera settings, they may be used in a variety of metropolitan environments and building styles with little modification. This guarantees dependable and consistent outcomes in various settings. Finally, machine learning models are not as good at detecting buildings that are partially or completely blocked as image processing techniques. Even when building characteristics are completely or partially obscured by other objects, they can still be identified by using sophisticated algorithms like morphological filtering, edge identification, and texture analysis. This lowers the likelihood that the findings will contain false positives or false negatives. Creating trustworthy algorithms for identifying and categorizing buildings from high-resolution satellite data [20] has been the subject of numerous studies. But putting in place a system that can function globally presents other difficulties that need to be resolved, like the fact that different parts of the world have different kinds, sizes, and shapes of buildings, and that handling big datasets with different quality and resolution levels is necessary. A unique Roadmap-to-Satellite Building Detector (RSBD) method is put forth to overcome these obstacles. It makes use of outlines from Google Maps as well as other cutting-edge image processing techniques to create a highly effective and scalable system for building detection and classification on a worldwide basis. The Roadmap view image from Google Maps [21], which includes building outlines [22], is transformed to grayscale for this study. The final image is next subjected to a threshold, the value of which is established by the desired degree of building outline colour.

The threshold image is then enhanced and minor details are eliminated using morphological processes like dilation and erosion [23].

The contours in the threshold image are then determined by features like area or aspect ratio, and those that are not building outlines are filtered out. The identified buildings are then displayed once the filtered outlines have been put on a Google Maps satellite view image [24].

This research aims to solve issues like quality and resolution, as well as the variations in building kinds, sizes, and shapes across the globe, intended for global application. In this work, six distinct global locations with diverse building kinds, sizes, shapes, and picture resolution were used to test the Roadmap-to-Satellite Building Detector (RSBD). The experimental results and quantitative validation in this research indicate the promising potential of the developed approach.

The rest of this paper is organized as follows: Section II is the Literature Review where we discuss the related work. Section III explains the Methodology used by the authors. Section IV provides a Use Case Analysis that illustrates the usefulness of our work. Section V presents the Results of our experiments and evaluations. Section VI contains a detailed Discussion that interprets the results and their implications. Section VII discusses the Threshold Value Analysis, which sheds light on the essential counts that directly affect the output of our analysis. Finally Section VIII concludes the paper.

The major contributions of this article can be summarized as follows:

- Utilizes gray-colored building outlines found in

Google Maps' "Roadmap" map type as a foundational element for building detection from satellite images.

- Introduces an adaptive thresholding technique to convert the grayscale Google Maps "Roadmap" view image into a binary representation.
- Morphological operations are applied to enhance the thresholded image, and contour filtering [25] is employed to remove non-building contours based on specific properties, such as area and aspect ratio, respectively.
- Presents a globally applicable methodology designed to address challenges related to variations in building types, sizes, shapes, image quality, and resolution across different regions worldwide, making it adaptable to diverse contexts.
- Validates the Roadmap-to-Satellite Building Detector (RSBD) through extensive testing on six distinct regions worldwide, encompassing diverse building characteristics and image resolutions. The experimental results and quantitative validation demonstrate the method's promise and potential for efficient and effective building detection.
- A comprehensive critical analysis of the existing work related to building detection in aerial images is presented providing insights into the strengths and weaknesses of the approaches.

II. RELATED WORK

Numerous studies have been conducted on urban areas and building detection from aerial images using advanced image processing [26], [27], [28], [29] and machine learning techniques [2], [12], [13], [14]. In recent years, with the increasing availability of high-resolution satellite imagery, research on urban areas and building detection from aerial images using image processing techniques has been extensively explored due to its importance in various fields, including urban planning, disaster response, and monitoring of construction and development activities. Zerubia et al. presented one of the first studies in this area [26]. They developed a texture parameter that takes into account the image's local conditional variations by modelling the luminance field using chain-based models.

To provide additional information on the likelihood that pixels would belong to a certain cluster, they created a modified fuzzy C-means method with an entropy term that does not require prior knowledge of the number of classes. This approach was tested on both simulated and real satellite images from CNES and ESA and was further applied to a Markovian segmentation model. Benediktsson et al. [27] suggested employing morphological and neural techniques to classify panchromatic high-resolution data from metropolitan regions. Three steps make up the method: feature extraction or selection, classification, and the creation of a differential morphological profile employing geodesic opening and closing operations. High-resolution Indian Remote Sensing 1C (IRS-1C) and IKONOS remote sensing data were used to test the suggested approach, which demonstrates better classification accuracy with comparatively few characteristics required. A

technique for detecting buildings from low-contrast satellite pictures was presented by Aamir et al. [28].

The suggested technique uses a line-segment detection system to precisely identify building line segments and uses singular value decomposition based on the discrete wavelet transform to improve image contrast. The entire building's contours are then obtained by hierarchically grouping the identified line segments. The suggested technique performs better than current methods when applied to high-resolution images with sufficient contrast. In order to extract building rooftops from satellite pictures, Avudaiammal et al. [29] introduced MBION-SVM, a system that combines morphological, spectral, form, and geometrical features with an SVM classifier. The technique employs the Normalized Difference Vegetation Index (NDVI) and Otsu thresholding to remove mislabeled rooftops and the Morphological Building Index (MBI) to identify likely buildings.

An SVM is trained using geometrical features of recognized rooftops, and self-correction is utilized to eliminate rooftops that have been incorrectly categorized and provide surface area data. Kohli et al. [30] used object-oriented image analysis and expert knowledge to present a built environment morphology-based urban slum detection approach. For slum detection, the technique employed spatial measurements and the contrast of textural features. Compared to the land cover classification accuracy of 80.8%, the agreement percentage between the reference layer and slum classification was only 60%. According to the study's findings, the approach is practical and might be successfully used in related situations.

A novel approach to building extraction from high-resolution satellite data is presented by Liu et al. [7] utilizing the probabilistic Hough transform and multi-scale object-oriented categorization. Building roof extraction and shape reconfiguration are the two stages of the system. Building roofs are extracted using a fuzzy rule decision tree classifier after the multispectral and panchromatic pictures are fused and segmented at various space scales. After determining the building roof's dominant line using the probabilistic Hough transform, the building boundary is fitted using a building squaring algorithm. Experimental results show that the approach can precisely identify and extract rectangular building roofs. A new method for automatically extracting building footprints from HRS pan-sharpened IKONOS multispectral pictures was presented by Gavankar et al. [31]. In order to extract buildings and remove incorrectly categorized urban elements, the method mainly concentrates on optimizing segmentation and shape parameters. Completeness, accuracy, and quality indicators are used to assess the technique's suitability. Automatic building detection from pan-sharpened very high spatial resolution satellite data was the main focus of Dey et al. [32].

In multi-level segmentation-based building detection, the suggested method makes use of shadow context, color tone, size, edge features, structural and geometric features, and prior information. Although the results are encouraging, they require modifications for real-world applications. Additionally, the study demonstrates the effectiveness of the UNB pan sharpening method in applications that make use of spectral and spatial data. A region-based level set segmentation technique was presented by Karantzalos et al. [33] for the automatic identification of artificial items in satellite and aerial photos.

The method measures information within regions according to their statistical description, optimizing the position and shape of an evolving geometric curve. Because of its rapid convergence and complete automation, the technique is appropriate for real-time applications. The algorithm was tested on various aerial and satellite photos. It correctly identified roads, buildings, and other man-made features, demonstrating its efficacy through both qualitative and quantitative evaluation. In order to create normalized Digital Surface Models (nDSM) and differentiate between ground and non-ground points, Cao et al. [34] used point cloud data processing techniques such as noise removal and point reduction. They then created a technique that uses characteristics including flatness, normal direction variance, and nDSM texture to designate structures at an object scale. A graph-cut technique was utilized to fuse and normalize these features. The impact of varying grid sizes on parameter correctness and detail was also investigated. In conclusion, the authors thoroughly examined point cloud data in order to construct labeling and characterization. Farhadi et al. [15] use satellite imagery to extract building footprints (BF) in order to address the difficult challenge of tracking the expansion of urbanization. They suggest a novel unsupervised method dubbed Feature-Based Building Footprint Extraction (F2BFE), which makes use of a Digital Elevation Model (DEM) and Sentinel-1 and 2 satellite photos. The process uses sophisticated thresholding techniques for feature extraction and generates a radar index (NRI) from Sentinel-1 data to extract main building footprints (PBF). Furthermore, spectral indices associated with various land cover categories are extracted from Sentinel-2 photos. In order to create precise and effective ways for identifying buildings in satellite data [35], machine learning approaches have recently gained popularity. Support vector machines (SVMs) are a common machine learning method for object detection. SVMs are binary classifiers that have been effectively used for a number of pattern recognition tasks, such as identifying objects in aerial photos. The suggested approach in a paper by Turker et al. [36] uses SVM classification to identify building patches in the image and sequential processing of edge detection, Hough transformation, and perceptual grouping to extract building boundaries.

The developed method is validated through experiments conducted on pan-sharpened and panchromatic Ikonos imagery, which demonstrate high accuracy in detecting industrial and residential buildings, achieving average detection rates of 93.45% for industrial and 95.34% for residential buildings. Cao et al. [14], addressed the challenge of accurately detecting changes in built-up areas (BAs) for a comprehensive understanding of urban development. They introduced a multi-scale weakly supervised learning approach that utilized image-level labels and high-resolution images. Creating multi-scale Class Activation Maps (CAM) for BA pseudo labels, reducing noise in the pseudo labels, and producing trustworthy pseudo labels for BA change detection were the three main components of the approach. Additionally, they used ZY-3 satellite pictures to create multi-view datasets that covered China's largest cities. This method, which uses multi-scale CAM and temporal correlations for increased accuracy, was beneficial because it was economical and efficient in situations with few labels. One machine learning method that has become more and more prominent in building detection is random forests (RFs).

RFs are an ensemble learning technique based on decision

trees that has been effectively used for a variety of remote sensing tasks. The efficiency of machine learning techniques in mapping Jeddah, Saudi Arabia's informal settlements using very-high resolution imagery and terrain data was investigated in a study by Fallatah et al. [13]. The study used an object-based RF technique to map 14 markers of settlement features. With an overall accuracy of 91%, the object-based RF method was found to be more successful than object-based image analysis. Building detection in satellite images has also made extensive use of artificial neural networks (ANNs) [37], in addition to SVMs and RFs. Large datasets can be used to teach ANNs, which are strong machine learning models, intricate patterns, and correlations. Building traits were automatically extracted from high-resolution Pleiades data using machine learning methods in a work by Idris et al. [38]. Building footprints were extracted using the Artificial Neural Network (ANN) with an accuracy rate of 80.13%, proving its efficacy and excellent computational efficiency. The findings of the study offer an automated method for building extraction that can streamline database and map updates for planning and decision-making.

Building detection in satellite photography has been accomplished through the use of convolutional neural networks (CNNs). One kind of deep learning model that is capable of extracting hierarchical features from huge datasets is CNN. A damaged building detection technique based on CNNs optimized with the Bayesian optimization approach was proposed by Ekici et al. [39]. The effectiveness of the improved CNN model is confirmed by performance evaluation metrics derived from balanced and unbalanced testing datasets, and testing and validation results demonstrate the robustness of the suggested approach. UNet-AP, a unique CNN architecture, was presented by Rastogi et al. [40] for the automatic extraction of building footprints from satellite data. The architecture was evaluated using multispectral satellite images and contrasted with the UNet and SegNet baseline implementations. The findings demonstrate that the suggested architecture consistently improves performance across various urban settlement classes, surpassing both UNet and SegNet.

A new model called SG-EPUNet was introduced by Geo et al. [14] for updating building footprints in bitemporal remote sensing pictures. Change detection, building extraction, and edge preservation are all combined into one framework in this approach. It uses a gated attention module (GAM) to improve building edges and an Edge-preservation building extraction network (EPUNet) for accurate building footprint extraction. By using semi-supervised self-training, SG-EPUNet overcomes the problem of limited post-change labels by updating building footprints using pre-change and post-change picture attributes along with a change saliency map.

The proposed approach leverages deep learning [41] and transfer learning to improve model robustness and generalization, making it suitable for automating building footprint updates in remote sensing imagery. However, the proposed SG-EPUNet show limitations in updating the small newly-built buildings, especially when the image resolution is low.

Zheng et al. [2] addresses the critical issue of rapid and accurate building damage assessment in the aftermath of sudden-onset natural and man-made disasters. the study introduces a novel framework called ChangeOS. In ChangeOS,

a deep object localization network replaces the conventional superpixel segmentation in OBIA to generate precise building objects. These objects are then integrated into a unified semantic change detection network along with a deep damage classification network, facilitating end-to-end building damage assessment. This approach not only ensures semantic consistency but also provides deep object features for more coherent feature representation. Ding et al. [42] introduce the Semi-LCD method to enhance Binary Change Detection (BCD) performance when labeled samples are limited. Semi-LCD combines sample perturbation, consistency regularization, and pseudo-labeling. It comprises a supervised module for labeled data and an unsupervised module for unlabeled data. They also propose a lightweight change detection network, LCD-Net, designed to maintain high performance while reducing model complexity. During training, a combined loss function balances supervised and unsupervised components. In testing, the unsupervised module is not used, and change probabilities are binarized to obtain BCD results.

This approach aims to improve BCD with limited labeled data and address model complexity issues.

Wang et al. [43]proposed a deep learning-based approach to detect structured building rooflines from satellite images. The proposed approach uses CNNs to detect corner and line segment primitives, and a collaborative branch of semantic annotation information to obtain the building segmentation map. Experiments on the SpaceNet dataset show that the proposed approach improves the accuracy of building extraction, and the planar graph representation promotes 3D reconstruction and other subsequent applications.

Mohammadian et al. [44] focus on building detection and change detection using remote sensing images, the authors propose a novel siamese model called SiamixFormer. This model utilizes both pre- and post-disaster images as inputs and features a hierarchical transformer architecture with two encoders. In SiamixFormer, each stage of both encoders contributes to a temporal transformer for feature fusion. This fusion involves generating a query from pre-disaster images and (key, value) pairs from post-disaster images, considering temporal features for enhanced performance.

The use of temporal transformers in feature fusion allows the model to maintain large receptive fields effectively, outperforming CNN-based approaches. Finally, the output from the temporal transformer is passed through a simple MLP decoder at each stage.

Although machine learning techniques have shown promise in detecting buildings in urban areas from aerial images, they have limitations. These limitations include heavy dependence on the quality and quantity of training data [16], difficulty in generalizing to different types of urban areas and building styles [6], and challenges in detecting partially or fully obstructed buildings [7]. Furthermore, gathering and classifying training data can be costly and time-consuming [17], [18].

On the other hand, image processing methods can get around these restrictions when it comes to detecting buildings in satellite photos. In order to overcome the aforementioned constraints, image processing techniques are employed in this study. A thorough critical evaluation of the corpus of research on building detection in aerial photos is provided in Table I. To

address the problem of detecting buildings in high-resolution satellite data, the research that are part of this investigation use a variety of approaches, such as image processing techniques, deep learning, and machine learning algorithms. For each study, its advantages and limitations are highlighted, providing insights into the strengths and weaknesses of the respective approaches. This critical assessment serves as a valuable reference for researchers, practitioners, and decision-makers in the fields of urban planning, disaster response, and construction monitoring, helping them make informed choices when selecting methodologies for building detection tasks.

III. METHODOLOGY

As shown in Fig. 1, Google Maps building outlines are the graphical representation of buildings on a map. These belong to the “Roadmap” map type of Google Maps which is intended to show the road network and various geographical features like building footprints. These outlines are lines that outline the shape of a footprint for buildings, and they’re typically presented in light gray or beige. Note that the building outlines shown in Google Maps are not precise: Google’s machine learning algorithms identify and extract building outlines from satellite and aerial images, an image processing technique [31]. These techniques are not always foolproof and often misidentify building footprints, mistaking them with shadows, vegetation or other features [45]. Building outlines are helpful for getting a high-level sense of the general area and so navigating in Google Maps, but they are likely not detailed enough to support urban planning, disaster response or construction tracking efforts. But still, those outlines can help kick-start the automated process of building detection using satellite images. In this research, several image processing operations are applied to the Roadmap image to extract and clean up the building boundaries. You then get an overlaid image, which you can also lay down on a color satellite image and see the structures. The Roadmap-to-Satellite Building Detector (RSBD) flowchart is in Fig. 2 and the articles below explain each step in detail.

A. Converting Google Maps Images to Grayscale for Simplified Image Processing

Let $I_q(r, c)$ represent the “Roadmap” image where $q \in \{1\}$ and pixel value at row r and column c . The I_q image of the target location is obtained by passing the parameters like coordinates, zoom level, and size to the Google Maps Static API [22]. To simplify the image processing operations and reduce the amount of data that needs to be processed [46], the image I_q is converted to grayscale using the Eq. (1):

$$G_q(r, c) = 0.114 \cdot I(r, c, 0) + 0.587 \cdot I(r, c, 1) + 0.299 \cdot I(r, c, 2) \quad (1)$$

where $I(r, c, 0)$, $I(r, c, 1)$, and $I(r, c, 2)$ represent the values of the respective color channels of each pixel, and $G_q(r, c)$ is the resulting grayscale image. The choice of weights used in Eq. (1) was motivated by the well-established phenomenon that the human eye is more sensitive to green light compared to red or blue light [47]. Therefore, the green channel was given a higher weight in the computation, followed by the red and blue channels.



Figure 1. Google map roadmap view with building outlines.

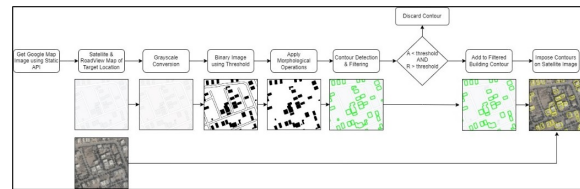


Figure 2. Flow diagram of the Roadmap-to-Satellite Building Detector (RSBD) process.

B. Thresholding Technique for Building Outline Extraction from Grayscale Images

Thresholding is a commonly used technique for converting a grayscale image into a binary image, where each pixel is classified as either foreground or background. Its goal is to make it easier to do additional picture analysis by separating the object of interest—in this case, building outlines—from the background. This study employed a binary thresholding approach, which allocates zero to all pixel values below the threshold and the maximum value to all pixel values above it. The maximum value of 255 and the empirically determined threshold value of 243 in Eq. (2) are based on the features of the building outlines in the grayscale image that was produced from Eq. (1). The following formula is used to apply the thresholding:

$$T(r, c) = \begin{cases} \text{maxval} & \text{if } G_q(r, c) > \text{thresh} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where $G_q(r, c)$ is the intensity value of the grayscale image at pixel (r, c) , thresh is the threshold value 243, maxval is the maximum value 255, and $T(r, c)$ is the resulting threshold image.

TABLE I. CRITICAL ANALYSIS OF EXISTING STUDIES ON BUILDING DETECTION IN AERIAL IMAGES

Year	Author	Method	Advantages	Limitations
2000	Zerubia et al. [26]	Chain-based models, fuzzy C-means algorithm, Markovian model.	<ul style="list-style-type: none"> • Texture parameter for luminance field. • No prior knowledge of classes required. • Tested on real satellite images. 	<ul style="list-style-type: none"> • Limited to texture-based features. • May not generalize well to all urban areas. • Specific to certain satellite images.
2003	Benediktsson et al. [23]	Morphological and neural approaches	<ul style="list-style-type: none"> • Improved classification accuracy. • Few features needed. • Tested on high-resolution data. 	<ul style="list-style-type: none"> • Specific to certain data sources (IRS-1C, IKONOS).
2018	Aamir et al. [28]	Singular value decomposition, line-segment detection	<ul style="list-style-type: none"> • Works with low contrast satellite images. • Accurate building line segment detection. 	<ul style="list-style-type: none"> • Focuses on line segments, not complete building shapes.
2020	Avudaiammal et al. [29]	Morphological Building Index (MBI), SVM classifier	<ul style="list-style-type: none"> • Integrates multiple features. • Eliminates mislabeled rooftops. • Geometrical features used. 	<ul style="list-style-type: none"> • Relies on multiple preprocessing steps. • Requires a labeled dataset for SVM training.
2016	Kohli et al. [30]	Object-oriented image analysis, textural feature contrast, spatial metrics	<ul style="list-style-type: none"> • Suitable for urban slum detection. • Qualitative and quantitative evaluation. 	<ul style="list-style-type: none"> • Lower accuracy compared to land cover classification.
2005	Liu et al. [7]	Multi-scale object-oriented classification, probabilistic Hough transform, building squaring algorithm	<ul style="list-style-type: none"> • Accurate detection and extraction of rectangular building roofs. 	<ul style="list-style-type: none"> • Specific to certain image types. • Multi-scale segmentation may be computationally expensive.
2019	Gavankar et al. [31]	Optimization of segmentation and shape parameters	<ul style="list-style-type: none"> • Focuses on building footprint extraction. • Evaluates completeness and correctness. 	<ul style="list-style-type: none"> • Specific to HRS pansharpened IKONOS images.
2011	Dey et al. [32]	Shadow context, color tone, size, edge features, structural and geometric features, multi-level segmentation	<ul style="list-style-type: none"> • Utilizes various spectral and spatial features. • Shows promising results. 	<ul style="list-style-type: none"> • Requires modifications for real-world applications. • Performance may vary with image quality.
2009	Karantzas et al. [33]	Region-based level set segmentation	<ul style="list-style-type: none"> • Automated and converges quickly. • Detects roads, buildings, and man-made objects. 	<ul style="list-style-type: none"> • Effectiveness may depend on image content and quality.
2020	Cao et al. [34]	Point cloud data processing, feature fusion	<ul style="list-style-type: none"> • Comprehensive analysis of point cloud data. • Addresses building characterization and labeling. 	<ul style="list-style-type: none"> • Sensitivity to parameter settings. • May require careful tuning for different scenarios.
2023	Farhadi et al. [15]	Feature-Based Building Footprint Extraction (F2BFE)	<ul style="list-style-type: none"> • Focuses on monitoring urbanization growth. • Utilizes Sentinel-1 and 2 satellite images. • Automated approach. 	<ul style="list-style-type: none"> • Dependent on Sentinel satellite data availability. • Effectiveness may vary with disaster types.
2015	Turker et al. [36]	SVM classification, edge detection, Hough transformation, perceptual grouping	<ul style="list-style-type: none"> • High accuracy in detecting industrial and residential buildings. • Sequential processing. 	<ul style="list-style-type: none"> • Specific to certain imagery (Ikonos).
2023	Cao et al. [12]	Multi-scale weakly supervised learning, Class Activation Maps (CAM), pseudo labels	<ul style="list-style-type: none"> • Cost-effective approach. • Leverages multi-scale CAM and temporal correlations. 	<ul style="list-style-type: none"> • Effectiveness may depend on label availability and quality. • May require large-scale datasets.
2020	Fallatah et al. [13]	Object-based RF approach	<ul style="list-style-type: none"> • Effective in mapping informal settlements. • High overall accuracy. 	<ul style="list-style-type: none"> • May not generalize well to different regions.
2021	Idris et al. [38]	Artificial Neural Network (ANN)	<ul style="list-style-type: none"> • High accuracy in building footprint extraction. • High computational efficiency. 	<ul style="list-style-type: none"> • Performance may vary with dataset and model complexity.
2021	Ekici et al. [39]	Convolutional Neural Networks (CNNs)	<ul style="list-style-type: none"> • Robust damaged building detection method. • Optimized using Bayesian optimization. 	<ul style="list-style-type: none"> • Effectiveness may depend on dataset and model optimization.
2022	Rastogi et al. [40]	UNet-AP architecture	<ul style="list-style-type: none"> • Improved building footprint extraction. • Outperforms baseline implementations. 	<ul style="list-style-type: none"> • Specific to multispectral satellite imagery.
2021	Geo et al. [14]	SG-EPUNet model	<ul style="list-style-type: none"> • Updates building footprints in bi-temporal remote sensing images. • Incorporates deep learning and transfer learning. • Addresses limited post-change labels. 	<ul style="list-style-type: none"> • May have limitations in updating small newly built buildings with low-resolution images.
2021	Zheng et al. [2]	ChangeOS framework	<ul style="list-style-type: none"> • Precise building object generation. • End-to-end building damage assessment. 	<ul style="list-style-type: none"> • Framework-specific and may require additional labeled data. • Effectiveness may vary with disaster types.
2023	Ding et al. [42]	Semi-LCD method	<ul style="list-style-type: none"> • Enhances Binary Change Detection (BCD) performance with limited labeled samples. • Addresses model complexity. 	<ul style="list-style-type: none"> • Effectiveness may depend on the availability of labeled data. • Complexity tradeoffs.
2021	Wang et al. [43]	CNNs for corner and line segment detection, collaborative branch for semantic annotation	<ul style="list-style-type: none"> • Detects structured building rooflines. • Promotes 3D reconstruction and other applications. 	<ul style="list-style-type: none"> • Specific to structured building rooflines. • Evaluation may vary with different datasets.
2023	Mohammadian et al. [44]	SiamixFormer siamese model	<ul style="list-style-type: none"> • Uses pre- and post-disaster images for building and change detection. • Utilizes hierarchical transformer architecture. 	<ul style="list-style-type: none"> • May require large datasets for optimal performance. • Performance depends on the quality of input images.

C. Morphological Operations for Binary Image Processing: Closing Operation with Structuring Elements

A crucial part of image processing is morphological operations, which are commonly used to work with binary images, where the pixels have binary values of 0 or 1. Because these procedures can change the shape and structure of binary images, they have a wide range of applications, such as object detection, smoothing, and noise removal [48]. These techniques offer ways to enhance image quality, extract significant information, and get images ready for further processing or analysis. The morphological operation carried out in Eq. (3) is closing, which entails applying erosion and dilation procedures one after the other. In order to improve object detection accuracy in later processing stages, the closing procedure is used to fill in tiny gaps in foreground objects [23].

$$M_q(r, c) = (T(r, c) \oplus K) \ominus K \quad (3)$$

The following is the mathematical expression (3) for the closing operation carried out in this investigation. Let K be the structuring element, let $T(r, c)$ be the input binary image, and let \oplus and \ominus stand for dilation and erosion operations, respectively. Image $T(r, c)$ is first dilated using the structuring element K , and then it is eroded using the same structuring element K . Following the operation, the final image is saved as $M_q(r, c)$.

D. Contour Detection for Object Recognition and Segmentation

Applications for contour detection include object recognition, tracking, and segmentation. It is an essential procedure for determining the borders that divide multiple objects or areas inside an image. In order to highlight picture features and make the $M_q(r, c)$ binary image from Eq. (3) suitable for contour detection, it is subjected to morphological processes such as erosion or dilation. As a result, this method may be applied to detect the borders between highways, buildings, and other objects in a picture [49]. In this study, by using Eq. (4), all the contours are retrieved and used to construct a full hierarchy of nested contours. The contour approximation method employed compresses horizontal, vertical, and diagonal segments, leaving only their endpoints. Mathematically, contour detection can be represented as follows:

$$C = \text{findContours}(M_q(r, c), \text{Mode}, \text{Method}) \quad (4)$$

In the mathematical Eq. (4) of contour detection, the binary image $M_q(r, c)$ is subjected to contour detection with the use of two parameters: Mode, which specifies the contour retrieval mode, and Method, which specifies the contour approximation method. The resulting output C is a list of detected contours.

E. Building Contour Filtering Based on Area and Aspect Ratio

As mentioned previously, Google's machine learning algorithms analyze satellite and aerial imagery to identify and map the shapes of buildings. However, these building outlines may not always be accurate, as shadows, vegetation, or other features can sometimes be misinterpreted as building outlines.

Commonly used geometric metrics, such as area or length-width ratio, can help remove small, noisy items or elongated objects such as roads [6].

To eliminate object contours in an image that are not classified as buildings, two filtering conditions are applied based on their area and aspect ratio. In Eq. (5), first, contours with an area less than 500 pixels are considered too small to be a building and are discarded. Second, contours with an aspect ratio of the bounding rectangle less than 0.5 are considered too narrow to be a building and are also discarded. The values of 500 for the area and 0.5 for the aspect ratio were chosen empirically based on the image resolution and the desired level of accuracy for detecting building outlines. These filtering conditions exclude contours that are unlikely to represent buildings, thus improving the accuracy of subsequent processing steps.

$$B = \begin{cases} \text{building} & \text{if area} > 500 \wedge \text{aspect_ratio} > 0.5 \\ \neg\text{building} & \text{otherwise} \end{cases} \quad (5)$$

Where,

$$\text{area} = 0.5 \times |(x_1y_2 - x_2y_1) + \dots + (x_ny_1 - x_1y_n)| \quad (6)$$

and,

$$\text{aspect_ratio} = \frac{w}{h} \quad (7)$$

The filtered list of building contours is represented by B , which is obtained by applying two conditions based on the area and aspect ratio of the contours. Here, the symbol \neg represents the logical NOT operator, and the caret symbol \wedge represents the logical AND operator. The resulting list B contains only the contours that satisfy both conditions and are identified as buildings. Eq. (6) calculates the area of a contour, where n is the number of points in the contour and (x_i, y_i) are the coordinates of the i th point in the contour. The vertical bars $|\dots|$ indicate the absolute value of the sum of the terms inside. The aspect ratio of the contour is then calculated in Eq. (7) as the ratio of the width (w) to the height (h), normalized by converting the w value to a floating-point number and dividing it by h .

F. Buildings Detection and Visualization of Identified Buildings on Satellite Images

Finally, the filtered contour list is superimposed on the satellite image of the target location, providing a visual representation of the identified buildings within the image. Building outlines from Google Maps are used as a baseline by the Roadmap-to-Satellite Building Detector (RSBD), which offers an effective method of detecting buildings from satellite photos. This approach may find use in construction monitoring, disaster response, and urban planning.

IV. TEST CASES ANALYSIS

The trials carried out to assess the effectiveness of the Roadmap-to-Satellite Building Detector (RSBD) methodology are detailed in this section. In order to evaluate RSBD's robustness and generalizability in identifying distinct building kinds in difficult situations, the study tests it on a varied collection of Google Maps photos from different parts of the world in Section VI(A). Furthermore, a quantitative comparison of the detection findings with ground truth data is provided in Section IV(B). Metrics like True Positives had to be calculated for this analysis. False Negatives, False Positives, Completeness, Correctness, and Quality to measure the accuracy and effectiveness of Roadmap-to-Satellite Building Detector (RSBD). Furthermore, Section VII explains our rationale for using a specific threshold value of 243 for thresholding grayscale images consistently throughout our experiments.

A. Qualitative Analysis

33 Google Map [20] photos taken from various parts of the world, including Pakistan, Canada, the United Arab Emirates, India, Yemen, and Thailand, were used to test the Roadmap-to-Satellite Building Detector (RSBD). These regions presented significant challenges due to variations in building types, number of buildings, materials, and occlusions by objects such as trees and shadows. Out of these 33 images, six were acquired from different sites in Pakistan, six from Canada, five from UAE, six from India, five from Yemen, and five from Thailand. The objective of testing the methodology on different regions of the world was to evaluate its robustness and generalizability to various urban areas with varying characteristics. Some building detection results can be found in the following paragraphs.

B. RSBD Performance in Identifying Small Buildings in Sub-Urban Areas: Test Case in Quetta, Pakistan

In this study, we took a series of actions through the roadmap-to-Satellite Building Detector (RSBD) process and reported the detailed results in Fig. 3 based on the test case concerning satellite low-rise building extraction in residential areas. The selected image, shown in the Fig. 3, is a view of an area, which is a suburb of Quetta, Pakistan, was taken using Google maps, [50](lat:30.2668639, and long: 66.9495658). The RSBD was tested at identifying small buildings, typically residential buildings that tend to be low and have small footprints with this case. The RSBD procedure consists of processes such as contour detection, filtering, and classification. First, detected contours on the roadmap view are filtered with two conditions, namely area and aspect ratio. This is done by establishing criteria to determine if the contour smatch what is regarded as a typical building in terms of shape and size. Contours that are too small or have aspect ratios that do not correspond to regular building sizes, for example, are eliminated. This step is important as it helps to reduce the occurrence of false positives wherein some other features which are not buildings, or other such life is wrongly detected as buildings. This filtering is shown on the output of Fig. 3(d) where some contours are filtered out based on less than meets the conditions set. By aggregating data from various sources through a rigorous selection process, the accuracy of the building detection mechanism drastically

improves, as only authentic buildings get recognized. Also, the result presented in Fig. 3(f) confirms the RSBD's ability to correctly pinpoint and delineate small structures. This ensures accuracy for urban planning and development in suburban areas where identification of the spatial distribution of residential structures is critically important. Of specific interest for application development, building detection can be highly beneficial for housing development, infrastructure deployment, resource allocation and disaster management strategies. The RSBD process plays an important role in enabling evidence-based urban & suburban development decisions by effectively mapping and monitoring these structures.

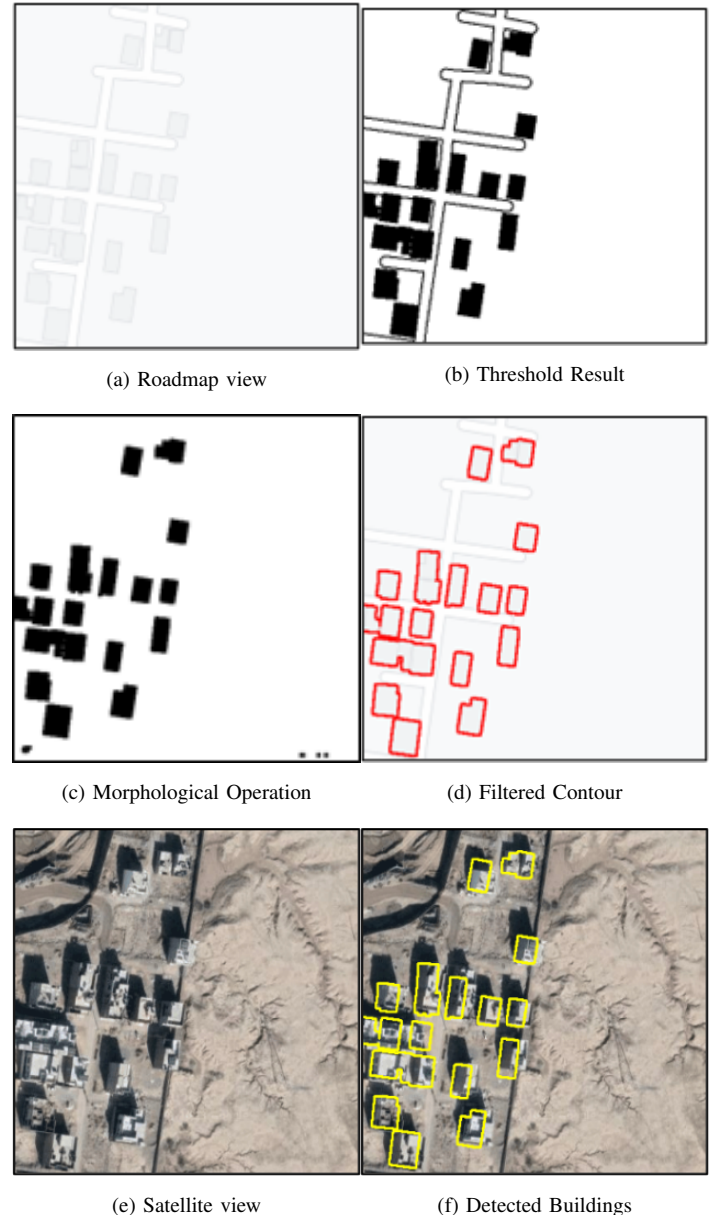


Figure 3. Roadmap-to-Satellite Building Detector (RSBD) successfully identifies small residential buildings in sub-urban areas: A test case in Quetta, Pakistan.

C. RSBD Performance in Identifying High-Rise Buildings: Test Case in Denver, Canada

Fig. 4 [51] is an example of a test scenario with satellite image of urban region with tall buildings. A satellite image from Google Maps [44] of the commercial buildings in Denver Canada along the heights at latitude 39.7491684 and longitude -104.980819. The specific test case was devised to test the strength of the Roadmap-to-Satellite Building Detector (RSBD) to detect several enormous high-rise buildings (tall commercial structures with large footprints) in the scene [5]. These structures are typically located in business districts or as part of a downtown area, making building detection particularly challenging due to their scale and architectural complexity. Therefore, the manual collection of such data is both rich in time and labor cost, which comprise limitations to collection of data and task automation, allowing the RSBDs ability for such high-rise buildings detect, classification as prerequisite for requirement on order chronicles such, namely, environmental, disaster, urban planning applications. With the rapid growth of urbanization, the accurate recognition and monitoring of high-rise buildings became crucial for sustainable city management. They hold significance as social and economic constructs in cities across the world. Their existence impacts the skyline and cityscape, infrastructure demands, emergency services and more. As shown in Fig. 4(f), the RSBD can accurately identify and delineate these structures, indicating its potential to advance in these areas. The successful outcome demonstrates the ability of the RSBD to accurately delineate large high-rise buildings with meaningful implications for urban development and management. Such literacy contributes to the sustainable development of cities by promoting more efficient infrastructure investment, urban planning and emergency responses. The RSBD provides valuable information about the stuff of high-rise buildings, records data on their location and dimensions, and allows urban planning decision-making to be better informed, leading to more efficient resource allocation and improved resilience to natural or human-made disasters.

D. RSBD Performance in Identifying Individual Structures: Test Case in Dubai, UAE

The performance of the Roadmap-to-Satellite Building Detector (RSBD) in identifying a single structure was evaluated using a test image urban areas with a single structure (see Fig. 5). Moving to the next step, we extracted the geographical coordinates of the building: a building in Dubai, United Arab Emirates with latitude (25.0980968) and longitude (55.2373434) [52]. This case was used to test the RSBDs ability to highlight on only one building from an image in an urban filled setting. The results show that RSBD was able to locate and delineate the only building in the image, suggesting it is effective on such images. This is useful in numerous use cases like disaster response, infrastructure assessment, urban planning, etc. In crowded urban centers such as Dubai, it is important to properly identify and track individual buildings. The RSBD supports these efforts through mapping and monitoring isolated buildings with a high degree of precision. Accurate identification is vital for work that includes the assessment of the state of individual buildings, urban planning optimization and effective emergency response in big cities. Focusing on individual buildings can improve the

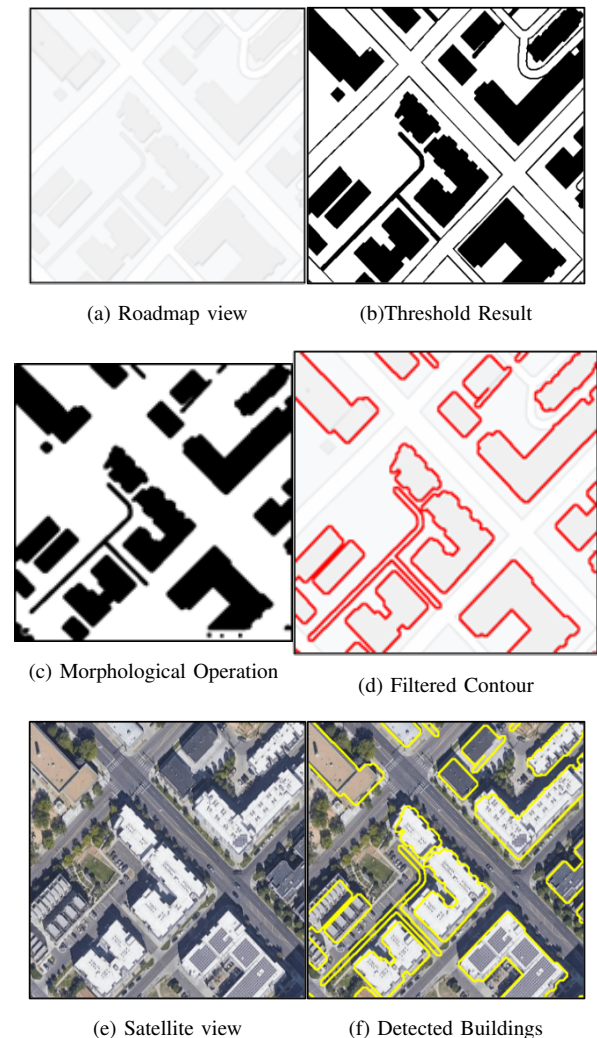


Figure 4. Roadmap-to-Satellite Building Detector (RSBD) successfully identifies high-rise commercial buildings: A test case in Denver, Canada.

accuracy and effectiveness of urban management strategies, from assessing structural integrity after a natural disaster to planning new infrastructure projects. This ability of the RSBD to compute such analyses positions it as a crucial tool for urban planners, emergency responders, and infrastructure modelers alike, providing them with valuable insights upon which they can rely confidently.

E. RSBD Performance in Detecting Multiple Buildings: Test Case in Mumbai, India

Satellite view from Google Maps [53] in Fig. 6 showing an urban area in Mumbai, India, latitude:19.088443, longitude: 72.9033463. This test case tested the capability of our Roadmap-to-Satellite Building Detector (RSBD) to identify multiple buildings that are closely clustered in a single image. Development of the test area covered urban and suburban buildings of varying height, shape, and type representative of the Mumbai skyline. The outcomes illustrated in Fig. 6(f) confirm the RSBD's ability to correctly label and segment multiple structural elements of the image. This capability is

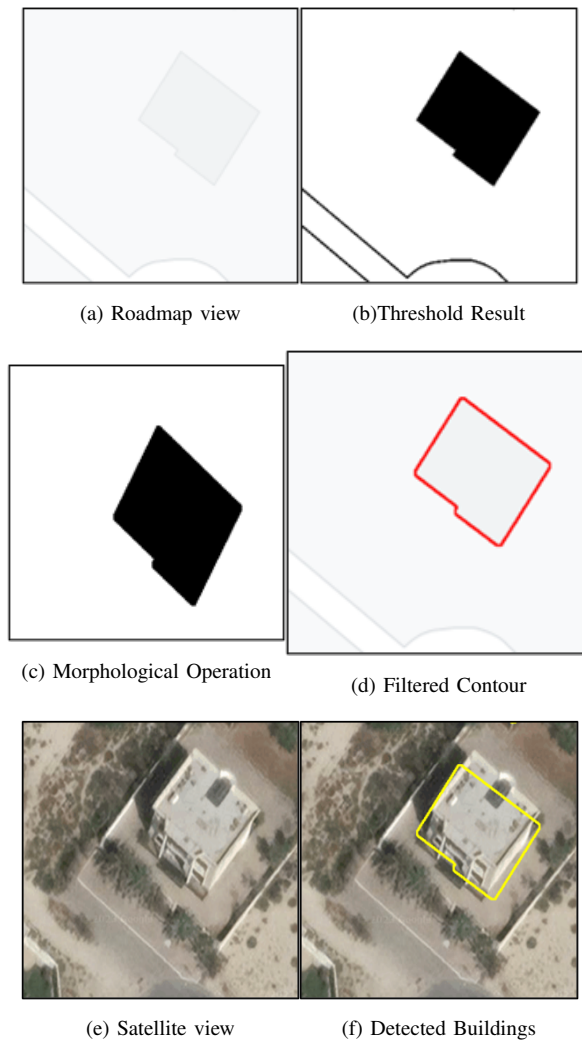


Figure 5. Roadmap-to-Satellite Building Detector (RSBD) successfully identifies high-rise commercial buildings: A test case in Dubai, UAE.

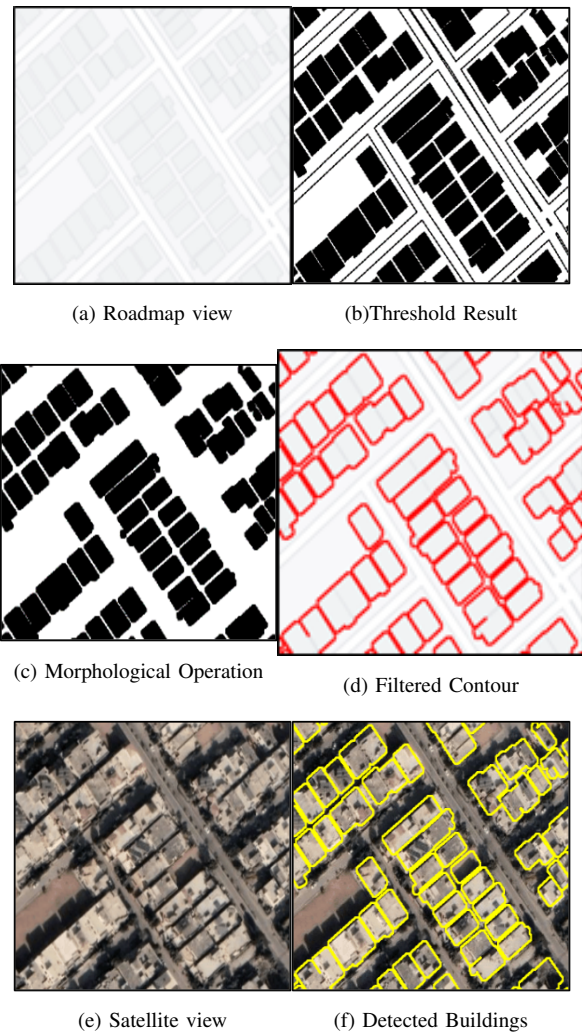


Figure 6. Roadmap-to-Satellite Building Detector (RSBD) successfully identifies high-rise commercial buildings: A test case in Mumbai, India.

especially vital in crowded areas such as Mumbai, where up-to-the-minute information about buildings is critical for all manner of urban management tasks. Precise building detection aids infrastructure development, land-use planning, and disaster management, critical elements for sustainable urban development and resilience. The success of the RSBD at detecting buildings of various sizes and types highlights its versatility and adaptability across urban environments. This functionality is a boon for urban analysts and urban planners worldwide, as it improves the eviction mapping with better accuracy and aids in decision making at various levels. Regardless of the definition, the RSBD's reliability at identifying numerous structures mean that it will be an important tool for urban planners, whether it be for efficiently formulating infrastructure needs in high-density urban areas or keeping track of the suburbs.

F. RSBD Performance in Detecting Earthen Buildings: Test Case in Shibam, Yemen

Houses are built from mud in many other parts of the world which we call earthen houses. This construction material is common in many areas since it is easily available and is comparatively cheap. However, many of these structures have spectral characteristics comparable to their environment, making them difficult to detect using conventional methods. The result of a case of satellite image of Shibam, Yemen with coordinates (15.9223003, 48.6393691) [54] is represented in Fig. 7. Shibam is famous for its mud-brick structures dating back centuries and representing the traditional building style of the area. This test aimed to evaluate the performance of RSBD in the detection and segmentation of buildings in cases where the spectral differences between the buildings and the surrounding terrain are weak. As shown in Fig. 7, it is apparent that the RSBD was able to accurately separate the mud houses from their surroundings, whilst also suppressing the background landscape in the process due to spectral similarity. This finding highlights the strength and versatility of the RSBD to identify

buildings built with natural materials, which are prevalent in rural, and some urban, areas across the globe. Capability of classifying such buildings is important for urban planning, heritage conservation, and disaster management, especially in areas of the world where earthen houses predominate. It aids efforts to keep current records of building inventories and to ensure appropriate measures are taken to decorate architectural heritage and for disaster preparedness. The success of the RSBD in these challenging detection scenarios validates its potential as a versatile tool that can be utilized in several distinct geographical and cultural settings.

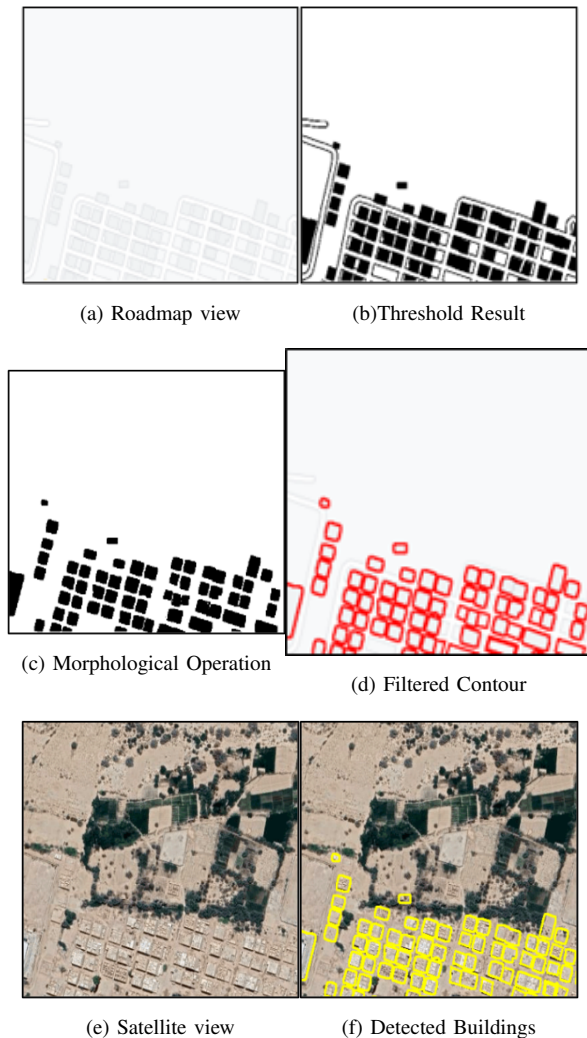


Figure 7. Roadmap-to-Satellite Building Detector (RSBD) successfully identifies high-rise commercial buildings: A test case in Shibam, Yemen.

G. RSBD Performance in Detecting Buildings Obstructed by Objects: Test Case in Thailand

The last test scenario was used to assess the ability of the roadmap-to-Satellite Building Detector (RSBD) S2 to detect buildings that are difficult to identify given the surrounding features of the environment, which may include various obstructions (trees and shadows). For many of these real-world scenario's buildings can be fully or partially hidden from

sight, leading to misclassification when detecting buildings in satellite data because of the Spectro-physical overlap between the elements hiding buildings. Fig. 8 shows the satellite image from Google Maps, is a suburb region in Thailand latitude 19.3287643, longitude 98.3887638 [47] [55]. House have forest,3D Rendering The difficulty of RSBD monitoring with the vegetative coverage of buildings, which may impede image processing conventional methods. Fig. 8(f) shows the overall effectiveness of the RSBD in accurately segmenting anatomy across all patients, even in such challenging scenarios. And even though some buildings were covered by trees, the RSBD has been able to tell the difference and remains an advanced detection tool in cases where natural elements hinder visibility. This feature is vital for applications like urban forestry management, land-use planning, and disaster response, where accurate recognition of concealed buildings is crucial for sound decision-making and resource allocation.

The RSBD has demonstrated strong performance in both obscured and unobscured conditions (79.9% and 73.1%, respectively), reinforcing the ability to reliably detect person-borne threats in different environmental contexts. Such robustness allows for its utilization for many remote sensing applications and urban studies and helps maintain accurate inventories of buildings and preparedness against natural or human-made disasters. Overall, this successful detection of hidden structures is a powerful enhancement of the utility of the RSBD in a wide range of settings, further validating its utility as a general-purpose solution to complex urban detection problems.

H. Quantitative Analysis

A quantitative comparison of the detection results with the ground truth was used to validate the Roadmap-to-Satellite Building Detector (RSBD). The evaluation results of the RSBD approach applied to a set of 33 test photos sourced from Google Map satellite imagery are displayed in Table II. We gathered satellite imagery for every nation, concentrating on certain categories like "Earthen Buildings," "Multiple Buildings," "Individual Building," "High-rising Buildings," "Small Buildings," and "Buildings Obstructed". True Positives (TP), False Negatives (FN), and False Positives (FP) are evaluation metrics that are derived from ground truth data and are essential parts of detection accuracy measurements. After a thorough, careful, and time-consuming process of photo interpretation, an expert manually constructed and annotated the ground truth, which includes the precise locations of the buildings. Furthermore, three quality metrics are presented and computed using the previously specified detection metrics: Completeness, Correctness, and Quality [56]. Specifically, FP stands for the number of buildings that were not found in the image, FN for the structures that were not found, and TP for the number of buildings that were correctly identified. According to Eq. (8), completeness is the number of real structures found in the picture. According to Eq. (9), correctness is a metric that quantifies the proportion of detected buildings that were, in fact, buildings. Completeness and Correctness are combined to create Quality, which is a measure of the algorithm's overall performance as given by Eq. (10). Therefore, one can assess an algorithm's efficacy and accuracy in identifying buildings

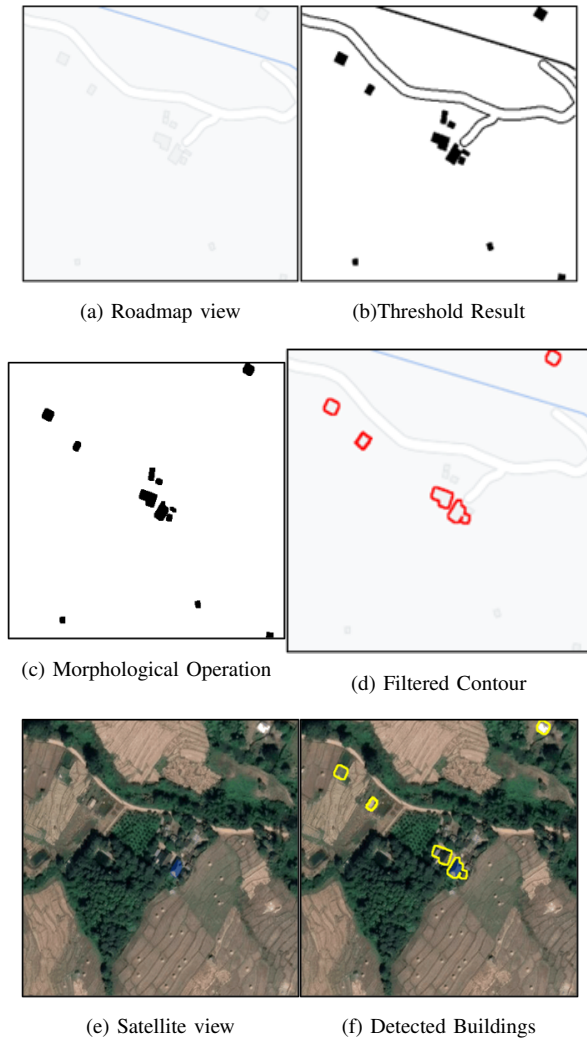


Figure 8. Roadmap-to-Satellite Building Detector (RSBD) successfully identifies high-rise commercial buildings: A test case in Thailand.

in an image by computing these three measures.

$$\text{Completeness} = \frac{TP}{TP + FN} \times 100\% \quad (8)$$

$$\text{Correctness} = \frac{TP}{TP + FP} \times 100\% \quad (9)$$

$$\text{Quality} = \frac{2 \times \text{Completeness} \times \text{Correctness}}{\text{Completeness} + \text{Correctness}} \times 100\% \quad (10)$$

With an average Completeness score of 79%, the Roadmap-to-Satellite Building Detector (RSBD) does a respectable job of identifying buildings in the test photos, according to the data shown in Table II. This suggests that over 80% of the real structures in the pictures can be identified by the RSBD. Furthermore, the majority of the recognized buildings appear to be real buildings, as indicated by the average Correctness score of 9%. The RSBD achieves a reasonable balance between correctness and completeness, as seen by its average Quality score of 85%.

TABLE II. EVALUATION OF THE DETECTION RESULTS IN THE TEST IMAGE SET

Country	Satellite Image	TP	FN	FP	Complete	Correct	Quality
Pakistan*	Small Buildings (1)	17	4	1	81%	94%	87%
	Small Buildings (2)	23	3	4	88%	85%	87%
	High-rising Buildings	6	1	0	86%	100%	92%
	Single Building	1	0	0	100%	100%	100%
	Multiple Buildings	22	4	2	85%	92%	88%
	Earthen Buildings	20	2	0	91%	100%	95%
Canada*	Small Buildings	16	3	0	84%	100%	91%
	High-rising Buildings	6	1	1	86%	86%	86%
	Single Building	1	1	0	50%	100%	67%
	Multiple Buildings (1)	19	5	2	79%	90%	84%
	Multiple Buildings (2)	21	5	0	81%	100%	89%
	Buildings Obstructed	8	2	0	80%	100%	89%
UAE*	Small Buildings	17	9	0	65%	100%	79%
	High-rising Buildings	8	1	0	89%	100%	94%
	Multiple Buildings (1)	20	6	1	77%	95%	85%
	Multiple Buildings (2)	23	3	3	88%	88%	88%
	Earthen Buildings	22	4	0	85%	100%	92%
India*	Small Buildings (1)	18	8	2	69%	90%	78%
	Small Buildings (2)	16	10	1	62%	94%	74%
	High-rising Buildings	4	0	1	100%	80%	89%
	Multiple Buildings (1)	19	7	1	73%	95%	83%
	Multiple Buildings (2)	21	5	1	81%	95%	87%
	Buildings Obstructed	6	3	1	67%	86%	75%
Yemen*	Small Buildings	23	3	2	88%	92%	90%
	Single Building	1	0	0	100%	100%	100%
	Multiple Buildings (1)	22	4	1	85%	96%	90%
	Multiple Buildings (2)	17	9	2	65%	89%	75%
	Earthen Buildings	20	6	4	77%	83%	80%
Thailand*	Small Buildings	18	8	1	69%	95%	80%
	High-rising Buildings	9	2	0	82%	100%	90%
	Multiple Buildings (1)	19	7	3	73%	86%	79%
	Multiple Buildings (2)	21	5	2	81%	91%	86%
	Buildings Obstructed	4	1	1	80%	80%	80%

It is important to note, too, that the RSBD performs differently in various geographical areas. In particular, the RSBD outperforms Yemen and India in terms of construction detection in Pakistan, Canada, the United Arab Emirates, and Thailand. This regional variation in performance suggests that variables like geographic features and differences in building kinds and densities may have an impact on the RSBD accuracy.

V. RESULTS

The performance of the Roadmap-to-Satellite Building Detector (RSBD) is demonstrated in Fig. 9 utilizing six distinct satellite pictures from Pakistan, with an emphasis on the identification of various building types. In terms of quality evaluation and detection accuracy, the data shows encouraging outcomes. Notably, RSBD received a 95% overall quality

score for “Earthen Buildings,” with 91% completeness and 100% accuracy. Likewise, for “Multiple Buildings,” the RSBD revealed an overall quality score of 88%, a completeness of 85%, and an accuracy of 92%. RSBD obtained a perfect completeness and correctness rate of 100% for “Individual Building” detection. Furthermore, RSBD demonstrated excellent completeness scores of 86% and 81% for “High-rising Buildings” and “Small Buildings,” respectively, in addition to high accuracy rates, yielding overall quality ratings of 9% and 87%, respectively. These results highlight how well the Roadmap-to-Satellite Building Detector (RSBD) can recognize a variety of building types across Pakistan’s regions.

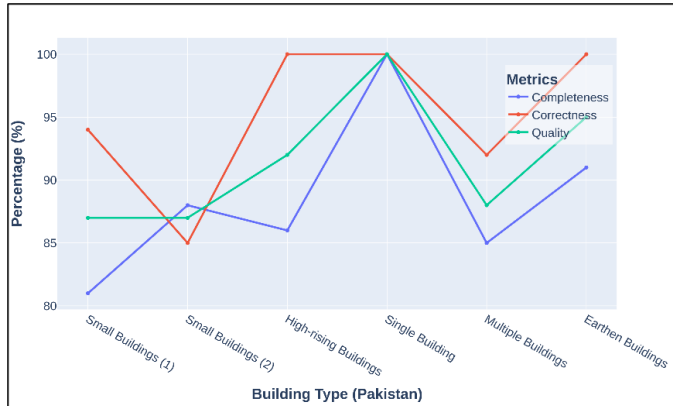


Figure 9. Roadmap-to-Satellite Building Detector (RSBD) Performance Across Different Satellite Images in Pakistan.

The performance of the Roadmap-to-Satellite Building Detector (RSBD) across six distinct satellite pictures in Canada is shown in Fig. 10. Among the many image categories, RSBD demonstrated a remarkable degree of accuracy, with correctness ranging from 86% to 100%. The RSBD technique is strong, as seen by its completeness, which ranges from 50% to 86% and assesses the capacity to discover true positives. The total RSBD quality ranges from 67% to 91%, demonstrating how well RSBD can recognize buildings in satellite imagery from a variety of Canadian locales.

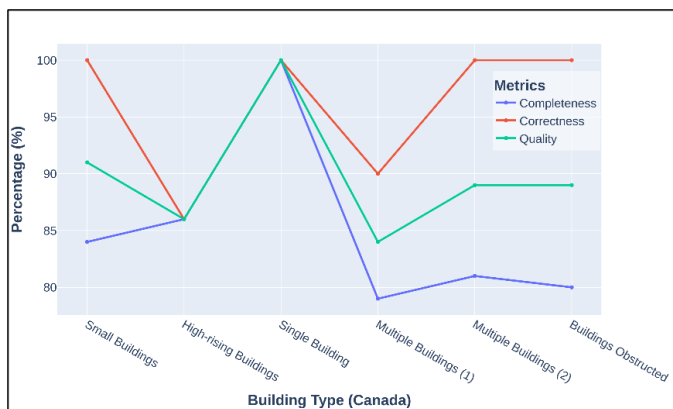


Figure 10. Roadmap-to-Satellite Building Detector (RSBD) Performance across different satellite images in Canada.

Findings from an examination of satellite imagery from different parts of the United Arab Emirates (UAE) are shown

in Fig. 11, with an emphasis on the identification of distinct building types. The Roadmap-to-Satellite Building Detector (RSBD) performance in these categories is shown in the graph, which shows encouraging outcomes. Notably, RSBD received an overall quality score of 85% for the “Multiple Buildings” category, with 77% completeness and 95% accuracy. Likewise, with “Earthen Buildings,” RSBD achieved a remarkable 85% completeness and 100% accuracy, yielding a 92% quality score. For “Small Buildings” and “High-rising Buildings,” respectively, RSBD demonstrated high accuracy rates of 100% and outstanding quality scores of 79% and 94%. These findings underscore the potential of Roadmap-to-Satellite Building Detector (RSBD) in accurately identifying diverse building types in UAE satellite imagery, contributing to advancements in remote sensing applications.

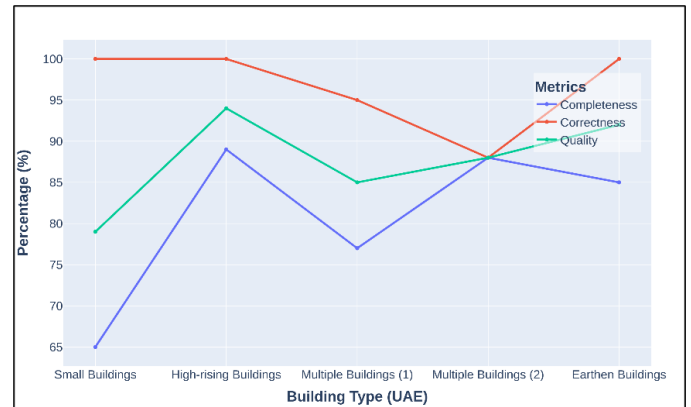


Figure 11. Roadmap-to-Satellite Building Detector (RSBD) Performance across different satellite images in UAE.

The performance of the Roadmap-to-Satellite Building Detector (RSBD) on six distinct satellite photos of India is shown in Fig. 12. According to the graph, when recognizing several buildings, the Roadmap-to-Satellite Building Detector (RSBD) obtained an exceptional average completeness rate of 77% and an accuracy rate of 95%, yielding a quality score of 83%. RSBD obtained a 78% overall quality score, a 66% completeness rate, and a 92% accuracy rate for small buildings. Furthermore, with 100% completeness and 80% correctness rate, RSBD demonstrated exceptional performance in identifying high-rise buildings, earning an 89% quality score. RSBD obtained a quality score of 75%, a correctness rate of 86%, and a completeness rate of 67% when working with obstructed buildings. These results highlight how well Roadmap-to-Satellite Building Detector (RSBD) can recognize and classify buildings in satellite photos, especially when it comes to seeing several, tall buildings. A thorough examination of satellite image data from multiple Yemeni regions is shown in Fig. 13, with an emphasis on the identification of distinct building types. With completeness ranging from 65% to 100% and correctness ranging from 83% to 100%, the graph shows encouraging results in terms of detection accuracy. With an average score of 87%, the overall quality of the buildings that were detected likewise shows excellent performance. With the best performance seen in the recognition of individual buildings, these results demonstrate the promise of the Roadmap-to-Satellite Building Detector (RSBD) for precise building detection in Yemen.

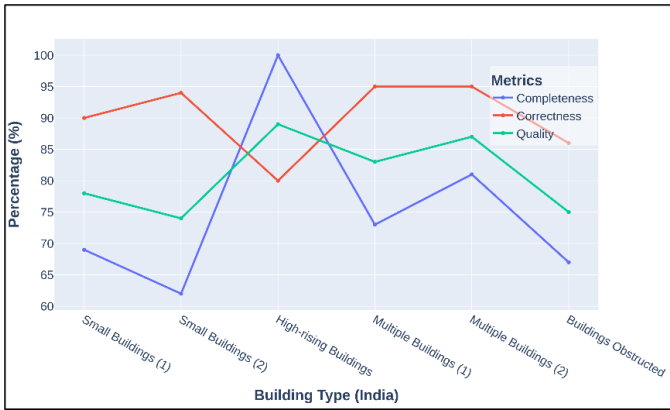


Figure 12. Roadmap-to-Satellite Building Detector (RSBD) Performance across different satellite images in India.

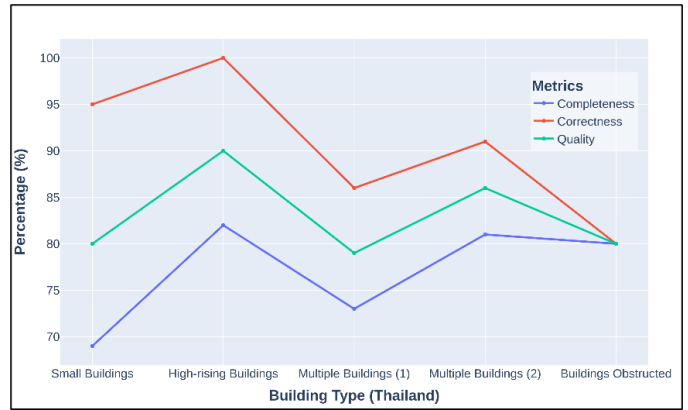


Figure 14. Roadmap-to-Satellite Building Detector (RSBD) Performance across different satellite images in Thailand.



Figure 13. Roadmap-to-Satellite Building Detector (RSBD) Performance across different satellite images in Yemen.

Findings from satellite photos of different parts of Thailand are shown in Fig. 14, with an emphasis on identifying structures and classifying them according to their kind. Significant differences in the performance metrics between the various building categories are shown in the graph. For example, Roadmap-to-Satellite Building Detector (RSBD) received a 79% overall quality score in the “Multiple Buildings” category, with 73% completeness and 86% accuracy. Conversely, the “Small Buildings” category had an overall quality score of 80% due to its higher accuracy of 95% and lower completeness of 69%. These results highlight how crucial it is to modify detection tactics according to particular building types when using satellite data for urban study in Thailand. Moreover, the “High-rising Buildings” category demonstrated exceptional performance with an 82% completeness, 100% correctness, and a remarkable overall quality score of 90%. This suggests that RSBBD excels in detecting taller structures in these satellite images.

VI. DISCUSSION

This section presents and analyzes the findings from the Roadmap-to-Satellite Building Detector (RSBD) approach. In addition to exploring the findings’ wider ramifications, the discussion will offer an interpretation of these results in light of earlier research and working ideas.

A. Robustness and Generalizability

The robustness and generalizability of RSBBD were demonstrated by the qualitative study conducted in several geographical areas. Despite differences in building kinds, sizes, materials, and occlusions, RSBBD was able to detect buildings in a variety of scenarios. The methodology’s flexibility to diverse urban settings is demonstrated by its high performance in several regions. These results are consistent with earlier research that emphasized the significance of creating reliable building detection techniques for satellite photography, considering the variety of urban settings found throughout the world.

B. Detection Accuracy

The quantitative analysis offered a thorough evaluation of the detection accuracy of RSBBD. The performance was assessed using the True Positives (TP), False Negatives (FN), and False Positives (FP) measures. With an average completeness score of 79%, the approach was able to identify roughly 79% of the real structures in the test photos. The bulk of the structures that were spotted were, according to the average accuracy score of 93%, genuine positives. A good balance between completeness and correctness was indicated by the quality score, which averaged 85%. One significant finding is the regional variance in performance, with RSBBD doing better in certain areas than others. Variations in image quality, building density, and geographic elements could all be responsible for this discrepancy. It highlights that in order to achieve the best results, the methodology must be modified to account for certain area features. Additionally, it is in line with earlier studies that have emphasized the difficulties in detecting buildings in various geographical locations.

C. Machine Learning vs. Image Processing

The fact that RSBBD relies on image processing methods rather than machine learning or deep learning algorithms is one of its noteworthy features. Benefits of this option include lower data needs, resilience to changes in weather and lighting, and efficiency when dealing with partially blocked structures. These benefits are consistent with the drawbacks of machine learning models that were covered in the introduction, where issues with data quality, generalization, and environmental

sensitivity were noted. Because machine learning and deep learning techniques work well on particular datasets, they have frequently been preferred in earlier research for constructing detection. Nevertheless, RSBD's findings imply that image processing methods can outperform machine learning models in certain areas while still producing competitive outcomes. This discovery adds to the continuing debate on whether methods are best suited for building detecting jobs.

VII. THRESHOLD VALUE ANALYSIS

As mentioned earlier, thresholding is a commonly used technique to convert grayscale images into binary images by classifying each pixel as foreground or background. This approach is particularly useful in separating the object of interest, which in this study pertains to building outlines, from the background and streamlining subsequent image analysis. In this study, a threshold value of 243 was consistently employed throughout all experiments. This choice was made after a thorough examination of the features of the building outlines in the grayscale pictures. Fig. 15 shows the histogram of pixel intensity values for two grayscale images acquired using Eq. (1) in various test scenarios to further clarify our choice. These graphs demonstrate that 243, 249, and 253 were the intensity values that appeared most frequently in the grayscale photographs. These specific intensity values were found to correlate with ground, roads, and building outlines, respectively, after empirical investigation.

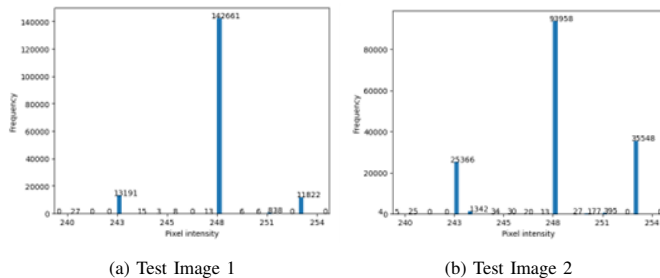


Figure 15. Threshold selection for building outlines in grayscale images: Using pixel intensity histogram analysis.

This study led to the selection of 243 as the threshold value for all studies. This choice was made since it was discovered that this specific intensity value worked best for recognizing building outlines in the grayscale pictures. Additionally, the Roadmap-to-Satellite Building Detector (RSBD) produced great results, showing that this method of detecting buildings from satellite photos has several uses, such as urban planning and catastrophe management. This method allows us to precisely recognize and examine building outlines from satellite photos, yielding insightful information for a range of uses.

VIII. CONCLUSION

The experimental findings show that the Roadmap-to-Satellite Building Detector (RSBD) has the ability to automatically identify and categorize buildings in satellite imagery from Google Maps. The approach successfully recognized and categorized buildings in six global locations, including low-rise

and high-rise, urban and rural, and successfully handled single and multiple structures in an image. To improve the precision and resilience of the detection process, this methodology makes use of sophisticated capabilities, such as the Google Maps Roadmap view, and uses contour filtering and morphological procedures. Furthermore, it is well-suited for universal applications due to its adaptability to different building kinds, sizes, and shapes throughout worldwide areas. Nevertheless, this suggested approach has a drawback. The RSBD method uses Google Maps Road Map view's footprints to identify structures in satellite photos. As a result, RSBD won't recognize buildings whose outlines Google has supplied are out-of-date or missed by Google's algorithm. The significance of regional adaptation is highlighted by the regional differences in RSBD's performance. Future studies might concentrate on adjusting the methodology to particular geographical areas while accounting for elements like construction types, regional materials, and environmental circumstances. This modification may result in improved precision and dependability in many settings. Even though RSBD mostly uses image processing, future studies might look into using machine learning or deep learning methods to improve its functionality even more. To increase detection accuracy, machine learning models could be trained to adjust to local variables. Even greater outcomes could be achieved by combining the advantages of machine learning with image processing. To sum up, the Roadmap-to-Satellite Building Detector (RSBD) presents a viable way to address the difficulties associated with automatically identifying and categorizing buildings in satellite imagery. The methodology's potential for worldwide applications is demonstrated by its resilience and flexibility in a variety of urban settings. Future research and development in the area of automatic building detection and classification from high-resolution satellite data can benefit greatly from the conclusions of this work.

AUTHORS' CONTRIBUTION

Mr. Arbab Sufyan Wadood: Conceptualization, Methodology, Data Collection, Writing – Original Draft, Dr. Ahasham Sajid: Data Analysis, Software Development, Validation, Dr. Muhammad Mansoor Alam: Literature Review, Formal Analysis, Writing – Review & Editing, Dr. Mazliham Mohd Su'ud: Supervision, Project Administration, Funding Acquisition, Mr. Arshad Mehmood: Formatting and camera ready preparation Dr. Inam Ullah Khan: reviews handling.

DATA AVAILABILITY STATEMENT

Data Available Upon Request: "The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request."

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

ABBREVIATIONS

- **RSBD** - Roadmap-to-Satellite Building Detector
- **GIS** - Geographic Information System
- **HRS** - High-Resolution Satellite
- **OBIA** - Object-Based Image Analysis

- NDVI - Normalized Difference Vegetation Index
- DEM - Digital Elevation Model
- CNN - Convolutional Neural Network

REFERENCES

- [1] B. Sirmacek and C. Unsalan, "Urban-area and building detection using sift keypoints and graph theory," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 4, pp. 1156–1167, 2009.
- [2] Z. Zheng, Y. Zhong, J. Wang *et al.*, "Building damage assessment for rapid disaster response with a deep object-based semantic change detection framework: From natural disasters to man-made disasters," *Remote Sensing of Environment*, vol. 265, p. 112636, 2021.
- [3] X. Zhang, P. Xiao, X. Feng *et al.*, "Separate segmentation of multi-temporal high-resolution remote sensing images for object-based change detection in urban area," *Remote Sensing of Environment*, vol. 201, pp. 243–255, 2017.
- [4] P. Saeedi and H. Zwick, "Automatic building detection in aerial and satellite images," in *2008 10th International Conference on Control, Automation, Robotics and Vision*. IEEE, 2008, pp. 623–629.
- [5] M. Wahbi, I. El Bakali, B. Ez-zahouani *et al.*, "A deep learning classification approach using high spatial satellite images for detection of built-up areas in rural zones: Case study of soussmassa region-morocco," *Remote Sensing Applications: Society and Environment*, vol. 29, p. 100898, 2023.
- [6] X. Huang and L. Zhang, "A multidirectional and multiscale morphological index for automatic building extraction from multispectral geoeye-1 imagery," 2011.
- [7] W. Liu and V. Prinet, "Building detection from high-resolution satellite image using probability model," in *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS'05.*, vol. 6. Citeseer, 2005, pp. 3888–3891.
- [8] M. Awrangjeb, M. Ravanbakhsh, and C. S. Fraser, "Automatic detection of residential buildings using lidar data and multispectral imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 5, pp. 457–467, 2010.
- [9] B. Sirmacek and C. Unsalan, "Building detection from aerial images using invariant color features and shadow information," in *2008 23rd International Symposium on Computer and Information Sciences*. IEEE, 2008, pp. 1–5.
- [10] A. Schneider, K. C. Seto, and D. R. Webster, "Urban growth in chengdu, western china: application of remote sensing to assess planning and policy outcomes," *Environment and Planning B: Planning and Design*, vol. 32, no. 3, pp. 323–345, 2005.
- [11] A. Asokan, J. Anitha, M. Ciobanu, A. Gabor, A. Naaji, and D. J. Hemanth, "Image processing techniques for analysis of satellite images for historical maps classification—an overview," *Applied Sciences*, vol. 10, no. 12, 2020. [Online]. Available: <https://www.mdpi.com/2076-3417/10/12/4207>
- [12] Y. Cao, X. Huang, and Q. Weng, "A multi-scale weakly supervised learning method with adaptive online noise correction for high-resolution change detection of built-up areas," *Remote Sensing of Environment*, vol. 297, p. 113779, 2023.
- [13] A. Fallatah, S. Jones, and D. Mitchell, "Object-based random forest classification for informal settlements identification in the middle east: Jeddah a case study," *International Journal of Remote Sensing*, vol. 41, no. 11, pp. 4421–4445, 2020.
- [14] H. Guo, Q. Shi, A. Marinoni *et al.*, "Deep building footprint update network: A semisupervised method for updating existing building footprint from bi-temporal remote sensing images," *Remote Sensing of Environment*, vol. 264, p. 112589, 2021.
- [15] H. Farhadi, H. Ebadi, and A. Kiani, "F2bfe: Development of feature-based building footprint extraction by remote sensing data and gee," *International Journal of Remote Sensing*, vol. 44, no. 19, pp. 5845–5875, 2023.
- [16] R. Qin, J. Tian, and P. Reinartz, "Spatiotemporal inferences for use in building detection using series of very-high-resolution space-borne stereo images," *International Journal of Remote Sensing*, vol. 37, no. 15, pp. 3455–3476, 2016.
- [17] Z. Zhang, W. Guo, M. Li *et al.*, "Gis-supervised building extraction with label noise-adaptive fully convolutional neural network," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 12, pp. 2135–2139, 2020.
- [18] B. Sirmacek and C. Unsalan, "A probabilistic framework to detect buildings in aerial and satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 1, pp. 211–221, 2010.
- [19] N. O'Mahony, S. Campbell, A. Carvalho *et al.*, "Deep learning vs. traditional computer vision," in *Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC)*, vol. 1. Springer, 2020, pp. 128–144.
- [20] T. Partovi, F. Fraundorfer, R. Bahmanyar, H. Huang, and P. Reinartz, "Automatic 3-d building model reconstruction from very high resolution stereo satellite imagery," *Remote Sensing*, vol. 11, no. 14, 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/14/1660>
- [21] Google, "Google maps," 2023, <https://www.google.com/maps>.
- [22] G. Svennerberg, *Beginning Google Maps API 3*. Apress, 2010.
- [23] A. Benhabana, M.-K. Kholadi, R. Bensaci *et al.*, "Building detection in high-resolution remote sensing images by enhancing superpixel segmentation and classification using deep learning approaches," *Buildings*, vol. 13, no. 7, p. 1649, 2023.
- [24] "From google maps to a fine-grained catalog of street trees," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 135, pp. 13–30, 2018.
- [25] C. Zhang, Y. Cui, Z. Zhu, S. Jiang, and W. Jiang, "Building height extraction from gf-7 satellite images based on roof contour constrained stereo matching," 2022. [Online]. Available: <https://www.mdpi.com/2072-4292/14/7/1566>
- [26] A. Lorette, X. Descombes, and J. Zerubia, "Texture analysis through a markovian modelling and fuzzy classification: Application to urban area extraction from satellite images," *International Journal of Computer Vision*, vol. 36, no. 3, pp. 221–236, 2000.
- [27] J. A. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 9, pp. 1940–1949, 2003.
- [28] M. Aamir, Y.-F. Pu, Z. Rahman *et al.*, "A framework for automatic building detection from low-contrast satellite images," *Symmetry*, vol. 11, no. 1, p. 3, 2018.
- [29] R. Avudaiammal, P. Elaveni, S. Selvan *et al.*, "Extraction of buildings in urban area for surface area assessment from satellite imagery based on morphological building index using svm classifier," *Journal of the Indian Society of Remote Sensing*, vol. 48, pp. 1325–1344, 2020.
- [30] D. Kohli, R. Sliuzas, and A. Stein, "Urban slum detection using texture and spatial metrics derived from satellite imagery," *Journal of Spatial Science*, vol. 61, no. 2, pp. 405–426, 2016.
- [31] N. L. Gavankar and S. K. Ghosh, "Object-based building footprint detection from high-resolution multispectral satellite image using k-means clustering algorithm and shape parameters," *Geocarto International*, vol. 34, no. 6, pp. 626–643, 2019.
- [32] V. Dey, Y. Zhang, and M. Zhong, "Building detection from pan-sharpened geoeye-1 satellite imagery using context-based multi-level image segmentation," in *2011 International Symposium on Image and Data Fusion*. IEEE, 2011, pp. 1–4.
- [33] K. Karantzas and D. Argialas, "A region-based level set segmentation for automatic detection of man-made objects from aerial and satellite images," *Photogrammetric Engineering & Remote Sensing*, vol. 75, no. 6, pp. 667–677, 2009.
- [34] S. Cao, Q. Weng, M. Du *et al.*, "Multi-scale three-dimensional detection of urban buildings using aerial lidar data," *GIScience & Remote Sensing*, vol. 57, no. 8, pp. 1125–1143, 2020.
- [35] M. Thaik, L. Bounoua, and H. Cherkaoui Dekkaki, "Using satellite data to characterize land surface processes in morocco," *Remote Sensing*, vol. 15, no. 22, 2023. [Online]. Available: <https://www.mdpi.com/2072-4292/15/22/5389>
- [36] M. Turker and D. Koc-San, "Building extraction from high-resolution optical spaceborne images using the integration of support vector machine (svm) classification, hough transformation, and perceptual grouping," *International Journal of Applied Earth Observation and Geoinformation*, vol. 34, pp. 58–69, 2015.

- [37] A. S. Buriboev, K. Rakhmanov, T. Soqiyev, and A. J. Choi, "Improving fire detection accuracy through enhanced convolutional neural networks and contour techniques," *Sensors*, vol. 24, no. 16, 2024. [Online]. Available: <https://www.mdpi.com/1424-8220/24/16/5184>
- [38] I. Idris, A. Mustapha, O. Caleb *et al.*, "Application of artificial neural network for building feature extraction in abuja," *International Journal of Multidisciplinary Education Research (IJMCER)*, vol. 3, no. 4, pp. 09–15, 2021.
- [39] B. B. Ekici, "Detecting damaged buildings from satellite imagery," *Journal of Applied Remote Sensing*, vol. 15, no. 3, pp. 032 004–032 004, 2021.
- [40] K. Rastogi, P. Bodani, and S. A. Sharma, "Automatic building footprint extraction from very high-resolution imagery using deep learning techniques," *Geocarto International*, vol. 37, no. 5, pp. 1501–1513, 2022.
- [41] M. Pepe, D. Costantino, V. S. Alfio, G. Vozza, and E. Cartellino, "A novel method based on deep learning, gis and geomatics software for building a 3d city model from vhr satellite stereo imagery," *ISPRS International Journal of Geo-Information*, vol. 10, no. 10, 2021. [Online]. Available: <https://www.mdpi.com/2220-9964/10/10/697>
- [42] Q. Ding, Z. Shao, X. Huang *et al.*, "Consistency-guided lightweight network for semi-supervised binary change detection of buildings in remote sensing images," *GIScience & Remote Sensing*, vol. 60, no. 1, p. 2257980, 2023.
- [43] J. Wang, H. Xiong, J. Gong *et al.*, "Structured building extraction from high-resolution satellite images with a hybrid convolutional neural network," in *2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2021, pp. 2417–2420.
- [44] A. Mohammadian and F. Ghaderi, "Siamixerformer: A fully-transformer siamese network with temporal fusion for accurate building detection and change detection in bi-temporal remote sensing images," *International Journal of Remote Sensing*, vol. 44, no. 12, pp. 3660–3678, 2023.
- [45] H. Fan, A. Zipf, Q. Fu *et al.*, "Quality assessment for building footprints data on openstreetmap," *International Journal of Geographical Information Science*, vol. 28, no. 4, pp. 700–719, 2014.
- [46] H. M. Bui, M. Lech, E. Cheng *et al.*, "Using grayscale images for object recognition with convolutional-recursive neural network," in *2016 IEEE Sixth International Conference on Communications and Electronics (ICCE)*. IEEE, 2016, pp. 321–325.
- [47] J. COBB, "Some reflections on color," *Optics and Photonics News*, vol. 6, pp. 51–51, 1995.
- [48] D. Chaudhuri, N. K. Kushwaha, A. Samal *et al.*, "Automatic building detection from high-resolution satellite images based on morphology and internal gray variance," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 9, no. 5, pp. 1767–1779, 2015.
- [49] G. Liasis and S. Stavrou, "Building extraction in satellite images using active contours and colour features," *International Journal of Remote Sensing*, vol. 37, no. 5, pp. 1127–1153, 2016.
- [50] Google, "Quetta, sub-urban area," <https://www.google.com/maps>, 2023.
- [51] —, "Canada, urban area," <https://www.google.com/maps>, 2023.
- [52] —, "Dubai, urban area," <https://www.google.com/maps>, 2023.
- [53] —, "Mumbai, urban area," <https://www.google.com/maps>, 2023.
- [54] —, "Shibam, urban area," <https://www.google.com/maps>, 2023.
- [55] —, "Thailand, sub-urban area," <https://www.google.com/maps>, 2023.
- [56] M. Vakalopoulou, K. Karantzos, N. Komodakis *et al.*, "Building detection in very high resolution multispectral data with deep learning features," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2015, pp. 1873–1876.

Exploring Machine Learning in Malware Analysis: Current Trends and Future Perspectives

Noura Alyemni, Mounir Frikha
College of Computer Sciences & Information Technology,
King Faisal University, Al-Ahsa 31982, Saudi Arabia

Abstract—Sophisticated cyberattacks are an increasing concern for individuals, businesses, and governments alike. Detecting malware remains a significant challenge, particularly due to the limitations of traditional methods in identifying new or unexpected threats. Machine Learning (ML) has emerged as a powerful solution, capable of analyzing large datasets, recognizing complex patterns, and adapting to rapidly changing attack strategies. This paper reviews the latest advancements in machine learning for malware analysis, shedding light on both its strengths and the challenges it faces. Additionally, it explores the current limitations of these approaches and outlines future research directions. Key recommendations include improving data preprocessing techniques to reduce information loss, utilizing distributed computing for greater efficiency, and maintaining balanced, up-to-date datasets to enhance model reliability. These strategies aim to improve the scalability, accuracy, and resilience of ML-driven malware detection systems.

Keywords—Machine learning; malware analysis; cybersecurity

I. INTRODUCTION

Malware evolves and adapts continuously as computer systems and internet connections continue to expand [1]. The interconnected nature of devices allows malware to spread rapidly, resulting in significant cybersecurity risks. In a sense, malware is similar to a digital virus; it is a sneaky program designed to harm your computer or network [2]. This term encompasses a variety of harmful programs, including viruses, worms, trojans, ransomware, adware, and others [3].

As cyberattacks become more sophisticated, there is an increasing need for advanced malware detection and analysis techniques. However, traditional methods face limitations in performance accuracy and often fail to detect unexpected malware variants. In malware analysis, techniques from a variety of fields are used, including program analysis and network analysis [4]. By examining malicious samples, analysts aim to gain a comprehensive understanding of malware behavior and how it evolves over time.

Researchers have developed various methods for malware detection, which can be broadly divided into two groups: signature-based techniques and machine learning (ML)-based techniques. Signature-based methods rely on recognizing predefined patterns from known malware, while ML-based approaches use algorithms to analyze both benign and malicious samples [5]. This allows ML models to detect both familiar threats and new unpredictable ones. The adaptability of ML-based techniques makes them more suitable for malware detection.

The application of machine learning in malware detection offers promising solutions by adapting to new and evolving threats. However, while machine learning offers significant potential, existing research often examines individual techniques in isolation, without providing a cohesive view of their combined strengths and weaknesses. Furthermore, practical challenges such as mitigating adversarial attacks, managing computational efficiency, and addressing dataset imbalances in real-world applications remain underexplored. These gaps highlight the need for a more integrated and comprehensive approach to fully realize the potential of machine learning in malware detection. ML-based methods enable systems to learn and improve from experience without requiring explicit programming for each task. Unlike signature-based techniques, which depend on predefined malware signatures, ML-based methods are more effective in identifying emerging threats. The effectiveness of these models depends heavily on the quality of features and training data, making them adaptable to the constantly changing nature of malware.

This paper aims to present a comprehensive overview of current trends in machine learning for malware analysis, including descriptions, challenges, and future directions. Specifically, the research aims to answer these questions:

- 1) What are the key trends in machine learning-based malware analysis techniques?
- 2) What are the challenges and issues associated with each of these trends?
- 3) What future research directions in this field require further exploration?

The paper is organized as follows. Section II provides an overview of machine learning in malware analysis. In Section III, we present the methodology used in our research. Section IV describes different trends in malware analysis using ML, followed by challenges associated with each trend in Section V. Finally, suggestions for future directions, countermeasures, and conclusions are discussed in Sections VI and VII.

II. ROLE OF MACHINE LEARNING IN MALWARE ANALYSIS

Machine learning has become a vital component in malware detection and analysis, offering solutions to the challenges posed by traditional methods. Its ability to identify unique patterns, adapt to emerging threats, and process vast amounts of data has positioned it as a cornerstone technology in the fight against cybercrime.

One of the key strengths of machine learning is its scalability. Unlike traditional malware analysis techniques, which

depend on manual processes that are both time-consuming and error-prone, machine learning algorithms can evaluate millions of files in just seconds [6]. This ability to quickly and efficiently identify potential threats is crucial in a landscape where the volume and complexity of malware are growing exponentially .

Another significant advantage of machine learning is its adaptability. As cybercriminals continuously develop sophisticated malware and zero-day attacks exploiting vulnerabilities that have not yet been identified machine learning models are uniquely equipped to detect hidden anomalies and respond to novel attack patterns [7]. Models that focus on analyzing dynamic behaviors are particularly effective at staying ahead of these evolving threats, making machine learning an indispensable tool in cybersecurity.

Pattern recognition is another area where machine learning excels. By analyzing the code, behavior, and attributes of malware, these models can uncover intricate patterns that would likely go unnoticed by human analysts. This capability is especially important for identifying zero-day malware, which exploits previously unknown vulnerabilities [8]. Moreover, the automation provided by machine learning frees security analysts to focus on higher-level tasks, such as strategic threat intelligence, thereby improving an organization's overall response to cyberattacks.

To explore how machine learning strengthens malware detection, the next section delves into the primary approaches to malware analysis, including static, dynamic, and hybrid techniques.

A. Overview of Malware Analysis Approaches

Understanding how malware operates, what it targets, and the potential damage it can cause is critical for developing effective defenses. Malware analysis helps achieve this by examining the behavior, structure, and impact of malicious programs. Over the years, several methods have been created to analyze and detect malware, each tailored to address evolving threats. This section discusses the primary approaches: static, dynamic, and hybrid analysis.

- **Static Analysis:** Static analysis involves inspecting the structure of a program without running it. This approach identifies key attributes of executable files, such as memory usage and file sections, to understand the malware's properties. It is often divided into basic and advanced techniques. Basic static analysis focuses on simple characteristics like file size, type, and header information, using tools such as PEiD, BinText, MD5deep, and PEview [9]. Advanced static analysis takes a deeper dive into the code itself, analyzing commands and instructions in detail to uncover malware's hidden functionality [10]. Machine learning often utilizes features extracted during static analysis, including opcode sequences, file headers, and structural patterns, to build models capable of identifying malware patterns. These features allow models to distinguish between malicious and legitimate software.
- **Dynamic Analysis:** Dynamic analysis examines the behavior of malware as it executes, often in a controlled environment such as a sandbox or virtual

machine. This method provides insights into how malware operates in real-world scenarios while keeping the host machine protected from infection. Tools like Process Monitor, API Monitor, Process Explorer, Regshot, and Wireshark are commonly used to observe basic malware behaviors [11]. Advanced dynamic analysis goes further by using debugging tools like OllyDbg and WinDbg, allowing analysts to step through code execution, modify parameters, and examine detailed system interactions. Once the analysis is complete, the environment is reset to its original state to ensure safety [12]. Behavioral data collected during dynamic analysis, such as API calls, system interactions, and network traffic patterns, plays a crucial role in training machine learning models. These insights help create algorithms that detect both known and previously unseen malware.

- **Hybrid Analysis:** Hybrid analysis combines static and dynamic techniques to provide a more comprehensive understanding of malware. It begins by analyzing the code and structure without executing it and then proceeds to observe its behavior in a controlled environment. This dual approach overcomes many limitations of using static or dynamic methods alone [13]. Features generated from both static and dynamic analysis, such as opcode sequences, behavioral patterns, and system interaction logs, are integrated into machine learning models. This combination enhances the adaptability and accuracy of malware detection frameworks, making them more effective against evolving threats.

Each of these methods has its strengths and weaknesses, as shown in Table I, and their integration with machine learning offers promising advancements in malware detection [14].

TABLE I. COMPARISON OF MALWARE ANALYSIS APPROACHES

Approaches	Advantages	Disadvantages
Static Analysis	<ul style="list-style-type: none">• Quick analysis• Low resource usage• Multi-path analysis• Enhanced security• High accuracy	<ul style="list-style-type: none">• Difficulty analyzing obfuscated and encrypted malware• Limited ability to detect unknown malware
Dynamic Analysis	<ul style="list-style-type: none">• Analysis of obfuscated and encrypted malware• Superior accuracy compared to static analysis• Detection of known and unknown malware	<ul style="list-style-type: none">• Slow and insecure• High resource usage• Time-consuming and vulnerable• Limited code analysis
Hybrid Analysis	<ul style="list-style-type: none">• Result in more accurate result	<ul style="list-style-type: none">• Requires significant time and resources• High level of complexity

Static, dynamic, and hybrid analysis approaches provide valuable features that significantly enhance machine learning

models used in malware detection. By integrating these methods into machine learning workflows, we can improve detection accuracy and tackle the challenges posed by increasingly sophisticated and evolving malware threats.

III. RESEARCH STRATEGY

This paper uses a systematic literature review (SLR) to explore the role of machine learning in malware detection. The goal is to gather a comprehensive set of relevant studies. The PRISMA 2020 framework (see Fig. 1) was followed to ensure a transparent and systematic approach to selecting and evaluating research articles.

The review focused on journal articles and conference papers published in the past four years. Databases such as IEEE Xplore, ScienceDirect, SpringerLink, and Google Scholar were searched using keywords like “machine learning” AND “malware analysis,” “AI-based malware detection,” and “deep learning” AND “malware classification”. The initial search retrieved 550 records, 500 from primary databases and 50 from secondary sources. After removing 100 duplicates, 450 unique papers remained.

Next, the studies were screened by reviewing their titles and abstracts. This step eliminated 350 papers that were not relevant, lacked full-text availability, were limited to abstracts, or were in languages other than English. The remaining 100 papers were reviewed in full, resulting in the exclusion of 70 papers due to insufficient relevance or methodological quality. Finally, 30 studies were selected based on their alignment with the research scope and their focus on recent challenges or innovative approaches in ML-based malware detection. The selection process is illustrated in Fig. 1.

IV. TRENDS IN MALWARE ANALYSIS USING MACHINE LEARNING

Machine learning has revolutionized malware detection by improving both accuracy and efficiency. Researchers have concentrated on three key approaches: deep learning, transfer learning, and explainable AI (XAI). Each of these techniques brings its own advantages and challenges, working together to tackle the complex demands of malware detection by striking a balance between precision, resource efficiency, and transparency. This section explores these methods, highlighting their applications and contributions to advancing malware detection.

A. Deep Learning-Based Malware Analysis

Deep learning is a sophisticated branch of machine learning that uses deep artificial neural networks to find hidden patterns and intricate correlations in data. These networks are made up of linked layers of cells that hierarchically learn representations directly from raw input data, simplifying the process and eliminating the need for manual feature engineering [15]. This automatic feature extraction allows deep learning models to efficiently manage large volumes of data, making them vital in fields such as image processing, healthcare and cybersecurity. Deep learning creates several levels of abstraction using supervised and unsupervised algorithms, facilitating complex analysis and decision-making. This has led to its broad acceptance in a variety of sectors [16].

Rhode et al. [17] investigated Recurrent Neural Networks (RNNs), especially Long Short-Term Memory (LSTM) networks, for early-stage malware prediction. The authors extracted static features from Portable Executable (PE) files, a common format used by Windows applications, and utilized the LSTM network to simulate the sequential nature of file execution, as a result of which the model was able to capture both short-term and long-term dependencies. By focusing on early-stage behaviors, their model predicted whether a file was malicious before it fully executed, preventing malware before it spreads. Study results demonstrated that LSTM networks are capable of learning temporal patterns, crucial for understanding malware behavior over time. In contrast to traditional machine learning models that rely on static analysis, Rhode et al.’s approach reduced the vulnerability window by detecting malicious intent earlier with LSTMs. In their study, they found that the RNN-based model outperformed traditional techniques such as decision trees and SVMs, which require more data to identify malware. Elayan and Mustafa [18], developed a deep learning-based approach for detecting Android malware. Using gated recurrent units (GRUs), they analyzed static features from Android apps, including API calls and permissions. Their model achieved a high accuracy of 98.2% on the CICAndMal2017 dataset, which supports the effectiveness of deep learning in identifying malicious Android apps.

Catak et al. [19], developed a sequential model using deep learning for analyzing Windows EXE files. Researchers gathered and analyzed a dataset of non-malicious and malicious EXE files and extracted API call sequences. They employed a Long Short-Term Memory (LSTM) network to model the sequential nature of these API calls, which enabled the model

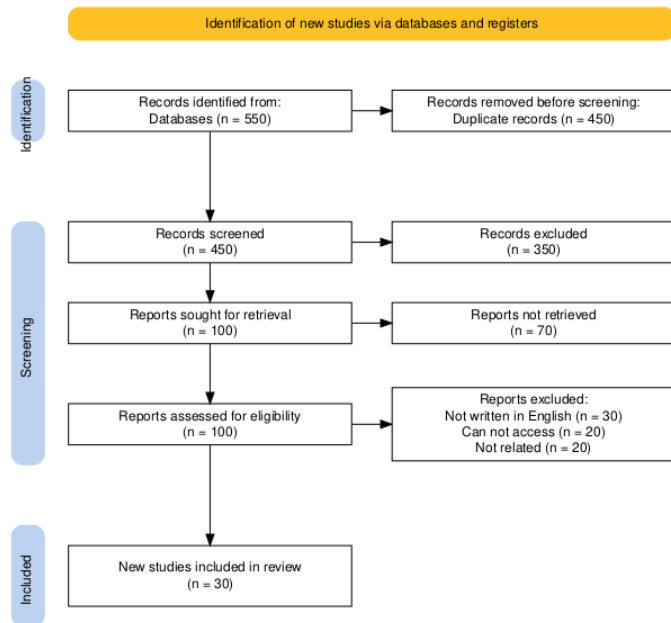


Fig. 1. Selection of papers for review using PRISMA model.

to detect temporal dependencies and patterns indicative of malicious behavior. Based on the dataset, the LSTM model was trained to distinguish malicious from benign EXE files with an impressive accuracy of 98.2%. As a result, deep learning has been demonstrated to be effective in detecting and classifying malware in Windows environments. McDole et al. [20] investigate the application of deep learning techniques for behavioral malware analysis in cloud Infrastructure-as-a-Service (IaaS) environments. The study shows how deep learning models are effective at analyzing malware behavior, making them more effective at detecting sophisticated attacks on cloud-based infrastructures. Similarly, Ravi et al. [21] developed a multi-view attention-based deep learning framework for malware detection in smart healthcare systems. By accurately identifying malicious activities and taking into account the unique operational dynamics of healthcare networks, their work demonstrates that deep learning plays a critical role in ensuring security within healthcare settings.

Calik Bayazit et al. [22] conducted a comprehensive comparative analysis of deep learning models for Android malware detection. They used the Drebin dataset, a publicly accessible collection of benign and malicious Android apps, to evaluate the performance of various models, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). They trained and evaluated the models effectively by extracting static features such as permissions, API calls, and opcode sequences from the apps. In this study, the results demonstrated that hybrid architectures, combining CNNs and RNNs, outperformed individual models, providing evidence that deep learning can enhance Android security.

Ibrahim et al. [23] proposed a malware detection method for Android applications that combines static analysis and deep learning. They extracted key features, including two newly defined features, from the applications. These features were then used as input for a custom-developed deep learning model. The method was evaluated using a classified dataset of Android apps. The extracted features included permissions, API calls, services, broadcast receivers, opcode sequences, application size, and fuzzy hash.

Patil and Deng [24] demonstrated the superior performance of deep learning (DL) networks over traditional machine learning models in malware analysis. They developed a neural network-based framework that achieved high accuracy in classifying malware. The researchers attributed the improved performance to the backpropagation and gradient descent mechanisms employed in DL, which enhance accuracy, true positive rate, and reduce false positive rate.

Rodrigo et al. [25] developed a hybrid machine learning model for Android malware detection. The model consisted of three fully connected neural networks: one for static features, one for dynamic features, and one for a combination of both. When trained on individual features, the static network achieved 92.9% accuracy and the dynamic network achieved 81.1% accuracy. However, the hybrid model, combining both static and dynamic features, outperformed the individual models with an accuracy of 91.1%. This suggests that a hybrid approach, considering both static and dynamic characteristics, is more effective for detecting Android malware.

Obaidat et al. [26] proposed the Jadeite framework to detect

Java-based malware by combining image analysis and behavior analysis with deep learning. The framework uses Java bytecode to create grayscale images that represent malware and identify malicious behavior in real time. Jadeite is composed of three primary components. The first component is the Bytecode Transformation Engine, which converts Java bytecode into grayscale images so malware can be visualized. The second is the Feature Extraction Engine that extracts critical features from bytecode. In addition to the two grayscale images and extracted features, the CNN Classifier Engine analyzes the entire file to determine whether it is malicious or benign using a Convolutional Neural Network (CNN) model.

Although these techniques have offered optimal results in modeling intricate patterns of different malware, these algorithms heavily depend on high-quality datasets and are fairly susceptible to adversarial inputs, as will be explained below. Moreover, based on the papers examined, CNNs emerge as the most often used deep learning technology for malware detection, recognized for their capacity to quickly extract features from binary or grayscale representations. As a result of this capability, CNNs are particularly effective at analyzing image-based malware. Additionally, in order to detect dangerous patterns in code or behavior, recurrent neural networks and LSTMs are frequently used for sequential data analysis. Other approaches, such as deep neural networks, help advance the area of malware detection by identifying intricate linkages in data. While deep learning has revolutionized malware detection, its practical deployment still faces significant challenges, as detailed in the next section.

Although deep learning excels at identifying complex patterns, its application often demands large datasets and significant computational resources, which can limit its practicality. To address these constraints, transfer learning has emerged as a promising alternative by reusing pre-trained models to enhance efficiency.

B. Transfer Learning-Based Malware Analysis

Transfer learning allows knowledge from one domain to be applied in another, minimizing the need for large training datasets and heavy computational demands. This method has garnered considerable interest in malware analysis for its ability to efficiently address challenges related to data scarcity and resource constraints. Researchers have demonstrated its potential in improving malware detection, particularly when datasets are limited—a common challenge in real-time applications [27].

A number of advantages can be derived from the use of transfer learning in malware analysis. First, it greatly minimizes the quantity of training data necessary. Because the model begins with pre-learned information fewer malware-specific samples are required to attain high accuracy. Additionally, Transfer learning improves feature representation by combining abstract and complicated patterns learnt during pre-training. Moreover, the process is computationally efficient which reduces the time and resources needed to train a model from the beginning [28], [29].

Chen et al. [30] demonstrated the effectiveness of transfer learning for static malware classification by adapting pre-trained CNNs to malware-specific datasets. By treating mal-

ware binaries as images, their approach significantly reduced training time while maintaining high accuracy. Bhodia et al. [31] employed VGG16, a deep learning model pre-trained on ImageNet, for malware image classification. Fine-tuning the model on malware-specific datasets improved detection accuracy and showed particular promise in identifying zero-day malware attacks.

Prima and Bouhorma [32] leveraged transfer learning to adapt CNN models for malware detection, converting binary malware files into grayscale images. Their results highlighted the efficiency of transfer learning in resource-constrained environments. Similarly, Ahmed et al. [33] proposed a framework that combines transfer learning with convolutional neural networks to classify malware binaries. They used data augmentation and fine-tuning techniques, which enhanced detection accuracy while reducing computational demands.

Zhao et al. [34] extended the concept of transfer learning by developing a multi-channel framework that visualizes malware binaries as images. By fine-tuning pre-trained CNN models for malware detection, their study emphasized the importance of integrating diverse data channels to improve robustness. Panda et al. [35] investigated transfer learning in IoT environments by using pre-trained models to classify malware image representations. They introduced preprocessing techniques to standardize input sizes but noted the challenge of information loss during the conversion process.

Ngo et al. [36] introduced a hybrid approach that combines transfer learning with static and dynamic feature analysis. Their method minimized computational overhead while improving malware detection accuracy, particularly for obfuscated malware. Tasyurek and Arslan [37] developed RT-Droid, a real-time Android malware detection framework based on transfer learning. By examining static features like API calls and permissions, their framework achieved 98.6% accuracy, demonstrating its effectiveness in real-time scenarios.

Transfer learning techniques such as grayscale image, multi-channel frameworks, and the combination of static-dynamic features have also improved malware detection. These methods enable the models to fine-tune with ease for malware related tasks without the need for large amounts of datasets or computational resources. However, preprocessing requirements, input standardization challenges, and dataset imbalances remain significant obstacles.

However, challenges such as information loss during preprocessing, dataset imbalances, and the complexity of fine-tuning pre-trained models must be addressed for transfer learning to realize its full potential in malware detection.

C. Explainable AI-Based Malware Analysis

Explainable Machine Learning (XAI) is a strong ally to improve the transparency and reliability of malware detection systems. Additionally, XAI helps explain how complex machine learning algorithms make decisions based on identifying key features and patterns [38]. A growing field of research has focused on explainability, aimed at clarifying and simplifying machine learning reasoning and decision-making processes. Explainability methods clarify how ML models work, assisting developers and users in understanding their behavior [39].

A variety of explainable AI methods including SHAP and LIME, provide interpretable explanations for model outputs which assist analysts in understanding how classifications are made. This transparency improves decision-making helps refine malware detection methods and supports the development of stronger more reliable malware detection systems [40]. Moreover, XAI aids in identifying potential inaccuracies in the models and data leading to fairer and more balanced systems. With XAI, the factors that influence a model's decision are clarified resulting in fewer false positives and negatives ultimately improving malware detection accuracy and effectiveness.

Ladarola et al. [41] developed a deep learning model for classifying malware families based on their visual representations, achieving 93.4% accuracy. They used LIME to explain model decisions and activation maps to assess model reliability, identify biases, and improve robustness. Alani and Awad [42] proposed PAIRED, an efficient Android malware detection system using XML techniques. By extracting static features from applications, PAIRED achieved an accuracy rate of over 98% while consuming minimal resources. SHAP values were utilized to explain the decision-making process, enhancing the transparency and interpretability of their model.

Liu et al. [43] focused on the interpretability of machine learning models for Android malware detection. They examined the internal workings of models, including decision trees, to identify key features and patterns involved in malware classification. Similarly, Kinkead et al. [44] utilized LIME to enhance the interpretability of CNN-based predictions. Their study validated the consistency between CNN's feature selection and LIME's interpretability framework, showcasing the utility of LIME in corroborating CNN-based malware detection.

According to H. Manthena [45], many malware analysis models lack transparency, making them difficult to trust. This problem was addressed by integrating XML techniques, such as KernelSHAP, TreeSHAP, and DeepSHAP, into online malware detection. These techniques evaluated performance metrics, improved interpretability, and improved the trustworthiness of security systems. In another study, Manthena et al. [46] developed a malware detection system using SHAP to reveal the inner workings of CNNs and Feedforward Neural Networks (FFNNs). The system provided insights into model predictions, improving transparency and trust in the results.

Lu and Thing [47] proposed an Android malware detection framework employing three model explanation methods: Modern Portfolio Theory (MPT), SHAP, and LIME. These methods were compared based on their ability to provide explanations, with MPT demonstrating utility in analyzing adversarial samples. Additionally, Pan et al. [48] developed a hardware-assisted malware detection framework using regression-based explainable machine learning techniques to overcome prediction inaccuracies and lack of transparency.

Sharma et al. [49] designed a traffic analysis-based malware detection system based on traffic analysis that utilizes human-readable network traffic features. Decision tree-based models were employed, enabling more interpretable malware detection. Iadarola et al. [50] proposed an interpretable approach for detecting and categorizing Android malware fami-

lies. By visually representing malware as images and feeding them into an explainable deep learning model, their system achieved both high performance and transparency.

Explainable machine learning techniques such as SHAP, LIME, and XML are very effective in increasing the interpretability and transparency of malware detection systems. These methods contribute to enhancing the interpretability of the factors behind the decisions and therefore enhancing the reliability and accuracy of the results. However, challenges such as scalability in real-time environments and computational overhead prevent broad adoption.

XAI plays a vital role in improving transparency and reliability in malware detection, but it struggles with challenges like real-time scalability, computational demands, and the trade-off between performance and interpretability.

Table II provides a summary of the analyzed papers.

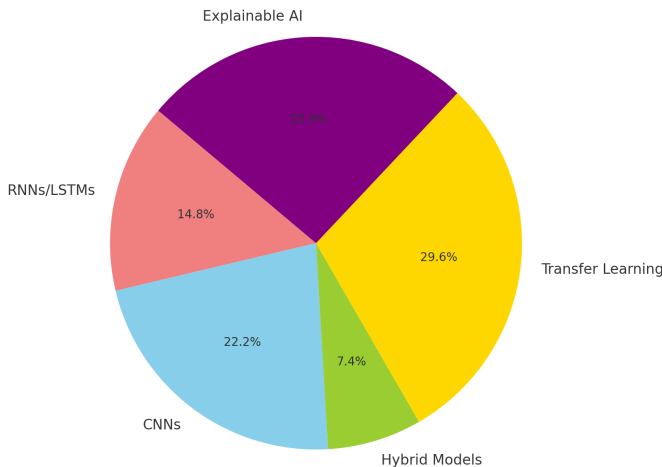


Fig. 2. Proportion of machine learning techniques used in malware detection studies.

Fig. 2 shows the use of various machine learning techniques in malware detection from the studies examined in the paper. Techniques like transfer learning and explainable XAI are prominently represented, indicating their increasing significance in addressing challenges such as limited resources and improving model interpretability. Deep learning methods, including CNNs and RNNs/LSTMs, also play a vital role in identifying complex patterns and managing sequential data, reflecting their foundational importance. Although hybrid models are less commonly employed, they showcase the potential of combining multiple approaches to achieve higher detection accuracy. This analysis underscores the diversity of methodologies adopted in malware detection research and their continuing progress.

While machine learning has great potential to enhance malware detection, it also comes with notable challenges. Deep learning demands significant computational resources, transfer learning contends with issues like imbalanced datasets, and explainable AI poses integration complexities. Tackling these obstacles is crucial to advancing malware detection methods. The next section delves into these challenges, highlighting

gaps in current approaches and exploring opportunities for improvement.

V. LIMITATIONS OF CURRENT APPROACHES

Although machine learning techniques have made significant advances in malware detection, there are still many limitations, affecting their practicality and scalability. Based on the studies reviewed, this section discusses the limitations of deep learning, transfer learning, and explainable AI.

A. Limitations of Deep Learning

Deep learning techniques have revolutionized malware detection, yet they are not without challenges:

- Rhode et al. [17] emphasized that their LSTM-based approach for early-stage malware detection heavily relied on large, high-quality datasets, limiting its practical applicability in real-world environments. Moreover, LSTMs are computationally intensive, susceptible to noise and adversarial attacks, and often face challenges in generalizing to previously unseen malware families. The complexity of interpreting LSTM models further complicates their adoption, as it undermines trust and impedes effective debugging
- Elayan and Mustafa [18] observed that their GRU-based model for Android malware detection, despite its high accuracy, posed challenges in resource-limited environments such as mobile devices or IoT systems. The model's computational demands resulted in higher energy consumption, longer processing times, and reduced battery efficiency on mobile devices. Addressing this issue may involve developing lightweight architectures or employing model compression techniques to enhance its suitability for resource-constrained settings.
- Catak et al. [19] highlighted the effectiveness of LSTMs in analyzing sequential data but noted their vulnerability to adversarial attacks. Subtle, malicious perturbations in input data can easily deceive LSTMs, resulting in misclassification. This weakness poses a significant security threat in practical applications, as attackers can exploit it to bypass detection mechanisms.
- McDole et al. [20] observed that while deep learning models provide scalability and flexibility in cloud environments, they come with high computational costs. This results in increased latency, elevated operational expenses, and lower energy efficiency, rendering them unsuitable for real-time applications with strict latency demands, such as autonomous vehicles or industrial control systems.
- Ravi et al. [21] identified that their multi-view attention framework, while effective, heavily relies on specialized hardware such as GPUs or TPUs for efficient training and inference. This dependence limits its usability in resource-constrained settings, including IoT and edge devices, where computational resources and memory are restricted. Additionally, the need for specialized hardware can elevate both deployment

TABLE II. RELATED WORK ANALYSIS

Ref.	Addressed Problems	Machine learning techniques used
[17]	Early detection of malware to predict malicious behavior in its initial stages	RNNs, LSTM
[18]	Detection of Android malware using deep learning methods for increased accuracy	CNNs
[19]	Sequential analysis of malware behavior through API call patterns in Windows executable	RNNs, LSTMs
[20]	Behavioral analysis of malware in cloud IaaS environments	CNNs, RNNs
[21]	Malware detection in smart healthcare systems using a multi-view attention-based approach	Attention-based DL Framework, Multi-view models
[22]	Comparative evaluation of deep learning techniques for detecting Android malware	CNNs, RNNs, DL models
[23]	Automatic detection of Android malware using static analysis techniques combined with deep learning	Static Analysis with Deep Neural Networks (DNNs)
[24]	Analysis of malware using a combination of traditional machine learning and deep learning methods	Machine Learning (Random Forest, SVM), CNNs
[25]	Development of a hybrid model for detecting malware on Android devices by combining multiple techniques	Hybrid Model combining Decision Trees and Neural Networks
[26]	Detection of Java-based malware using a combination of image-based and behavior-based features	CNN
[30]	Static malware classification by leveraging pre-trained models to improve accuracy	Transfer Learning with Deep Neural Networks
[31]	Malware classification using image-based representations of malware and transfer learning	Transfer Learning with CNNs
[32]	Malware classification leveraging pre-trained models for enhanced detection	Transfer Learning
[33]	Malware classification by leveraging the Inception V3 architecture and transfer learning	Transfer Learning with Inception V3
[34]	Visual malware classification using a multi-channel approach combined with transfer learning	Transfer Learning with CNNs
[35]	Malware detection in IoT environments through image-based transfer learning techniques	Transfer Learning
[36]	Efficient malware detection using combined static and dynamic features enhanced by transfer learning	Transfer Learning
[37]	Real-time Android application analysis utilizing transfer learning for malware detection	Transfer Learning with CNN Models
[41]	Understanding deep learning predictions in image-based malware detection using activation maps	Deep Learning with Activation Maps
[42]	Lightweight and explainable approaches for Android malware detection	Lightweight Explainable AI for Android Malware Detection
[43]	Enhancing understanding of Android malware detection models performance through explainable AI approaches	Explainable AI applied to Android Malware Detection
[44]	Improving interpretability of CNNs in Android malware detection	CNNs, Explainability
[45]	Development of explainable machine learning frameworks for malware analysis	Explainable Machine Learning
[46]	Providing insights into black-box models used for online malware detection, improving transparency and trust	Explainable Machine Learning
[47]	Providing explanations for predictions of AI-based malware detectors, especially for malicious Android apps	Explainable AI for Malware Detection
[48]	Utilizing hardware-assisted techniques to detect malware with explainable machine learning models	Explainable Machine Learning
[49]	Developing an extensible and explainable system for analyzing network traffic and detecting malware	TTP-based Explainable Systems, Machine Learning
[50]	Enhancing interpretability in deep learning models for detecting and categorizing mobile malware families	Deep Learning, Interpretability

costs and energy consumption, posing challenges for broader adoption.

- Calik Bayazit et al. [22] highlighted the effectiveness of hybrid architectures that leverage the advantages of various deep learning models. However, these architectures tend to be highly complex, presenting challenges in terms of training, optimization, and efficient deployment. Additionally, identifying the ideal combination of models and hyperparameters for a given task can be both time-intensive and computationally demanding.
- Ibrahim et al. [23] observed that integrating static analysis with deep learning improved malware detection accuracy. Despite its benefits, this approach often demands considerable domain expertise to effectively derive and refine features from static analysis outputs. Additionally, the integration of static analysis tools with deep learning models presents challenges in complexity and resource requirements, necessitating significant computational power and specialized infrastructure.
- Patil and Deng [24] highlighted that, although deep learning models outperform traditional approaches, their high training costs and demanding hardware requirements pose significant challenges. These scalability limitations restrict their usability in resource-constrained settings and complicate their application in scenarios requiring frequent retraining, such as adapting to emerging threats.
- Rodrigo et al. [25] observed that while their hybrid model, which integrates static and dynamic features, enhanced malware detection accuracy, it also introduced increased inference time. This limitation makes the approach less practical for real-time applications with strict latency demands, such as intrusion detection systems and real-time threat monitoring.
- Obaidat et al. [26] emphasized the effectiveness of their CNN-based method for detecting Java-based malware through visual bytecode representations. However, the conversion of bytecode into visual formats may result in information loss, which can hinder the model's ability to accurately identify subtle behavioral patterns and nuances of malware.

B. Limitations of Transfer Learning

Transfer learning has shown to be effective in reducing training time and resource consumption, but it faces the following challenges:

- Chen et al. [30] noted that converting malware binaries into image representations for the use of convolutional neural networks can result in substantial information loss. This reduction in critical data may compromise the model's accuracy and its capacity to effectively analyze complex malware behaviors and traits.
- Bhodia et al. [31] highlighted that, although transfer learning demonstrated high accuracy in malware detection, its effectiveness is significantly affected by

class imbalance in the training dataset. When certain malware classes are underrepresented, the resulting models may become biased, leading to reduced generalization capabilities for unseen samples belonging to these underrepresented categories.

- Prima and Bouhorma [32] noted that although transfer learning provides an effective initial framework, it often necessitates substantial fine-tuning on specific malware datasets. This process can be both computationally intensive and time-consuming, demanding considerable resources and potentially delaying the quick deployment and adaptation needed to address emerging threats.
- Zhao et al. [34] highlighted the effectiveness of multi-channel frameworks in combining diverse data sources. However, merging information from channels like static analysis, dynamic analysis, and network traffic introduces considerable computational overhead and complexity. Effectively processing and integrating these varied data streams necessitates meticulous optimization of both the model architecture and the training methodology.
- Panda et al. [35] emphasized the difficulties of pre-processing and standardizing input formats for IoT malware detection. The variation among IoT devices and the diversity of malware samples introduce significant challenges in creating consistent input structures, which can complicate data preprocessing workflows and potentially affect the overall performance of the models.
- Ngo et al. [36] demonstrated that integrating static and dynamic features within transfer learning models enhances accuracy. However, this methodology presents notable challenges, including increased training complexity, extended training durations, and difficulties in fine-tuning hyperparameters. Furthermore, the combined use of static and dynamic analysis may lead to longer inference times, potentially hindering the system's efficiency in real-time scenarios.
- Tasyurek and Arslan [37] highlighted that their RT-Droid system demonstrated effectiveness in real-time Android malware detection. However, maintaining its efficacy in the face of rapidly evolving malware requires frequent updates to its models and feature sets. This ongoing need for retraining and redeployment poses significant challenges, including increased resource demands and operational complexity.

C. Limitations of Explainable AI

Explainable AI techniques have enhanced the interpretability of malware detection models, but there are still several limitations:

- Ladarola et al. [41] demonstrated that LIME effectively enhances interpretability by offering localized explanations of model predictions. However, its substantial computational demands render it impractical

for real-time applications with strict latency constraints, such as online malware detection or intrusion detection systems.

- Alani and Awad [42] highlighted the utility of SHAP values in offering valuable insights into the elements shaping model predictions, thereby improving transparency. However, incorporating SHAP value computations into resource-limited environments, such as mobile or edge devices, poses challenges due to the substantial computational resources required for their calculation.
- Liu et al. [43] observed that decision trees, despite their inherent interpretability, encounter scalability challenges when dealing with large-scale malware datasets. The expansion in the number of features and data samples can significantly increase the tree's complexity, resulting in prolonged training durations, higher memory usage, and reduced overall efficiency.
- Kinkead et al. [44] demonstrated the effectiveness of LIME in explaining CNN-based malware detection model predictions. However, they identified scalability as a significant limitation, particularly when handling large and complex malware datasets. This constraint poses challenges for its practical application in real-world scenarios requiring swift analysis and explanation of extensive malware samples.
- Lu and Thing [47] investigated various explainability techniques, including MPT. However, they identified that MPT has shortcomings in effectively handling adversarial attacks. Such adversarial examples, designed to exploit vulnerabilities in the model, can undermine the reliability of explainability methods, resulting in distorted or inaccurate interpretations.
- Pan et al. [48] introduced a hardware-assisted framework aimed at enhancing the transparency and interpretability of deep learning models for malware detection. Despite its advantages, the reliance on specialized hardware restricts its use in general-purpose systems. Additionally, this dependency may elevate deployment costs, presenting a barrier to broader implementation.
- Manthena et al. [46] highlighted that SHAP enhances the interpretability of deep learning models by generating feature importance scores. However, calculating SHAP values introduces substantial computational overhead, which can adversely affect real-time system performance. This limitation poses a bottleneck in high-throughput malware analysis workflows, hindering their efficiency in time-sensitive applications.
- Sharma et al. [49] emphasized the effectiveness of decision-tree-based models in traffic analysis and malware detection. However, these models are highly susceptible to obfuscation techniques, which are employed by malware developers to modify the code's structure while retaining its functionality. Such obfuscation methods can compromise the model's ability to accurately detect and classify malware, posing a significant challenge in maintaining detection reliability.

The analyzed papers demonstrate that considerable progress has been made in employing machine learning approaches to detect malware; still, various limitations make it difficult to implement the obtained outcomes.

- 1) Deep Learning:
 - Datasets must be large and high quality to be effective in environments with limited data.
 - Vulnerable to adversarial attacks that manipulate model predictions.
 - The computational requirements make it difficult to deploy in resource-constrained systems.
- 2) Transfer Learning:
 - In preprocessing steps, such as converting malware binaries into images, losing information can be increased.
 - Dataset imbalances affect model generalizability.
 - Fine-tuning pre-trained models is computationally expensive and time-intensive.
- 3) Explainable AI (XAI):
 - High computational overhead deter scalability for real time applications.
 - Finding a balance between transparency and efficiency is still difficult.
 - Compatibility with existing security systems is a high level of integration and thus calls for domain-specific solutions.

While these challenges present significant hurdles, they also highlight critical areas that require further exploration and innovation. Overcoming these barriers is vital to realize the full potential of machine learning in malware detection. With advancements in techniques such as preprocessing optimization, improved dataset balancing, enhanced computational efficiency, and seamless system integration, limitations can be addressed effectively. The next section delves into specific strategies and emerging possibilities that aim to enhance the scalability, reliability, and transparency of machine learning-based malware detection systems.

VI. FUTURE DIRECTIONS AND COUNTERMEASURES

Detecting and analyzing malware has made significant progress; however, a number of challenges still exist. This section presents potential future research directions and actionable countermeasures for improving machine learning-based malware detection systems' scalability, robustness, and transparency.

A. Future Directions

- 1) Deep Learning:
 - Federated Learning for Privacy: Federated learning enables collaborative model training while ensuring data privacy by retaining data on individual devices, thus reducing the risk of sensitive information exposure. However, privacy concerns persist, prompting ongoing research into methods such as differential privacy to enhance protection and address these challenges [51].

- **Hybrid Architectures:** Combining CNNs and RNNs allows for combining their complementary capabilities, with CNNs excelling at identifying spatial patterns and RNNs adept at analyzing sequential data [67]. This integration enables the model to effectively capture both spatial and temporal relationships within malware datasets, offering enhanced accuracy and robustness in malware detection.
- **Optimized Lightweight Models:** Design models tailored for resource-constrained environments such as IoT devices or edge platforms by employing techniques like model pruning, quantization, and knowledge distillation. These methods significantly reduce model size and computational demands while maintaining acceptable levels of accuracy. Sze et al. [68] highlighted the value of optimizing deep neural networks for embedded systems, showcasing how such strategies can enhance energy efficiency and make models more suitable for real-time malware detection in low-power, latency-sensitive scenarios.

2) Transfer Learning:

- **Improved Preprocessing Techniques:** Advancing preprocessing methods is essential to retain critical features while minimizing the loss of information during data transformation. Techniques such as adaptive feature scaling and intelligent data augmentation can strike a balance, ensuring that key data characteristics are preserved for better model accuracy [69]. This approach has proven beneficial in applications where maintaining high-dimensional data integrity is crucial.
- **Cross-Domain Adaptability:** Developing models that perform effectively across varying domains, such as IoT and cloud infrastructures, is a vital research direction. Leveraging strategies like domain adaptation and transfer learning can enable these models to generalize efficiently across diverse environments, addressing discrepancies in data distributions and ensuring consistent performance [70].
- **Streamlined Fine-Tuning Processes:** Streamlining fine-tuning procedures is critical to enhance efficiency and performance. Automated tools such as AutoML and hyperparameter optimization frameworks can simplify this process by automating the search for optimal model parameters [72]. This reduces manual intervention and significantly improves the model's overall effectiveness.
- **Standardized Datasets for Malware Detection:** Ensuring the availability of standardized and balanced datasets is essential for reliable evaluation and benchmarking of malware detection models. These datasets should include diverse malware samples and simulate real-world scenarios to enhance the generalizability and robustness of the models [74].

3) Explainable AI:

- **Efficient Explanation Models:** Creating lightweight XAI frameworks tailored for real-time applications is essential. These models should focus on reducing

computational overhead while delivering clear, actionable insights. Streamlining algorithms like SHAP or LIME for efficient processing can make XAI more applicable in scenarios requiring immediate decision-making.

- **Adaptive Explanations:** Develop Dynamic explainability frameworks are crucial for addressing evolving malware patterns. By continuously learning and adapting to new threats, these systems can provide context-specific explanations that remain relevant over time. Such adaptability ensures that cybersecurity measures evolve in tandem with emerging challenges.
- **Integration with Security Frameworks:** Modular XAI tools designed for seamless integration with existing cybersecurity systems can significantly enhance decision-making [73]. These tools can act as plug-and-play components, working in harmony with established security workflows to improve detection accuracy and transparency .

B. Countermeasures

1) Deep Learning:

- **Defensive Mechanisms for Adversarial Inputs:** Building robust defenses against adversarial attacks is essential for bolstering the security of machine learning models. A widely adopted technique is adversarial training, which incorporates adversarial examples into the training process to enhance the model's resilience. For example, the study [52] introduces an Ulam-stability-based method that significantly improves model robustness against such attacks. Another promising method involves using an anti-adversarial module, as outlined in [53]. This approach applies targeted counter-adversarial treatments to input samples, effectively reducing the impact of adversarial perturbations. By employing these advanced techniques, machine learning models can achieve heightened resistance to adversarial inputs, ultimately increasing their reliability in security-critical applications.
- **Augmentation Techniques for Data:** Data augmentation has emerged as a vital technique to address limitations in dataset sizes. By using GANs, synthetic but realistic data can be generated, significantly enriching training datasets. For instance, [62] discuss how GANs can create synthetic malware samples, which help balance datasets and improve the robustness of machine learning models in detecting malware. This approach not only reduces reliance on large datasets but also enhances model generalizability by diversifying training inputs.
- **Improving Adversarial Resilience:** Adversarial training has become essential in strengthening models against adversarial attacks. For example, Madry et al. propose incorporating adversarial samples during the training process to improve a model's robustness [63]. By simulating potential attacks, this method ensures that models can better resist manipulation, making them more reliable in security-critical environments

- **Leveraging Hardware Solutions:** To address the computational demands of deep learning models, leveraging specialized hardware like Tensor Processing Units (TPUs) and GPUs has proven effective. Jouppi et al. demonstrate the use of TPUs to accelerate deep learning tasks, showing how such hardware can reduce training times and energy consumption while maintaining high performance [64].

2) Transfer Learning:

- **Handling Dataset Imbalances:** Addressing dataset imbalances is crucial for developing effective machine learning models. Techniques such as oversampling the minority class and generating synthetic data have proven effective in mitigating these imbalances. For instance, the Synthetic Minority Over-sampling Technique (SMOTE) creates synthetic samples by interpolating between existing minority instances, thereby enhancing model performance on imbalanced datasets [54]. Moreover, recent advancements have introduced methods like Localized Random Affine ShadowSampling (LoRAS), which oversamples from an approximated data manifold of the minority class, addressing limitations associated with traditional techniques [55]. Through the implementation of these strategies, models can achieve better balance and improved prediction accuracy.
- **Optimized Preprocessing Workflows:** Effective preprocessing is critical to ensuring the success of machine learning models in malware detection. Optimizing these workflows not only preserves essential data features but also reduces computational overhead, enabling efficient and scalable model deployment. Techniques such as feature selection and dimensionality reduction, as presented in [56], can streamline preprocessing by focusing on the most informative attributes while discarding redundant data. Additionally, leveraging automated preprocessing pipelines, as highlighted in [57], can dynamically adapt preprocessing strategies to diverse datasets and application requirements.
- **Distributed Training Systems:** Distributed training systems enable the efficient processing of large datasets and complex machine learning models by leveraging the computational power of multiple machines. This approach not only reduces resource bottlenecks but also accelerates the training process, making it ideal for scaling malware detection models to meet real-world demands. For instance, distributed frameworks such as Apache Spark and TensorFlow Distributed offer robust architectures for handling extensive data and computations across multiple nodes [58], [59]. These systems optimize training by partitioning tasks, balancing workloads, and parallelizing computations. Additionally, advancements in federated learning and edge computing can complement distributed systems, enabling secure and decentralized training of models without compromising data privacy [60], [61].

3) Explainable AI:

- **Real-Time XAI Models:** Real-time XAI frameworks are essential for applications requiring rapid decision-making. Simplified versions of SHAP and LIME can reduce computational overhead, enabling real-time processing. Accordingly, in the study by [65] real-time SHAP implementation demonstrated effective trade-offs between interpretability and speed, ensuring timely insights without significant computational delays.
- **Combining Explanation Approaches:** Integrating localized explanation strategies, like LIME, with global methods, such as SHAP, provides a comprehensive understanding of model decisions. This hybrid approach balances detailed insights with overarching trends, improving both interpretability and model validation. A study by [66] highlights the effectiveness of combining explanation techniques to enhance trust in machine learning models without compromising accuracy.

According to the outlined future directions and countermeasures, researchers can go a long way in enhancing the detection of malware. It seeks to optimise the ML approaches to increase their applicability on the current and emerging complex cybersecurity challenges.

VII. CONCLUSIONS

This paper offers a detailed review of the latest trends and challenges in applying machine learning to malware detection and analysis, with a focus on its increasing role in combating complex cyber threats. Machine learning has shown great promise as a versatile tool, providing scalability, adaptability, and improved pattern recognition for identifying and analyzing malware. However, significant challenges remain, including vulnerabilities to adversarial attacks, biases in datasets, and a lack of transparency in many deep learning models.

By exploring methods such as deep learning, transfer learning, and explainable AI, this review highlights both their strengths and the challenges they face, including high computational requirements and reliance on feature extraction. These obstacles underscore the need for innovative approaches to improve the effectiveness and dependability of machine learning systems in malware detection.

To overcome these limitations, this paper proposes several novel strategies, such as leveraging distributed computing, refining preprocessing methods, and enhancing the integration of explainability techniques. As a result of these advancements, machine learning models will become more robust, efficient, and transparent, ensuring their effectiveness in addressing malware threats as they evolve.

FUNDING

This work was funded by King Faisal University, Saudi Arabia [Grant No. KFU250089].

ACKNOWLEDGMENT

This work was supported through the Annual Funding track by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. KFU250089].

CONFLICTS OF INTEREST

All authors declare no conflict of interest.

REFERENCES

- [1] N. Z. Gorment, A. Selamat, L. K. Cheng, and O. Krejcar, "Machine learning algorithm for malware detection: Taxonomy, current challenges and future directions," *IEEE Access*, 2023.
- [2] M. S. Akhtar and T. Feng, "Malware analysis and detection using machine learning algorithms," *Symmetry*, vol. 14, no. 11, pp. 2304, 2022.
- [3] M. S. Akhtar and T. Feng, "IOTA-based anomaly detection machine learning in mobile sensing," *EAI Endorsed Transactions on Creative Technologies*, vol. 9, pp. 172814, 2022, doi:10.4108/eai.9-12-2022.172814.
- [4] D. Ucci, L. Aniello, and R. Baldoni, "Survey of machine learning techniques for malware analysis," *Computers & Security*, vol. 81, pp. 123–147, 2019, doi:10.1016/j.cose.2018.11.001.
- [5] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *SN Computer Science*, vol. 2, no. 3, pp. 160, 2021.
- [6] R. Komatwar and M. Kokare, "A survey on malware detection and classification," *Journal of Applied Security Research*, pp. 1–31, Aug. 2020.
- [7] O. Aslan, M. Ozkan-Okay, and D. Gupta, "A review of cloud-based malware detection system: Opportunities, advances and challenges," *European Journal of Engineering and Technology Research*, vol. 6, no. 3, pp. 1–8, Mar. 2021.
- [8] R. Komatwar and M. Kokare, "Retracted article: A survey on malware detection and classification," *Journal of Applied Security Research*, vol. 16, no. 3, pp. 390–420, 2021.
- [9] M. Sikorski and A. Honig, *Practical Malware Analysis: The Hands-On Guide to Dissecting Malicious Software*. San Francisco, CA, USA: No Starch Press, 2012.
- [10] O. Aslan and A. A. Yilmaz, "A new malware classification framework based on deep learning algorithms," *IEEE Access*, vol. 9, pp. 87936–87951, 2021.
- [11] O. Aslan and R. Samet, "Investigation of possibilities to detect malware using existing tools," in *Proc. IEEE/ACS 14th Int. Conf. Computer Systems and Applications (AICCSA)*, 2017, pp. 1277–1284.
- [12] M. Ijaz, M. H. Durad, and M. Ismail, "Static and dynamic malware analysis using machine learning," in *Proc. 16th Int. Bhurban Conf. Applied Sciences and Technology (IBCAST)*, 2019, pp. 687–691.
- [13] N. Tarar, S. Sharma, and C. R. Krishna, "Analysis and classification of android malware using machine learning algorithms," in *Proc. 3rd Int. Conf. Inventive Computation Technologies (ICICT)*, 2018, pp. 738–743.
- [14] V. Rao and K. Hande, "A comparative study of static, dynamic and hybrid analysis techniques for android malware detection," *Int. J. Eng. Dev. Res.*, vol. 5, no. 2, pp. 1433–1436, 2017.
- [15] U. H. Tayyab *et al.*, "A survey of the recent trends in deep learning based malware detection," *Journal of Cybersecurity and Privacy*, vol. 2, no. 4, pp. 800–829, 2022.
- [16] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no. 1, pp. 41–50, 2018.
- [17] M. Rhode, P. Burnap, and K. Jones, "Early-stage malware prediction using recurrent neural networks," *Computers & Security*, vol. 77, pp. 578–594, 2018.
- [18] O. N. Elyan and A. M. Mustafa, "Android malware detection using deep learning," *Procedia Computer Science*, vol. 184, pp. 847–852, 2021.
- [19] F. O. Catak *et al.*, "Deep learning-based sequential model for malware analysis using Windows EXE API calls," *PeerJ Computer Science*, vol. 6, pp. e285, 2020.
- [20] A. McDole *et al.*, "Deep learning techniques for behavioral malware analysis in cloud IaaS," in *Malware Analysis Using Artificial Intelligence and Deep Learning*, Springer, 2021, pp. 269–285.
- [21] V. Ravi *et al.*, "A multi-view attention-based deep learning framework for malware detection in smart healthcare systems," *Computer Communications*, vol. 195, pp. 73–81, 2022.
- [22] E. C. Bayazit *et al.*, "Deep learning-based malware detection for Android systems: A comparative analysis," *Tehnicki Vjesnik*, vol. 30, no. 3, pp. 787–796, 2023.
- [23] M. Ibrahim *et al.*, "A method for automatic Android malware detection based on static analysis and deep learning," *IEEE Access*, vol. 10, pp. 117334–117352, 2022.
- [24] R. Patil and W. Deng, "Malware analysis using machine learning and deep learning techniques," in *Proc. 2020 SoutheastCon*, vol. 2, pp. 1–7.
- [25] C. Rodrigo, S. Pierre, R. Beaubrun, and F. El Khoury, "A hybrid machine learning-based malware detection model for Android devices," *Cybersecurity and Data Science*, pp. 194, 2021.
- [26] I. Obaidat, M. Sridhar, K. M. Pham, and P. H. Phung, "Jadeite: A novel image-behavior-based approach for Java malware detection using deep learning," *Computers & Security*, vol. 113, pp. 102547, 2022.
- [27] R. Ribani and M. Marengoni, "A survey of transfer learning for convolutional neural networks," in *Proc. 32nd SIBGRAPI Conf. Graphics, Patterns and Images Tutorials (SIBGRAPI-T)*, 2019, pp. 47–57.
- [28] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Artificial Neural Networks and Machine Learning–ICANN 2018*, 2018, pp. 270–279.
- [29] H. M. K. Barznji, "Transfer learning as a new field in machine learning," *Int. Archives Photogrammetry, Remote Sensing Spatial Information Sciences*, vol. 44, pp. 343–349, 2020.
- [30] L. Chen, "Deep transfer learning for static malware classification," *arXiv preprint arXiv:1812.07606*, 2018.
- [31] N. Bhodia, P. Prajapati, F. Di Troia, and M. Stamp, "Transfer learning for image-based malware classification," *arXiv preprint arXiv:1903.11551*, 2019.
- [32] B. Prima and M. Bouhorma, "Using transfer learning for malware classification," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 44, pp. 343–349, 2020.
- [33] M. Ahmed, N. Afreen, M. Ahmed, M. Sameer, and J. Ahamed, "An Inception V3 approach for malware classification using machine learning and transfer learning," *International Journal of Intelligent Networks*, vol. 4, pp. 11–18, 2023.
- [34] Z. Zhao, S. Yang, and D. Zhao, "A new framework for visual classification of multi-channel malware based on transfer learning," *Applied Sciences*, vol. 13, no. 4, pp. 2484, 2023.
- [35] P. Panda, O. K. CU, S. Marappan, S. Ma, and D. V. Nandi, "Transfer learning for image-based malware detection for IoT," *Sensors*, vol. 23, no. 6, pp. 3253, 2023.
- [36] M. V. Ngo, T. H. Truong, D. Rabadi, J. Y. Loo, and S. G. Teo, "Fast and efficient malware detection with joint static and dynamic features through transfer learning," in *Proc. Int. Conf. Applied Cryptography and Network Security*, pp. 503–531, 2023.
- [37] M. Tasyurek and R. S. Arslan, "Rt-droid: A novel approach for real-time Android application analysis with transfer learning-based CNN models," *Journal of Real-Time Image Processing*, vol. 20, no. 3, pp. 55, 2023.
- [38] U. Bhatt, A. Xiang, S. Sharma, *et al.*, "Explainable machine learning in deployment," in *Proc. 2020 Conf. Fairness, Accountability, and Transparency*, 2020, pp. 648–657.
- [39] S. M. Lundberg *et al.*, "Explainable machine-learning predictions for the prevention of hypoxaemia during surgery," *Nature Biomedical Engineering*, vol. 2, no. 10, pp. 749, 2018.
- [40] X. Zhong, B. Gallagher, S. Liu, *et al.*, "Explainable machine learning in materials science," *npj Computational Materials*, vol. 8, no. 1, pp. 204, 2022.
- [41] G. Ladarola, F. Mercaldo, F. Martinelli, and A. Santone, "Assessing deep learning predictions in image-based malware detection with activation maps," in *Proc. Security and Trust Management: 18th Int. Workshop, STM 2022*, vol. 13867, pp. 104, 2023.
- [42] M. M. Alani and A. I. Awad, "Paired: An explainable lightweight Android malware detection system," *IEEE Access*, vol. 10, pp. 73214–73228, 2022.
- [43] Y. Liu, C. Tantithamthavorn, L. Li, and Y. Liu, "Explainable AI for Android malware detection: Towards understanding why the models perform so well?" in *Proc. IEEE 33rd Int. Symp. Software Reliability Engineering (ISSRE)*, pp. 169–180, 2022.

- [44] M. Kinkad, S. Millar, N. McLaughlin, and P. O’Kane, “Towards explainable CNNs for Android malware detection,” *Procedia Computer Science*, vol. 184, pp. 959–965, 2021.
- [45] H. Manthena, *Explainable Machine Learning Based Malware Analysis*, Ph.D. dissertation, North Carolina Agricultural and Technical State University, 2022.
- [46] H. Manthena, J. C. Kimmel, M. Abdelsalam, and M. Gupta, “Analyzing and explaining black-box models for online malware detection,” *IEEE Access*, vol. 11, pp. 25237–25252, 2023.
- [47] Z. Lu and V. L. Thing, “How does it detect a malicious app? Explaining the predictions of AI-based malware detector,” in *Proc. IEEE 8th Int. Conf. Big Data Security on Cloud (BigDataSecurity)*, pp. 194–199, 2022.
- [48] Z. Pan, J. Sheldon, and P. Mishra, “Hardware-assisted malware detection using explainable machine learning,” in *Proc. IEEE 38th Int. Conf. Computer Design (ICCD)*, pp. 663–666, 2020.
- [49] Y. Sharma, S. Birnbach, and I. Martinovic, “Radar: A TTP-based extensible, explainable, and effective system for network traffic analysis and malware detection,” 2023.
- [50] G. Iadarola, F. Martinelli, F. Mercaldo, and A. Santone, “Towards an interpretable deep learning model for mobile malware detection and family identification,” *Computers & Security*, vol. 105, pp. 102198, 2021.
- [51] Y. Bi, H. Wang, J. Liu, and X. Zhang, “Enabling privacy-preserving cyber threat detection with federated learning,” *arXiv preprint arXiv:2404.05130*, 2023.
- [52] K. Yan, L. Yang, Z. Yang, and W. Ren, “Enhancing adversarial robustness through stable adversarial training,” *Symmetry*, vol. 16, no. 10, p. 1363, 2024.
- [53] Z. Qin, G. Liu, and X. Lin, “Enhancing model robustness against adversarial attacks with an anti-adversarial module,” in *Pattern Recognition and Computer Vision (PRCV 2023)*, Lecture Notes in Computer Science, vol. 14433, Springer, 2023, pp. 66–78.
- [54] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.
- [55] B. Kovács, I. Bagyinszki, and J. Abonyi, “LoRAS: Localized Random Affine Shadowsampling to Address Class Imbalance,” *Applied Sciences*, vol. 9, no. 16, pp. 3334, 2019, doi:10.3390/app9163334.
- [56] J. Smith and A. Kumar, “Efficient Feature Selection for Malware Detection Using Recursive Feature Elimination,” *Journal of Cybersecurity*, vol. 10, no. 3, pp. 45–60, 2021.
- [57] L. Wang, M. Zhang, and H. Liu, “AutoML-Driven Preprocessing for Scalable Malware Detection,” *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, no. 2, pp. 173–183, 2022.
- [58] M. Zaharia, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, “Apache Spark: A Unified Engine for Big Data Processing,” *Communications of the ACM*, vol. 59, no. 11, pp. 56–65, 2016.
- [59] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, et al., “TensorFlow: A System for Large-Scale Machine Learning,” in *Proc. 12th USENIX Symp. Operating Systems Design and Implementation (OSDI)*, pp. 265–283, 2016.
- [60] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” in *Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS)*, pp. 1273–1282, 2017.
- [61] P. Kairouz, B. McMahan, D. Alistarh, et al., “Advances and Open Problems in Federated Learning,” *Foundations and Trends® in Machine Learning*, vol. 14, no. 1–2, pp. 1–210, 2021.
- [62] J. Wang, W. Xu, J. Chen, and S. Liu, “Data Augmentation for Deep Learning Using Generative Adversarial Networks: A Review,” *IEEE Access*, vol. 9, pp. 141061–141076, 2021.
- [63] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, “Towards Deep Learning Models Resistant to Adversarial Attacks,” in *Proc. Int. Conf. Learning Representations (ICLR)*, 2018.
- [64] N. P. Jouppi, C. Young, N. Patil, D. Patterson, and G. Agrawal, “In-Datcenter Performance Analysis of a Tensor Processing Unit,” in *Proc. 44th Annual Int. Symp. Computer Architecture (ISCA)*, 2017, pp. 1–12.
- [65] X. Zhong, B. Gallagher, S. Liu, et al., “Explainable machine learning in materials science,” *npj Computational Materials*, vol. 8, no. 1, pp. 204, 2022.
- [66] U. Bhatt, A. Xiang, S. Sharma, et al., “Explainable machine learning in deployment,” in *Proc. 2020 Conf. Fairness, Accountability, and Transparency*, pp. 648–657, 2020.
- [67] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [68] V. Sze, Y. Chen, T. Yang, and J. S. Emer, “Efficient processing of deep neural networks: A tutorial and survey,” *Proceedings of the IEEE*, vol. 105, no. 12, pp. 2295–2329, 2017.
- [69] D. Jha, K. W. Liang, and T. Singh, “Advances in preprocessing techniques for deep learning applications,” *IEEE Access*, vol. 8, pp. 34512–34523, 2020.
- [70] W. Li, L. Wang, and E. H. Xing, “Domain adaptation in the era of deep learning,” *Nature Machine Intelligence*, vol. 1, no. 6, pp. 335–346, 2019.
- [71] B. Zoph and Q. V. Le, “AutoML: A method for efficient hyperparameter optimization,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 25, pp. 5862–5869, 2018.
- [72] B. Zoph and Q. V. Le, “AutoML: A method for efficient hyperparameter optimization,” *Proceedings of the National Academy of Sciences*, vol. 115, no. 25, pp. 5862–5869, 2018.
- [73] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, and F. Herrera, “Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI,” *Information Fusion*, vol. 58, pp. 82–115, 2020.
- [74] H. S. Anderson and P. Roth, “EMBER: An open dataset for training and evaluating machine learning models on malware detection,” *arXiv preprint arXiv:1804.04637*, 2018.

SM9 Key Encapsulation Mechanism for Power Monitoring Systems

Chao Hong^{*1}, Peng Xiao², Pandeng Li³, Zhenhong Zhang⁴, Yiwei Yang⁵, Biao Bai⁶
Electric Power Research Institute, China Southern Power Grid, Guangzhou 510663, China^{1,3,5}
Guangdong Provincial Key Laboratory of Power System Network Security, Guangzhou 510663, China^{1,3,5}
Information Center of Yunnan Power Grid Co., Ltd., Kunming 650000, China^{2,4,6}

Abstract—The boundaries of the new power system network are blurred, and data privacy and security are threatened. Although the SM9 algorithm is widely used in power systems to protect data security, its efficiency and security remain the main issues in application. Therefore, an SM9 key encapsulation mechanism (OSM9-KEM-CRF) was proposed to support outsourced decryption and cryptographic reverse firewall. In order to resist the backdoor attacks, we deployed cryptographic reverse firewalls at the terminals and proved that the proposed OSM9-KEM-CRF is ID-IND-CCA2 secure. The cryptographic reverse firewalls maintain functionality, weakly retain security, and weakly resist penetration, thereby enhancing the security of the scheme. In addition, considering the limited computing resources of terminal devices, decryption operations are outsourced to cloud servers in order to reduce the computational burden on the terminals. Compared with other SM9-KEMs, the proposed mechanism not only reduces computational and communication overhead, but also lowers energy consumption. Therefore, the proposed mechanism is more suitable for power monitoring systems.

Keywords—SM9; Outsourced decryption; cryptographic reverse firewall; power monitoring systems

I. INTRODUCTION

With the wide application of IoT technology in power systems, the boundaries of new power system networks are becoming increasingly blurred, and a large number of terminal monitoring devices with limited resources have emerged in power monitoring networks. Although data can be stored in the cloud and pre-processed by cloud servers, thus reducing the storage and computational burden on these terminal devices. However, once the data is out of the direct control of the user, it will face the risk of privacy and security. Information security measures will become the main means of protection. Therefore, there is an urgent need to carry out research on power control systems and lightweight security protection technology.

Chen et al. [1] developed a power monitoring system based on the SM2 algorithm in 2022. However, SM2 algorithm requires complex public key certificate management, while identity-based cryptographic algorithm can avoid complex public key certificate management, and is more sui for new power monitoring systems with many members and dynamic changes in members.

The SM9 is an identity based cryptographic algorithm, which was officially released in 2016 and identified as the

standard algorithm for the cryptographic industry of China [2]. Cheng et al. [3] formally analyzed the security of the SM9 key agreement and the SM9 encryption scheme. Lai et al. [4] proposed Twin-SM9 key encapsulation mechanism using Twin-Hash-ElGamal technique.

A. Related Work

In power monitoring systems, SM9 cryptographic algorithms are favored for their simplified public key certificate management, but their high computational demand on resource-constrained monitoring devices and sensors highlights the need for efficiency optimization. This is especially true for resource-constrained end devices that are widely deployed in power system networks. These devices, such as smart meters, surveillance cameras, and other sensors, have limited computational power, making the complex bilinear mapping operations in the SM9 algorithm the key to improving the decryption efficiency. Ji et al. [5] pointed out that the operation of the SM9 encryption algorithm consumes a large amount of time and computational resources, which makes it challenging to run it on resource-constrained devices. Wang et al. [6] improved the complex operations in SM9 cryptographic algorithm, which improved the computational efficiency of SM9 algorithm to a certain extent, but it is still a large burden for resource-constrained lightweight devices, and could not solve the problem fundamentally. Lai et al. [7] proposed an efficient online/offline identity-based encryption for this purpose, which provides an idea for the implementation of SM9 cryptographic algorithm on lightweight devices. Sun et al. [8] investigated the SM9-IBE encryption scheme based on online/offline techniques. Peng et al. [9] developed an efficient certificate-free online/offline signature scheme and created a lightweight data authentication protocol specifically for WBAN. Liu et al. [10] introduced outsourcing decryption technique in attribute encryption scheme to reduce the computational overhead of the user. This considers the use of outsourcing technique to solve the problem of computational difficulties in SM9 encryption and decryption. Liu [11] proposed an OSM9 key encapsulation mechanism that supports decryption outsourcing, outsourcing the decryption part of SM9 cryptographic algorithm to the cloud service center for decryption operation, which reduces the computational burden of the terminal equipment, but the mechanism requires the cloud server to generate its own public-private key pairs, which increases the requirements of the system's initialization settings.

However the Snowden incident [12] showed that even

^{*}Corresponding authors.

provably secure cryptographic algorithms can be subject to backdoor attacks that threaten the security and privacy of user data. Mironov et al. [13] proposed the Cryptographic Reverse Firewall (CRF), which is an entity deployed on the user side to re-randomize the information received and sent by users, in order to prevent the leakage of the user's private information. Therefore, CRF can be deployed between the cloud server and the user, and even if the algorithm is tampered with by a backdoor, it will not threaten the security and privacy of the user's data. Therefore, constructing a cryptographic reverse firewall for the SM9 cryptographic algorithm is a very important task. Chen et al. [14] constructed several CRF-based cryptographic protocols by relying on a malleable smooth projective hash function with key malleability and element re-randomization. Zhou et al. [15] proposed an identity-based encryption scheme with CRF. Zhou et al. [16] designed a single-round, certificate-less public key encryption scheme incorporating CRF with reduced communication overhead. Furthermore, Zhou [17] suggested a searchable public key encryption approach based on CRF. Zhou et al. [18] designed an identity-based proxy re-encryption scheme with CRF that can resist leakage attacks. Jin et al. [19] designed a blockchain and CRF-based proxy re-encryption scheme. Xiong et al. [20] designed an SM9 encryption scheme with CRFs and supports equation testing, but the scheme only sets CRFs for data owners. Li et al. [21] designed an online/offline attribute-based encryption scheme with CRFs for IoT communication.

As can be seen from the above, The OSM9 key encapsulation mechanism proposed by Liu et al. [11], although considering outsourced decryption, prolongs user waiting time and does not take into account the threat of information leakage. On the other hand, although Xiong et al. [20] proposed the SM9 algorithm with cryptographic reverse firewall, this algorithm does not support outsourced decryption and does not consider the situation where the key generation center and data users are subjected to backdoor attacks. At present, there is no cryptographic reverse firewall built for outsourced decryption. A new mechanism needs to be proposed to consider the potential threat of backdoor attacks during cloud server outsourcing decryption.

B. Research Contributions

This paper focuses on the power monitoring system based on the SM2 cryptographic system proposed by Chen et al. [1], and constructs an SM9-KEM suitable for power monitoring systems, which not only supports outsourcing decryption but also has the function of CRF. The primary contributions include:

1) *Improve the efficiency of SM9 algorithm:* The bilinear mapping in the decryption operation of SM9 is outsourced to the cloud, and the cloud service center is not required to generate its own public-private key pair. It reduces the computational burden of users and greatly improves the efficiency of the scheme.

2) *Enhanced security of the SM9 algorithm:* Not only has CRF been set up on the data user side to re-randomize ciphertext, but CRF has also been set up on the KGC and data owner sides to re-randomize public parameters and user keys. This enables the OSM9-KEM-CRF proposed in this

paper, which supports outsourced decryption, to maintain its functionality and resist leakage even under backdoor attacks, further improving the security of the scheme.

C. Paper Organization

The remainder of this paper is organized as follows. Section II covers the fundamental concepts related to elliptic curves and reverse firewalls. Section III outlines the system model and the security model of OSM9-KEM-CRF. Section IV details the encapsulation mechanism of OSM9-KEM-CRF along with its security. Section V provides a comparison between our proposed scheme and existing schemes in terms of computational overhead, communication overhead, and energy consumption overhead. The conclusion in Section VI.

II. RELEVANT THEORETICAL FOUNDATIONS

A. Elliptic Curve

For an elliptic curve $\mathbb{E}: y^3 = x^3 + ax + b \pmod{p}$, where $a, b \in \mathbb{F}_p$, $(4a^3 + 27b^2) \pmod{p} \neq 0$, \mathbb{F}_p is a finite field of order prime $p > 3$, let \mathbb{G} be the group over \mathbb{E} , $p \in \mathbb{G}$, q on an elliptic curve, where $p \in \mathbb{G}$, q is the order of \mathbb{G} and O is the infinity point of \mathbb{G} . The operations on the elliptic curve are as follows:

1) *Addition of points:* let $P(x_1, y_1) \in \mathbb{E}$, $Q(x_2, y_2) \in \mathbb{E}$, where $P \neq O, Q \neq O, P \neq -Q$, let $R(x_3, y_3)$ is equal to $P+Q$, then the calculation of R can be expressed as $x_3 = \lambda^2 - x_1 - x_2, y_3 = \lambda(x_1 - x_3) - y_1$, where $\lambda = \begin{cases} \frac{y_2 - y_1}{x_2 - x_1}, P \neq Q \\ \frac{3x_1^2 + a}{2y_1}, P = Q \end{cases}$.

2) *Scalar multiplication:* given a point $P(x, y)$ on an elliptic curve and an integer k , scalar multiplication can be defined as $kP = \sum_{i=1}^k P_i$.

B. Bilinear Mapping

Let $\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T$ be cyclic groups, respectively. Then the bilinear mapping $e: \mathbb{G}_1 \times \mathbb{G}_2 \rightarrow \mathbb{G}_T$ has the following properties:

- 1) *Bilinearity:* for $a, b \in \mathbb{Z}_p, P_1 \in \mathbb{G}_1, P_2 \in \mathbb{G}_2$ there is $e(aP_1, bP_2) = e(P_1, P_2)^{ab}$.
- 2) *Non-degeneracy:* there exist elements $P_1 \in \mathbb{G}_1, P_2 \in \mathbb{G}_2$, such that $e(P_1, P_2) \neq 1$.
- 3) *Computability:* for any elements $P_1 \in \mathbb{G}_1, P_2 \in \mathbb{G}_2$, there exists an efficient polynomial time algorithm to evaluate $e(P_1, P_2)$.

C. Cryptographic Reverse Firewall (CRF)

CRF is a stateful algorithm \mathcal{W} with states and messages as inputs and updated states and messages as outputs. Simply, the state information of \mathcal{W} is not explicitly represented. For participant P and cryptographic reverse firewall \mathcal{W} in the system, $\mathcal{W} \circ P$ is defined as the composed party. If \mathcal{W} is composed of participant P , then we call \mathcal{W} cryptographic reverse firewall P . There are three security requirements for cryptographic reverse firewalls, namely Functionality maintaining, weak security preserving, and weak resistance to exfiltration, as described in [22].

III. FORMAL DEFINITION AND SECURITY MODEL OF OSM9-KEM-CRF

A. OSM9-KEM-CRF System Model

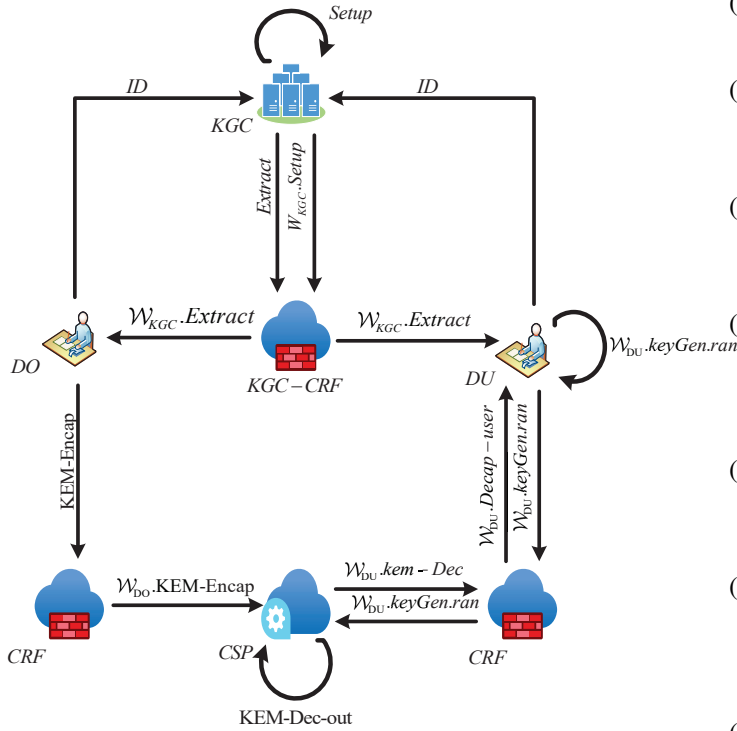


Fig. 1. Illustration of OSM9-KEM-CRF.

The OSM9-KEM-CRF for power monitoring system is shown in Fig. 1, which supports outsourced decryption and CRF and contains four members and three CRFs, that is the cloud service center (CSP), the key generation center (KGC) and its cryptographic reverse firewall \mathcal{W}_{KGC} , the data owner (DO) and its cryptographic reverse firewall \mathcal{W}_{DO} , the data user (DU) and its cryptographic reverse firewall \mathcal{W}_{DU} .

Specifically, KGC generates the master private key and the global public parameter pp . If the process is compromised then \mathcal{W}_{KGC} randomizes pp and broadcasts it globally. The KGC is also responsible for generating the private keys of the users (DO, DU), and if the process is compromised, then \mathcal{W}_{KGC} randomizes the user's private key. The CSP is responsible for storing the user's encrypted data and outsourcing the decryption of the data. The DO encrypts the data and uploads it to the CSP for storage. When the encryption process is compromised then \mathcal{W}_{DO} randomizes the encrypted ciphertext. DU downloads the ciphertext from CSP and decrypts it. If the outsourced decryption key generation process is compromised then \mathcal{W}_{DU} randomizes the outsourced decryption key.

B. OSM9-KEM-CRF System Model

The OSM9-KEM-CRF consists of the following 11 algorithms:

- (1) $\text{Setup}(1^\lambda) \rightarrow (msk, pp)$. The algorithm is run by KGC. Input security parameter λ , output global public

parameter pp and KGC master private key msk .

- (2) $\mathcal{W}_{GA}.\text{Setup}(pp) \rightarrow pp'$. The algorithm is run by KGC's Cryptographic Reverse Firewall \mathcal{W}_{KGC} . Input the system public parameters pp and output the updated system public parameters pp' .
- (3) $\text{Extract}(pp', msk, ID) \rightarrow sk$. The algorithm is run by KGC. Inputs pp', msk and user identity ID and outputs private key sk for user ID .
- (4) $\mathcal{W}_{KGC}.\text{Extract}(sk) \rightarrow sk'$. The algorithm is run by KGC Cryptographic Reverse Firewall \mathcal{W}_{KGC} . It inputs the private key sk of the user ID , outputs the updated sk' , and returns it to the user ID .
- (5) $\text{KEM-Encap}(pp', ID) \rightarrow (K, C_1)$. The algorithm is run by the data owner DO with input pp' and outputs the encapsulated key K and encapsulated ciphertext C_1 .
- (6) $\mathcal{W}_{DO}.\text{KEM-Encap}(K, t, C_1) \rightarrow (K', C'_1)$. The algorithm is run offline by the cryptographic reverse firewall \mathcal{W}_{DO} of the data owner DO. Input (K, C_1) , output updated encapsulated key K' and encapsulated ciphertext C'_1 .
- (7) $\text{KenGen.ran}(sk') \rightarrow (TK, RK)$. The algorithm is run by the user DU, which inputs its own private key sk' and outputs the transformation key TK and retrieval RK .
- (8) $\mathcal{W}_{DC}.\text{TKUpdate}(TK) \rightarrow (TK', \beta)$. The algorithm is run by the password reversal firewall \mathcal{W}_{DU} of the user user DU. Input TK . Output the updated conversion key TK' , keeping the corresponding random number β .
- (9) $\text{KEM-Decap-out}(pp', TK', C'_1) \rightarrow TCT$. The algorithm is run by CSP. Input pp', TK', C'_1 , Output convert ciphertext.
- (10) $\mathcal{W}_{DU}.\text{Decrypt}(TCT, \beta) \rightarrow TCT'$. The algorithm is run by the Cryptographic Reverse Firewall of the data user DU. Input TCT, β and output TCT' .
- (11) $\text{KEM-Decap-user}(pp', RK, TCT') \rightarrow K'$. The algorithm is run by the data user DU. Input pp', TCT', RK , Output updated encapsulated key K' .

Correctness: For security parameters and encapsulated keys, correctness is required for all

- $$\begin{aligned} &\text{Setup}(1^\lambda) \rightarrow (msk, pp), \\ &\mathcal{W}_{GA}.\text{Setup}(pp) \rightarrow pp', \\ &\text{Extract}(pp', msk, ID) \rightarrow sk, \\ &\mathcal{W}_{KGC}.\text{Extract}(sk) \rightarrow sk', \\ &\text{KEM-Encap}(pp', ID) \rightarrow (K, C_1), \\ &\mathcal{W}_{DO}.\text{KEM-Encap}(K, t, C_1) \rightarrow (K', C'_1), \\ &\text{KenGen.ran}(sk') \rightarrow (TK, RK), \\ &\mathcal{W}_{DC}.\text{TKUpdate}(TK) \rightarrow (TK', \beta), \\ &\text{KEM-Decap-out}(pp', TK', C'_1) \rightarrow TCT, \\ &\mathcal{W}_{DU}.\text{Decrypt}(TCT, \beta) \rightarrow TCT' \end{aligned}$$
- satisfy $\text{KEM-Decap-out}(RK, TCT') \rightarrow K'$.

C. Security Model for OSM9-KEM-CRF

Based on the security models of [3] and [22], this paper defines the security model of OSM9-KEM-CRF. In OSM9-KEM-CRF, it is assumed that KGC, DO and DU are fully trusted and the cloud service provider CSP is semi-trusted. Since the algorithms (Setup, Extract, KEM-Encap, KEM-Decap-out, KEM-Decap-user) of OSM9-KEM-CRF remain functional

after the implantation of a malicious trapdoor, it is necessary to take into account that these algorithms can be attacked without the knowledge of the executor. Also considering that \mathcal{W}_{DO} and \mathcal{W}_{DU} would be curious about the user's data, it is assumed that \mathcal{W}_{DO} and \mathcal{W}_{DU} are semi-trustworthy. Since \mathcal{W}_{KGC} can access to the user's decryption key, it is assumed to be fully trusted. In addition, all cryptographic reverse firewalls are considered to be trusted areas and will not be tampered with by any outsiders.

The ID-IND-CCA2 security of the OSM9-KEM-CRF is defined by a game between Challenger \mathcal{C} and Adversary \mathcal{A} . The game is played by the challenger and the adversary.

Initialization. The adversary sends function maintenance algorithm $\text{Setup}^*, \text{Extract}^*, \text{KEM} - \text{Encap}^*, \text{KeyGen.ran}^*, \text{KEM} - \text{Decap} - \text{out}^*$, and $\text{KEM} - \text{Decap} - \text{user}^*$ to the challenger \mathcal{C} .

Setup. Challenger \mathcal{C} runs $\text{Setup}(1^\lambda) \rightarrow (msk, pp)$, $\mathcal{W}_{KGC}.\text{Setup}(pp) \rightarrow pp'$, then sends pp' to adversary \mathcal{A} .

Phase 1. Adversary \mathcal{A} can adaptively query the private key oracle. For the query identity entered by the adversary, the challenger \mathcal{C} runs

$\text{Extract}(pp', msk, ID) \rightarrow sk$,
 $\mathcal{W}_{KGC}.\text{Extract}(sk) \rightarrow sk'$,
 $\text{KenGen.ran}(sk') \rightarrow (TK, RK)$,
 $\mathcal{W}_{DC}.\text{TKUpdate}(TK) \rightarrow (TK', \beta)$,

then returns sk' and TK' to adversary \mathcal{A} . Challenge. Adversary \mathcal{A} sends a challenge identity ID^* to challenger \mathcal{C} . Challenger \mathcal{C} runs $\text{KEM} - \text{Encap}(pp', ID^*) \rightarrow (K_0, C_1^*)$, $\mathcal{W}_{DO}.\text{KEM} - \text{Encap}(K_0, C_1^*) \rightarrow (K'_0, C'^*_1)$ and then randomly selects a key K'_1 in the key space, bit $b \leftarrow \{0, 1\}$, and then sends (K'_b, C'^*_1) to adversary \mathcal{A} .

Phase 2. As in Phase 1, adversary \mathcal{A} can adaptively query the private key of the user, but not the private key of user ID^* . Additionally adversary \mathcal{A} can adaptively query the decapsulation oracle. For the (ID, C) inputted by adversary, the challenger runs $\text{KEM} - \text{Decap} - \text{out}(pp', TK', C'_1) \rightarrow TCT$, $\mathcal{W}_{DU}.\text{Decrypt}(TCT, \beta) \rightarrow TCT'$, $\text{KEM} - \text{Decap} - \text{user}(RK, TCT') \rightarrow K'$, returns the corresponding decapsulation key K' . but at this point the adversary cannot access the decapsulation key for (ID^*, C'^*_1) .

Guess. Adversary \mathcal{A} outputs a guess $b' \in \{0, 1\}$ to send to challenger \mathcal{C} .

DEFINITION: OSM9-KEM-CRF is said to be ID-IND-CCA2-secure if for all probabilities polynomial time adversary \mathcal{A} has a negligible advantage of $\varepsilon = |\Pr[b = b'] - \frac{1}{2}| \leq \text{negl}(\lambda)$ in winning the above game.

IV. OSM9-KEM-CRF ENCAPSULATION MECHANISMS

A. Description of OSM9-KEM Mechanism

OSM9-KEM consists of the following six algorithms:

- (1) $\text{Setup}(1^\lambda)$. Input the security parameter λ , the algorithm performs the following operations.
 - ① Choose 3 groups $\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T$ of order prime r , a bilinear mapping $e : \mathbb{G}_1 \times \mathbb{G}_2 \rightarrow \mathbb{G}_T$, and randomly choose generators $P_1 \in \mathbb{G}_1, P_2 \in \mathbb{G}_2$.

- ② Randomly select $s \leftarrow \mathbb{Z}_r^*$ and compute $P_{pub} = sP_1$.

- ③ Make $g = e(P_{pub}, P_2)$.

- ④ Choose hash function $H_v : \{0, 1\}^* \rightarrow \{0, 1\}^v$ and an identifier hid .

- ⑤ Output global public parameters $pp = (\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, e, P_1, P_2, P_{pub}, g, H_v, hid)$ and master private key $msk = s$.

- (2) $\text{Extract}(pp, msk, ID)$. Input user identities $ID \in \{0, 1\}^*$, pp and msk , KGC calculates $t_1 = H_v(ID || hid, r) + s$, if $t_1 = 0$, recalculates the master private key, otherwise calculates $sk = t_2 P_2$, where $t_2 = st_1^{-1}$.

- (3) $\text{KEM} \rightarrow \text{Encap}(pp, ID)$. With inputs pp and ID , the algorithm performs the following.
 - ① Let $t_1 = H_v(ID || hid, r) + s, Q = h_1 P_1 + P_{pub} = (h_1 + s)P_1$.

- ② Random select $x \leftarrow \mathbb{Z}_r^*$, let $C_1 = xQ, t = g^x$.

- ③ Let $K = \text{KDF}_2(H_v, \text{EC2OSP}(C_1) || \text{FE2OSP}(t) || ID, l)$, where l is the key length of DEM .

- ④ Output (K, C_1) .

- (4) $\text{KenGen.ran}(sk) \rightarrow (TK, RK)$. Input sk . Randomly select $\alpha \leftarrow \mathbb{Z}_r^*$, compute $TK = \frac{1}{\alpha} sk = \frac{t_2}{\alpha} P_2$, and output conversion key TK and retrieval key $RK = \alpha$.

- (5) $\text{KEM} - \text{Decap} - \text{out}(pp, TK, C_1)$. Input pp' , the user identity ID and its conversion key TK , the encapsulation portion C_1 . The cloud service center computes the conversion ciphertext TCT , where $TCT = e(C_1, TK) = e(xQ, \frac{t_2}{\alpha} P_2) = e(s(h_1 P_1 + sP_1), \frac{s}{t_1 \alpha} P_2) = e(P_{pub}, P_2)^{\frac{s}{\alpha}} = g^{\frac{s}{\alpha}}$.

- (6) $\text{KEM} - \text{Decap} - \text{user}(pp, RK, TCT)$. Input pp , retrieval key RK for user identity ID , transformed ciphertext TCT , user ID computes $t = (TCT)^\alpha = g^x$, and lets $K = \text{KDF}_2(H_v, \text{EC2OSP}(C_1) || \text{FE2OSP}(t) || ID, l)$, where l is the key length of DEM . Output the encapsulated key K .

Theorem 1 If SM9-KEM is ID-IND-CCA2 secure, then the above OSM9-KEM is ID-IND-CCA2 secure.

Proof In this section, OSM9-KEM is constructed based on SM9-KEM by utilizing the key blinding technique of [23]. From [3], it is known that SM9-KEM is ID-IND-CCA2 secure. Thus it can be proved similarly to [23] that OSM9-KEM is ID-IND-CCA2 secure.

B. Description of OSM9-KEM-CRF Mechanism

Based on the above OSM9-KEM mechanism, this section constructs an OSM9-KEM-CRF mechanism.

After KGC runs $\text{Setup}(1^\lambda)$ to generate msk and pp , KGC first sends pp to \mathcal{W}_{KGC} . \mathcal{W}_{KGC} Run Algorithm $\mathcal{W}_{GA}.\text{Setup}$.

(1) $\mathcal{W}_{GA}.\text{Setup}(pp) \rightarrow pp'$. For pp , \mathcal{W}_{KGC} randomly selects $a, b, c \leftarrow \mathbb{Z}_r^*$ and computes $P'_1 = aP_1, P'_2 = aP_2, P'_{pub} = aP_{pub} = sP'_1, g' = e(P_{pub}, P_2)^{abc} = e(P'_{pub}, P_2)^c$. Output $pp' = (\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T, e, P'_1, P'_2, P'_{pub}, g', H_v, hid)$ but keep c . KGC carries out $\text{Extract}(pp', msk, ID) \rightarrow sk$ after receiving pp' and user identity $\text{Extract}(pp', msk, ID) \rightarrow sk$, sends sk to \mathcal{W}_{KGC} . \mathcal{W}_{KGC} runs algorithm $\mathcal{W}_{KGC}.\text{Extract}$.

(2) $\mathcal{W}_{KGC}.Extract(sk) \rightarrow sk'$. For sk , \mathcal{W}_{KGC} computes $sk' = c \cdot sk = \frac{cs}{s+h} P'_x$ by the previous random selected c , where $h_1 = H_v(ID || hid, r)$.

User ID runs $KEM - Encap(pp', ID) \rightarrow (K, C_1)$ after receiving pp' , sends (K, C_1) to \mathcal{W}_{DO} , \mathcal{W}_{DO} runs algorithm $\mathcal{W}_{DO}.KEM - Encap$.

(3) $\mathcal{W}_{DO}.KEM - Encap(K, t, C_1) \rightarrow (K', C'_1)$. For K, t and C_1 . \mathcal{W}_{DO} randomly selects $f \leftarrow \mathbb{Z}_r^*$ and compute $C'_1 = fC_1$, $t' = t^f - g^{fx}$ and $K' = KDF_2(H_v, EC2OSP(C'_1) || FE2OSP(t') || ID, l)$, where l is the key length of DEM . Output (K', C'_1) .

The user sends TK to \mathcal{W}_{DO} after running $KenGen.ran(sk') \rightarrow (TK, RK)$, and \mathcal{W}_{DO} runs Algorithm $\mathcal{W}_{DO}.TKUpdate$.

(4) $\mathcal{W}_{DC}.TKUpdate(TK) \rightarrow (TK', \beta)$. For TK , \mathcal{W}_{DO} randomly selects $\beta \leftarrow \mathbb{Z}_r^*$, computes $TK' = \frac{1}{\beta}TK$, and outputs TK' but keeps β .

The cloud service center runs $KEM - Decap - out(pp', T, K', C'_1) \rightarrow TCT$ after receiving TK' , sends TCT to \mathcal{W}_{DU} . \mathcal{W}_{DU} runs algorithm $\mathcal{W}_{DU}.Decrypt$.

(5) $\mathcal{W}_{DU}.Decrypt(TCT, \beta) \rightarrow TCT'$. For input TCT and reserved β , \mathcal{W}_{DU} computes $TCT' = (TCT)^\beta$.

After receiving TCT' , DU runs $KEM - Decap - user(pp', RK, TCT')$, gets $t' = (g^{\frac{fx}{\alpha}})^\alpha = g^{fx}$ and $K' = KDF_2(H_v, EC2OSP(C'_1) || FE2OSP(t') || ID, l)$.

C. Security Analysis

Theorem 2: If OSM9-KEM is ID-IND-CCA2 secure, then OSM9-KEM-CRF is ID-IND-CCA2 secure and the cryptographic reverse firewalls of KGC, DO, and DU maintain functionality, weakly retain security, and weakly resist penetration.

Proof the security of OSM9-KEM-CRF is proved by the following three sections.

(1) Functionality-maintaining. Because

$$\begin{aligned} TCT &= (TCT)^\beta = e(C'_1, TK')^\beta \\ &= e\left(fx(h_1 + s)P'_1, \frac{1}{\beta}TK\right)^\beta \\ &= e\left(fx(h_1 + s)P'_1, \frac{cs}{\alpha t_1}P'_2\right) \\ &= e\left(fxP'_1, \frac{cs}{\alpha}P'_2\right) = e(P'_{pub}, \frac{cs}{\alpha}P'_2) \\ &= g^{\frac{fx}{\alpha}} \end{aligned}$$

Thus the data user, after receiving

$K' = KDF_2(H_v, EC2OSP(C'_1) || FE2OSP(t') || ID, l)$, runs $K' = KDF_2(H_v, EC2OSP(C'_1) || FE2OSP(t') || ID, l)$ and can calculate $K' = KDF_2(H_v, EC2OSP(C'_1) || FE2OSP(t') || ID, l)$ which in turn yields the encapsulation key $K' = KDF_2(H_v, EC2OSP(C'_1) || FE2OSP(t') || ID, l)$. The encapsulated key is then obtained. Thus the mechanism satisfies the maintenance functionality.

(2) ID-IND-CCA2 Security. For any tampering algorithms $Setup^*$, $Extract^*$, $KEM - Encap^*$, $KeyGen.ran^*$, $KEM - Decap - out^*$ and $KEM - Decap - user^*$ on KGCs, DOs and DUs that maintain

functionality, we prove that OSM9-KEM-CRF is ID-IND-CCA2 secure by the indistinguishability of the secure game of OSM9-KEM from the secure game of OSM9-KEM-CRF. Consider the following game.

Game0. The security game same as OSM9-KEM-CRF in Section 3.3.

Game1. Same as Game0 except that pp and msk in the Setup phase are generated by the algorithm of OSM9-KEM instead of $Setup^*$ and $\mathcal{W}_{KGC}.Setup$.

Game2. Same as Game1 except that sk and TK in Phase 1 and Phase 2 are generated by the Extract and KeyGen.ran algorithms of OSM9-KEM, not by $Extract^*$, $\mathcal{W}_{KGC}.Extract$, $KeyGen.ran^*$ and $\mathcal{W}_{DC}.TKUpdate$.

Game3. It is the same as Game2 except that the challenge key ciphertext pair (K'_b, C'^*_1) in the Challenge phase is generated by $KEM - Encap$, not by $KEM - Encap^*$ and $\mathcal{W}_{DO}.KEM - Encap$.

It can be seen that Game3 is a secure game for OSM9-KEM, so it is only necessary to prove that Game0 is indistinguishable from Game3 to prove the security of OSM9-KEM-CRF.

In fact, since a, b, c is randomly chosen in Algorithm $\mathcal{W}_{KGC}.Setup$, regardless of the distribution of pp generated by $Setup^*$, the pp obtained after the processing of the reverse firewall $\mathcal{W}_{KGC}.Setup$ is uniformly random and consistent with the distribution of pp generated by Setup. Thus Game0 is indistinguishable from Game1. Also due to the extensibility of the key, it is similarly known that Game1 is indistinguishable from Game2.

For the challenge key ciphertext pair (K'_b, C'^*_1) , the distribution is randomized since K'_b is generated by KDF_2 . For C'^*_1 , even though C'^*_1 generated by $KEM - Encap^*$ is not random, since f is randomly selected in $\mathcal{W}_{DO}.KEM - Encap$, C'^*_1 after $\mathcal{W}_{DO}.KEM - Encap$ post-processing is random, which is consistent with the distribution of the ciphertext generated by $KEM - Encap$, thus the indistinguishability of Game2 from Game3 can be obtained. From Game0 and Game1, Game1 and Game2, and Game2 and Game3 are indistinguishable respectively, it can be known that Game0 and Game3 are indistinguishable.

(3) Weak Security Preserving, weak Resistant to Exfiltration. According to the ID-IND-CCA2 security of OSM9-KEM-CRF, it is shown that the cryptographic reverse firewalls \mathcal{W}_{KGC} , \mathcal{W}_{DU} , and \mathcal{W}_{DO} of KGC, DU, and DO are weakly preserve security. Also the proof of ID-IND-CCA2 security of OSM9-KEM-CRF shows that \mathcal{W}_{KGC} , \mathcal{W}_{DU} and \mathcal{W}_{DO} are weakly resistant to exfiltration.

V. COMPARATIVE ANALYSIS

In order to ensure the same security strength, the traditional RSA encryption algorithm requires a larger number of key bits than the elliptic curve cipher, resulting in longer encryption time and lower monitoring efficiency in power information systems. Chen et al. [1] proposed a power information system monitoring scheme based on the SM2 algorithm, in which an SM2 encryption component is connected to the server interface, which not only determines the user's access to resources

but also records information about user activities. When the user inserts the SM2 encryption device into the client, the client uses the HTTP protocol and the digital certificate to log in to the server, and then starts to access the server. When accessing the server, the system verifies the certificate by calling the suite “iaccount”, and if the verification is unsuccessful, the client’s “imidware” will be automatically directed to the security support platform, which supports validation of SM2 digital certificates. The certificate is generated after verifying SM2 digital certificates, and the user’s information is sent to the client, the client is redirected to the standby power supply system again through the “imidware”, and then returns to the electric power secondary system by submitting a one-time signature certificate and a one-time authorization code verification and destroys the one-time certificate, and decrypts the user information by verifying the authenticity of the user signature information, and finally logs in. The user information is decrypted and finally logged into the power system.

In the above scheme, although the SM2 encryption algorithm has advantages over the RSA algorithm, however, it has some limitations in practical applications.

- (1) Before using SM2 for encryption, the public key certificate of the other party must be obtained, otherwise the encryption operation cannot be performed. This requirement increases the complexity of certificate management, which in turn increases the management overhead of the overall power system.
- (2) In terms of decryption, the SM2 algorithm has some complexity when decrypting on the Web side.
- (3) When communicating securely across domains, it is necessary to establish a chain of trust for certificates.

Unlike SM2, SM9 is an identity-based encryption algorithm with the following advantages:

- (1) No certificate management is required, effectively solving the complexity of certificate management in SM2 and significantly reducing the management burden of public key infrastructure (PKI).
- (2) When decrypting on the web, there is no need for pre-registration.
- (3) It only requires the publication of security parameters without a chain of trust for certificates, and the user’s identity is his/her public key.

Therefore replacing the SM9 encryption algorithm with the SM2 encryption algorithm proposed by Chen et al. [1] for the power monitoring scheme not only enables more efficient data encryption, but also reduces the management cost of the power information system.

An outsourcing decryption is introduced on the basis of SM9 algorithm to further improve the decryption efficiency of SM9 algorithm in this paper. In addition, both SM2 and SM9 encryption algorithms have backdoor attacks, so this paper introduces cryptographic reverse firewall into SM9 algorithm to improve the security. Therefore the proposed OSM9-KEM-CRF based on SM9 key encapsulation is more suitable for power monitoring system.

In this section, the proposed OSM9-KEM-CRF is compared with other schemes in terms of computational overhead, communication overhead and energy consumption overhead.

TABLE I. EXECUTION TIME OF DIFFERENT CRYPTOGRAPHIC PRIMITIVES

Symbol	Operation	Times(ms)
T_{pa}	Bilinear-Pairing	13.8196
T_{pm}	ECC Point Addition	0.0110
T_e	ECC Point Exponent	12.2007
T_m	ECC Point Multiplication	2.2001
T_h	Time of hash function	0.4702

TABLE II. COMPUTATION OVERHEAD COMPARISON

Scheme	Computational overhead
[3]	$T_{pa} + T_h \approx 14.2898ms$
[4]	$2T_{pa} + T_h \approx 28.1049ms$
[11]	$T_e + T_h \approx 12.6709ms$
[20]	$T_{pa} + T_h \approx 14.2898ms$
Ours	$T_e + T_h \approx 12.6709ms$

A. Computation Cost Comparison

To evaluate the performance of our OSM9-KEM-CRF mechanism, we consistently used the Python programming language to test decryption operation times, employing 256-bit Barreto-Naehrig (BN) elliptic curves and R-ate bilinear pairings. The specific test setup was a personal desktop computer with the following configurations: 32GB of RAM, Windows 10 operating system (version 10.0.19045), Intel Cor i5-13400 CPU running at 2.5GHz, Visual Studio Code as the development environment, and the Charm cryptographic library. The notation for the operation times of cryptographic algorithms is defined in Table I.

The computational overhead of the decryption phase of each mechanism (scheme) are shown in Table II. In literature [3], one hash operation and bilinear pairing operation need to be run, and the time required is $T_{pa} + T_h \approx 14.2898ms$. In literature [4], one hash operation and two bilinear pairing operations need to be run, and the time required is $2T_{pa} + T_h \approx 28.1049ms$. In literature [11], one exponentiation operation and hash operation need to be performed, and the time required is $T_e + T_h \approx 12.6709ms$. However, in cloud services, the cloud is required to generate its own public-private key pairs, which increases the cloud’s computational overhead. In the proposed OSM9 mechanism, the decryption phase needs to perform one exponential operation and one hash operation on the multiplicative group, and the total time required is $T_e + T_h \approx 12.6709ms$. In the literature [20], it needs to perform one bilinear pairing operation and one hash operation, and the time required is $T_{pa} + T_h \approx 14.2898ms$.

The comparison of the time consumed in the decryption phase of each mechanism (scheme) is shown in Fig. 2, which shows that the time consumed in decryption of this paper’s mechanism and the scheme of literature [11] is lower than other schemes, and this paper’s scheme does not need to generate public-private key pairs in the cloud server, which reduces the time of the cloud computation and the waiting time of the user, compared to the scheme of literature [11].

Fig. 3 and 4 show the time overhead of each algorithm in the OSM9-KEM and OSM9-KEM-CRF mechanisms, respectively, from which it is clear that the addition of the Cryptographic Reverse Firewall to the OSM9-KEM mechanism does

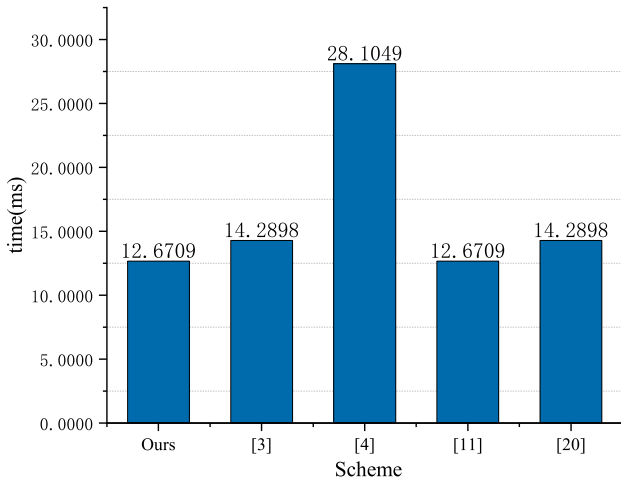


Fig. 2. Comparison of decryption time cost for users in different mechanisms (schemes).

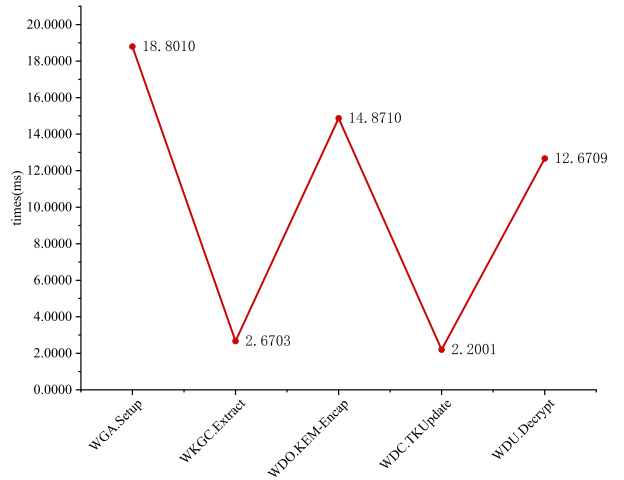


Fig. 4. OSM9-KEM-CRF algorithm time overhead

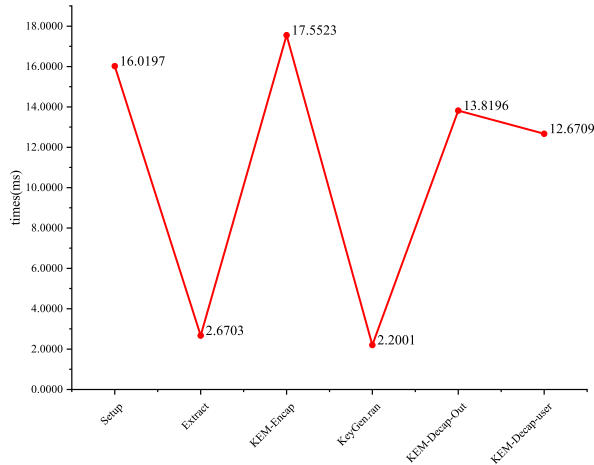


Fig. 3. The algorithm time cost used in OSM9-KEM.

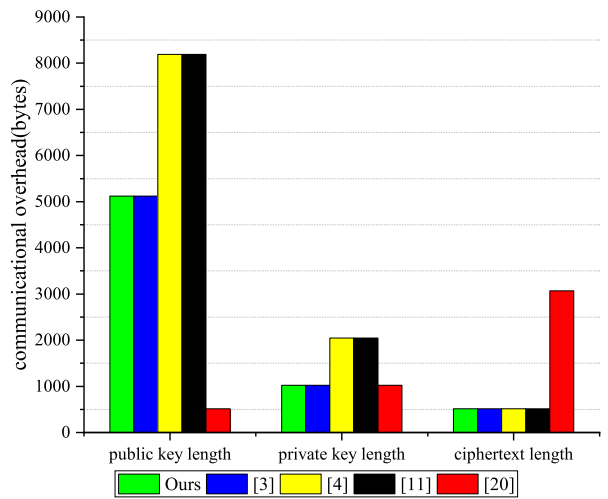


Fig. 5. Comparison of communication costs for different schemes.

not have a significant impact on the time overhead, but greatly increases the security.

B. Communication Cost Comparison

In terms of communication overhead, $|\mathbb{G}_1|, |\mathbb{G}_2|, |\mathbb{G}_T|, |\mathbb{Z}_p|$ denote the size of the elements in the $\mathbb{G}_1, \mathbb{G}_2, \mathbb{G}_T$ and \mathbb{Z}_p , respectively. Specifically, the 256-bit BN curve [24] is used, that is $|\mathbb{G}_1|=512\text{bit}, |\mathbb{G}_2|=1024\text{bit}, |\mathbb{G}_T|=3072\text{bit}, |\mathbb{Z}_p|=256\text{bit}$. Table III compares the bit requirements of the key encapsulation mechanism proposed in this paper with those of other schemes across public parameters, private keys, and ciphertexts. Furthermore, as illustrated in Fig. 5, our mechanism demonstrates a significant reduction in communication overhead for public keys, private keys, and ciphertexts.

C. Energy Cost Comparison

In terms of energy overhead, the calculation method in [25] is used, with the formula $vol \times cur \times T$, where vol represents the voltage, cur represents the current, T represents the execution time ($vol = 3V, cur = 1.8\mu A$), and the energy consumed for sending 1bit messages is $0.72\mu J$, and the energy consumed for receiving messages is $0.81\mu J$. In literature [3], the energy overhead related to computation is $vol \times cur \times (T_{pa} + T_h) \approx 77.1649\mu J$, and the energy overhead related to communication is $|\mathbb{G}_2| \times 0.81\mu J + |\mathbb{G}_1| \times 0.72\mu J \approx 1198.0800\mu J$, so the total energy overhead is $1275.2449\mu J$; in literature [4], the energy overhead related to computation is $vol \times cur \times (2T_{pa} + T_h) \approx 151.7664\mu J$, and the energy overhead related to communication is $2|\mathbb{G}_2| \times 0.81\mu J + |\mathbb{G}_1| \times 0.72\mu J \approx 2027.5200\mu J$, so the total energy overhead is $2179.2864\mu J$; in literature [11], the energy overhead related to computation

TABLE III. COMMUNICATION OVERHEAD OF DIFFERENT SCHEMES

Scheme	Public parameter length	Private key length	Ciphertext length
Ours	$2 \mathbb{G}_1 + \mathbb{G}_2 + \mathbb{G}_T \approx 5120\text{bits}$	$ \mathbb{G}_2 \approx 1024\text{bits}$	$ \mathbb{G}_1 \approx 512\text{bits}$
[3]	$2 \mathbb{G}_1 + \mathbb{G}_2 + \mathbb{G}_T \approx 5120\text{bytes}$	$ \mathbb{G}_2 \approx 1024\text{bytes}$	$ \mathbb{G}_1 \approx 512\text{bits}$
[4]	$2 \mathbb{G}_1 + \mathbb{G}_2 + 2 \mathbb{G}_T \approx 8192\text{bits}$	$2 \mathbb{G}_2 \approx 2048\text{bits}$	$ \mathbb{G}_1 \approx 512\text{bits}$
[11]	$2 \mathbb{G}_1 + \mathbb{G}_2 + 2 \mathbb{G}_T \approx 8192\text{bits}$	$2 \mathbb{G}_2 \approx 2048\text{bits}$	$ \mathbb{G}_1 \approx 512\text{bits}$
[20]	$ \mathbb{G}_1 \approx 512\text{bits}$	$ \mathbb{G}_2 \approx 1024\text{bits}$	$3 \mathbb{G}_1 + \mathbb{G}_2 + 2 \mathbb{Z}_p \approx 3072\text{bits}$

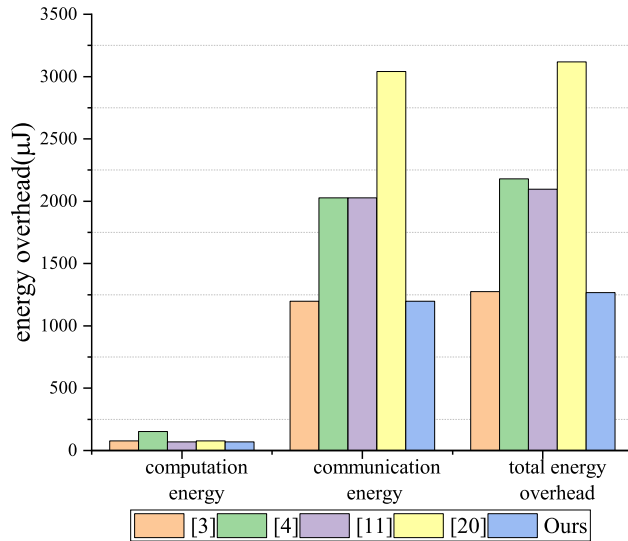


Fig. 6. Comparison of energy consumption of different schemes.

is $vol \times cur \times (T_e + T_h) \approx 68.4228\mu\text{J}$, and the energy overhead related to communication is $2|\mathbb{G}_2| \times 0.81\mu\text{J} + |\mathbb{G}_1| \times 0.72\mu\text{J} \approx 2027.5200\mu\text{J}$, so the total energy overhead is $2095.9428\mu\text{J}$; in literature [20], the computation-related energy overhead is $vol \times cur \times (T_{pa} + T_h) \approx 77.1649\mu\text{J}$ and the communication-related energy overhead is $|\mathbb{G}_2| \times 0.81\mu\text{J} + (3|\mathbb{G}_1| + |\mathbb{G}_2| + 2|\mathbb{Z}_p|) \times 0.72\mu\text{J} \approx 3041.28\mu\text{J}$, thus the total energy overhead is $3118.4449\mu\text{J}$; in this paper, the computation-related energy overhead is $vol \times cur \times (T_e + T_h) \approx 68.4228\mu\text{J}$ and the communication-related energy overhead is $|\mathbb{G}_2| \times 0.81 + |\mathbb{G}_1| \times 0.72 \approx 1198.0800\mu\text{J}$, thus the total energy overhead is $1266.5028\mu\text{J}$. The comparison of energy overhead of each mechanism (scheme) is shown in Fig. 6. In power monitoring system, less energy overhead is especially important in power system due to limited equipment resources, in the above comparison, the mechanism in this paper has less energy overhead and is more suitable for power system, and it incorporates a reverse firewall to block backdoor attacks and improve the security of the system.

VI. CONCLUSIONS

This paper proposes an SM9 key encapsulation mechanism that supports outsourced decryption and CRF, improving the efficiency and security of the SM9 key encapsulation mechanism. The proposed OSM9-KEM-CRF mechanism outsources the tedious bilinear mapping calculation in the decryption process to cloud servers, and cloud servers do not need

to generate its own public-private key pairs, improving the efficiency of the mechanism. In addition, the key encapsulation mechanism adds the cryptographic reverse firewall function for KGC and users respectively, and the deployment of CRF also makes the mechanism resistant to backdoor attacks, resistant to information leakage, protects user privacy, and improves the security of the key encapsulation mechanism. The security proof and comparative analysis comparison show that the mechanism is more suitable for the power monitoring system.

In future work, in order to further reduce the computational burden on users and enrich the functionality of the SM9 algorithm, we will research how to use smart contracts to verify the correctness of outsourced decryption, thereby further reducing users' computational overhead. In addition, the SM9 algorithm will be functionally extended to construct an attribute based encryption scheme based on SM9, achieving fine-grained access control of encrypted data in the cloud.

ACKNOWLEDGMENT

This research was funded by Science and Technology of Yunnan Power Grid (YNKJXM20222419, YNKJXM20222425).

REFERENCES

- [1] F. Chen, H. Zou, Y. Wu, X. Liu, D. Liang, Design of power information security monitoring system based on SM2 cryptosystem, *Electronic Design Engineering* 30 (05) (2022) 100-103+108.
- [2] F. Yuan; Z.H. Cheng, Review of SM9 identity cipher algorithm, *Information Security Research*, 2 (11) (2016) 1008-1027.
- [3] Z. Chen, Security analysis of SM9 key agreement and encryption, in: *Information Security and Cryptology: 14th International Conference, Inscrypt 2018, Fuzhou, China, December 14-17, 2018, Revised Selected Papers*, Proceedings, Springer, 2019, pp. 3-25.
- [4] J. Lai, X. Huang, D. He; W. Wu, Security analysis of state secret SM9 digital signature and key encapsulation algorithm, *Science China: Information Science*, 51 (11) (2021) 1900-1913.
- [5] H. Ji, H. Zhang, L. Shao, D. He, M. Luo, An efficient attribute-based encryption scheme based on SM9 encryption algorithm for dispatching and control cloud, *Connection Science*, 33 (04) (2021) 1094-1115.
- [6] M.D. Wang, W.G He, J. Li, R. M, Optimized design of state-secret SM9 algorithm R-ate pair computation, *Communication Technology*, 53 (11) (2020) 2241-2244.
- [7] J. Lai, Y. Mu, F. Guo, Efficient identity-based online/offline encryption and signcryption with short ciphertext, *International Journal of Information Security*, 16 (2017) 299-311.
- [8] Y. Sun, Z. Lu, J. Zhao, M.Q. Fan, Research on SM9-IBE encryption scheme based on online/offline technology, *Journal of Qiqihar University (Natural Science Edition)*, 39 (01) (2023) 26-30.
- [9] C. Peng, M. Luo, L. Li, K.K.R. Choo, D. He, Efficient certificateless online/offline signature scheme for wireless body area networks, *IEEE Internet of Things Journal*, 8 (18) (2021) 14287-14298.
- [10] P. Liu, Q. He, W.Y. Liu, X. Cheng, A CP-ABE scheme supporting revocation of attributes and outsourced decryption, *Information Network Security*, 20 (03) (2020) 90-97.

- [11] Liu, K. An OSM9 identity key encapsulation mechanism supporting decryption outsourcing, *Industrial Technology Innovation*, 10 (01) (2023) 106-113.
- [12] C. Patsakis, A. Charemis, A. Papageorgiou, D. Mermigas, S. Pirounias, The market's response toward privacy and mass surveillance: The Snowden aftermath, *Computers Security*, 73 (2018) 194-206.
- [13] I. Mironov, N. S. Davidowitz, Cryptographic reverse firewalls, in: *Advances in Cryptology – EUROCRYPT 2015: 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Sofia, Bulgaria, April 26-30, 2015, Proceedings, Part II, Proceedings, Springer, 2015, pp. 657-686.
- [14] R. Chen, Y. Mu, G. Yang, W. Susilo, F. Guo, M. Zhang, Cryptographic reverse firewall via malleable smooth projective hash functions, in: *Advances in Cryptology – ASIACRYPT 2016: 22nd International Conference on the Theory and Application of Cryptology and Information Security*, Hanoi, Vietnam, December 4-8, 2016, Proceedings, Part I, Proceedings, Springer, 2016, pp. 844-876.
- [15] Y. Zhou, Y. Guan, Z. Zhang, F. Li, Cryptographic reverse firewalls for identity-based encryption, in: *Frontiers in Cyber Security: Second International Conference, FCS 2019, Xi'an, China, November 15–17, 2019*, Proceedings, Springer, 2019, pp. 36-52.
- [16] Y. Zhou, J. Guo, F. Li, Certificateless public key encryption with cryptographic reverse firewalls, *Journal of Systems Architecture*, 109 (2020) 101754.
- [17] Y. Zhou, Z. Hu, F. Li, Searchable public-key encryption with cryptographic reverse firewalls for cloud storage, *IEEE Transactions on Cloud Computing*, 11 (01) (2021) 383-396.
- [18] Y. Zhou, L. Zhao, Y. Jin, F. Li, Backdoor-resistant identity-based proxy re-encryption for cloud-assisted wireless body area networks, *Information Sciences*, 604 (2022) 80-96.
- [19] C. Jin, Z. Chen, W. Qin, K. Sun, G. Chen, L. Chen, A Blockchain-Based Proxy Re-Encryption Scheme with Cryptographic Reverse Firewall for IoV, *International Journal of Network Management*, 34 (2024) 80-96.
- [20] H. Xiong, Y. Lin, T. Yao, An SM9 logo encryption scheme supporting equation testing and cryptographic reverse firewall, *Computer Research and Development*, 61 (04) (2024) 1070-1084.
- [21] J. Li, Y. Fan, X. Bian, Q. Yuan, Online/Offline MA-CP-ABE with Cryptographic Reverse Firewalls for IoT, *Entropy*, 25 (4) (2023) 616.
- [22] M.H. Ma, R. Zhang, G. Yang, Z. Song, S. Sun, Y. Xiao, Concessive online/offline attribute based encryption with cryptographic reverse firewalls—Secure and efficient fine-grained access control on corrupted machines, in: *Computer Security: 23rd European Symposium on Research in Computer Security, ESORICS 2018, Barcelona, Spain, September 3-7, 2018*, Proceedings, Part II, Proceedings, Springer, 2018, pp. 507-526.
- [23] M. Green, S. Hohenberger, B. Waters, Outsourcing the decryption of abe ciphertexts, In: *Proceedings of the 20th USENIX Conference on Security (USENIX'11)*, USENIX Association, 2011, pp.1–11.
- [24] G.C.C.F Pereira, Jr.M.A Simplício, M. Naehrig, P.S.L.M. Barreto, A family of implementation-friendly BN elliptic curves, *Journal of Systems and Software*, 84 (08) (2011) 1319-1326.
- [25] J. Du, C. Dai, P. Mao, W. Dong, X. Wang, Z. Li, An Efficient Lightweight Authentication Scheme for Smart Meter, *Mathematics*, 12 (8) (2024) 1264.

A Review of Analyzing Different Agricultural Crop Yields Using Artificial Intelligence

Vijaya Bathini¹, K. Usha Rani²

Research Scholar, Department of Computer Science, SPMVV (Women's University), Tirupati, AP, India¹

Professor, Department of Computer Science, SPMVV (Women's University), Tirupati, AP, India²

Abstract—The advancement of Artificial Intelligence (AI), in particular Deep Learning (DL), has made it possible to interpret gathered data more quickly and effectively in this new digital era. To draw attention to development advancements in deep learning across many industries. Agriculture has been one of the most affected areas in recent advancements of the current globalized world agriculture plays a vital role and makes significant contributions. Over the years, agriculture has faced several difficulties in meeting the growing demands of the global people, which has creased over the last 50 years. Different forecasts have been made regarding this extraordinary population expansion which is expected to grasp almost 9 billion persons worldwide by 2050. More than a century ago, different technologies were brought into agriculture to solve issues related to crop cultivation. Many mechanical technologies are accessible today, and they are evolving at an amazing rate. To support their demands and help them optimize their crop yields based on data and task automation need innovative techniques to aid farmers. This will transform the agricultural industry into a new dimension. Therefore, this study's primary goal was to present a thorough summary of the most current developments based on research interconnected with the digitization of agriculture for crop yields including fruit counting, crop management, water management, weed identification, soil management, seed categorization, disease detection, yield forecasting and harvesting of yields based on Artificial Intelligence Techniques.

Keywords—Agriculture; artificial intelligence; deep learning; crop yields; management

I. INTRODUCTION

Because of the population, the agriculture sector has to meet a wide range of food needs along with social, environmental and economic factors like the lack of workers, water, biodiversity, and land degradation [1]. Since the seasons are hard to predict and the environment is harsh, there are now a number of limits on its growth. For agricultural business growth, it is important to find new methods that will last.

Farmers' understanding of field management has changed by using cutting-edge technologies like robots, drones and sensors on farm equipment. Scientists who study data and farming are getting ideas from these new technologies to make better analytical tools and methods for managing fields and dealing with problems more correctly [2]. Today's technology makes it hard to make sure that everyone has access to a steady supply of high-quality food without putting natural environments at risk. To meet and support farmers' needs help to get the most out of their farming by automating tasks and data.

New developments in uses based on Artificial Intelligence (AI) had a big effect in this area [3]. They have made a big

difference in the progress of computer vision, ML(Machine Learning) and DL(Deep Learning) methods for building automated and reliable systems. But Agriculturalists still confront formidable challenges in making affordable, scalable, and ecologically sound solutions to the world's food crisis a reality, despite recent advances. This emphasizes the significance of studies that cover both the theoretical and practical aspects of incorporating technological advances into actual agricultural systems.

As a result, this study main goal is to give an in-depth overview on the latest advances in AI research that has to do with digitizing agriculture for crop yields. To identify existing gaps in the current review of Digitizing agriculture includes fruit counting, crop management, water management, weed identification, soil management, seed categorization, disease detection, yield forecasting, and harvesting of yields.

II. DIGITIZING AGRICULTURE CROP YIELDS USING AI TECHNIQUES

A. Fruit Counting

A vital component of the world economy is the fruit business. Food security, economic growth, nutritional diversity, processing, shipping and retail are just a few companies that benefit from it, and millions of farmers rely on it for income. Fruits have high vitamin, mineral, fiber and antioxidant content and plays a vital part in healthy diet. As per the FAOSTAT report fruit industry is in the rise worldwide. Producing approximately 909.644 million metric tons of fruits in 2023, the world continued its growing trend in food output, accounting for 19% of total food production. Maintaining accuracy and efficiency in huge fields or orchards becomes increasingly challenging when the volume of agriculture increases, rendering manual counting impracticable defined by Pathan and Rehman [4]. Manual fruit counting can be challenging in outdoor settings due to weather factors including rain and low vision as explained by Hunt and Doraiswamy [5]. The spatial coverage of manual counting is limited since people can't physically inspect every portion of a crop. It also makes coping with different crop architectures more difficult as it makes it harder to address differences in fruit size, shape, or distribution. The consistency and comparability of the data could be compromised due to inconsistent counting techniques caused by the absence of established counting standards.

Fruits calculating or counting flower thickness on images using Computer Vision (CV) algorithms is a commonly used method for autonomous yield estimation. There are two main types of CV-based approaches to estimating agricultural yields:

(1) methods that focus on specific regions or areas, and (2) methods that rely on counting. An automated method for estimating crop production in apple farms was created by Wang et al. [6] using stereo cameras. To lessen the impact of the erratic daylight lighting, they took the photos at night. An in-field cotton recognition system was created by Li et al. [7] using region-based semantic image segmentation. Joint maize tassel and crop segmentation was accomplished by Lu et al. [8] using region based color modelling. Yield estimation approaches based on counting have received surprisingly little attention, in comparison to methods based on regions [9]. Estimating the quantity of apples harvested in fields with natural light was done by Linker et al. [10] using color photographs. There were a lot of false positives because of the problems with direct light and color saturation. A technique for apple fruit segmentation [21] from video utilizing backdrop modelling was developed by Tabb et al. [11].

Counting problems demand one to reason about the total occurrences of an object in a scene, as opposed to the usual picture classification procedure that aims to identify the existence or nonexistence of an object. Multiple real-world applications encounter the counting problem: counting cells in microscopic imaginings, counting wildlife in aerial photos [12], counting fish [13], and crowd monitoring [14] in surveillance systems. Kim et al. [15] presented a system that uses a fixed-shot camera to recognize and follow moving subjects. To improve loss optimization during learning, Lempitsky et al. [16] presented a novel supervised learning structure for pictorial object counting jobs that considers MESA distance. The authors Giuffrida et al. [17] put forward a method for leaf counting that relies on learning in plants that grow in rosette sets. They connected image-based descriptors learned unsupervisedly to leaf counts using a supervised regression model. The present method of estimating production, which involves workers physically counting fruits or flowers, is impractical for vast fields due to its high cost and time requirements. Here, a practical answer is provided via robotic agriculture-based automatic yield estimation.

Nowadays, AI is playing a bigger part in fruit counting as it provides more precise and efficient answers for farming. Automating fruit counting in fields is possible with the usage of AI technologies, especially CV and ML. DL algorithms allows for automated interpretation of captured images or films. Based on visual features like size, shape, texture and color these systems are able to recognize and tally fruits developed by Koirala and Zhang [18]. DL algorithms have been taught to recognize and quantify fruits in images. These algorithms include Convolutional Neural Networks (CNNs) with massive datasets these models gradually get more accurate results by Sa et al. [19]. According to Wang et al. [24] AI is used in combination of LiDAR and 3D imagery to make three-dimensional fruit count estimates. A more precise evaluation of the distribution and volume of fruit can be achieved with this method. The author also explained that the method can be used for vast agricultural fields. According to Anand et al. and Kumar et al. [20] this combination enables thorough counting and monitoring through effective aerial surveys of vast agricultural fields. Adapting it to different orchard settings and fruit varieties is a breeze. These systems may be adjusted to various situations, which means they can be used with a variety of crops. Methods for counting objects using deep

learning have recently become more prominent. Seguíet al. [19] investigated the use of CNN for the job of counting instances of an interest notion. A system for microscopy cell counting was created by Xie et al. [22] using a convolutional regression network. Using deep CNN, Zhang et al. [23] created a framework for cross-scene crowd counting. So far as we are aware, no studies have addressed the topic of deep simulated learning fruit counting. All of the counting algorithms that have included deep learning have focused on object detection and subsequent counting of those instances.

B. Water Management

From legislators to end users, everyone involved in water usage and management is worried about the impending water scarcity. The opinions of many shareholders or a shortage of revised strategies and plans to increase efficacy can make it difficult to execute any freshwater conservation strategy Marston & Cai, [25]. These concerns about effective freshwater management are of particular importance in agriculture, where they may help alleviate sustainability and environmental concerns while also cutting expenses for farmers. According to Salmoral et al. [26], public institutions and lawmakers play a crucial role in this context, specifically under the EU's Common Agricultural Policy (CAP).

Actually, circa the agricultural sector drew around 70% of the world's water. In the Asian and African regions (81%), as well as in Oceania (65%), this is a very pertinent subject. This issue warrants particular attention in southern countries, while it is not as serious in European and American countries (25% and 48%, respectively) Aquastat [27]. Several stakeholders, including farmers, must be involved in the planning and execution of any strategy or plan to improve water efficiency on farms for it to be effective.

According to Koscielniak et al. [28], Nazari et al. [29] the agricultural sector of the European Union relies on proper water management, so it's important to shed light on these factors. Several factors influence the efficiency of water management in irrigation methods. These elements include pertaining to the environmental, social, technical, legal, and political aspects. In view of Castanedo et al. [30], such settings, considerations such as the depth of application and modified drainage systems may be important. Agricultural methods including energy usage and soil management strategies are interdependent on irrigation practices Lee et al. [31]. Surface irrigation agricultural output, and soil yields Kim et al. [32] are three areas where irrigation practices can substantially affect water management efficiency. The morphology and spatial circulation of roots from perpetual crops Deng et al. [33] and economic indices of farms Kumar et al. [34] are also affected.

Regardless, irrigation technology and methods have advanced, allowing farmers more leeway in their decisions and options Roth et al. [35]. Nonetheless, there is always room for improvement in this area van Steenberg et al. [36]. The selection of water-efficient cultivars in the turf business is another important consideration Githinji et al. [37]. In this context, effective strategies for managing water resources are crucial.

According to Preite et al. [38], 4.0 technologies are being considered as a possible result to enhance the agricultural

sector's sustainability. These technologies include blockchain, the Internet of Things (IoT), DL algorithms, ML and other computer applications. The simple, scalable automation that predictive algorithms offer makes them ideal for a 4.0 scenario that spans many different fields Mazzei & Ramjattan [39]. Meshram et al. [41], Liakos et al. [40], and others have grouped the machine learning methods used in agriculture into three distinct phases: before, during, and after harvesting. The first set of applications included topics related to irrigation, with identification of water scarcity, prediction of water demand, and scheduling of irrigation. Conventional irrigation scheduling takes a set period into account while ignoring the fact that environmental and plant variables can vary. In particular, water scarcity identification processes thermal infrared images, weather, and soil data to assess stem water potential, drought stress and plant water gratified.

According to Zhou et al. [41], the models mostly used in this scenario were gradient-boosted random forests, decision trees and CNN. Using support vector machines, gradient-boosting, artificial neural networks and decision trees algorithms, reference evapotranspiration, soil moisture contented and sap flow possessions were estimated using multispectral and thermal imageries in conjunction with meteorological and soil data. By analyzing sensor data, the authors of Corell et al. [42] present an outline for irrigation that compares three regression models to find optimal irrigation amount for olive farming. The emergent degree days, water provided to plants, and evapotranspiration rate were used in a fuzzy decision support system to evaluate appropriate irrigation quantity for corn, kiwi, and potato crops Giusti & Marsili-Libelli [43]. In order to give watering suggestions for lemon trees, Navarro-Hellín et al. [44] utilized an adaptive neural fuzzy inference system in conjunction with a partial least-square regression to analyze evapotranspiration, soil moisture and humidity. Chandrappa et al. [45] use DL algorithms (Long Short-Term Memory) and ML techniques (Support Vector Regression and Linear Regression) to evaluate soil moisture changes in depth and time. Against this backdrop, a multi-depth link between wind speed and soil moisture was brought to light. By training an artificial neural network to use data from soil sensors and meteorological stations to calculate the optimal irrigation period, a 20% reduction in water use was accomplished by Gu et al. [46]. Kavya et al. [47] have investigated the use of AI for short-term water demand prediction. In particular, using both univariate and multivariate time series assessed the prediction ability of deep learning and machine learning. While the multivariate scenario also took weather into account, the univariate series was applied just to the flow meter data. A probabilistic framework was created by Srivastava et al. [48] to ascertain irrigation methods using three distinct parameters: leaf area index, soil moisture, and evapotranspiration. These indicators show water deficiency in the soil, water stress in crops, and the water demand, in that order. Here they utilized a Recurrent Artificial Neural Network (long short-term memory) to make predictions, and employed a random forest regression to find good predictors for each parameter. The last step was compared the expected and actual numbers to tweak the resulting weights.

Aly et al. [49] used a super learning ensemble to predict the evapotranspiration with limited meteorological data. They achieved good accuracy by utilizing additional tree regression,

k-nearest neighbour, support vector regression, and AdaBoost regression. Yong et al. [50] also noted the latter difficulty as the primary obstacle to evapotranspiration rate prediction and proposed a hybrid neuro-fuzzy inference method to overcome it. Adnan et al. [51] examined practicality of hybrid support vector regression models from this angle. These models integrate ML methods with optimization meta-heuristic algorithms, such as Particle Whale Optimization, Swarm Optimization, Differential Evolution, and Covariance Matrix Adaptation Evolution Approach. By integrating ML and feature engineering, Považanová et al. [52] enhanced prediction accuracy for reference evapotranspiration estimation, shedding light on the efficacy and generalizability of the suggested models. Using a variety of machine learning algorithms including k-nearest neighbors, support vector machine, decision tree, and multilinear regression the authors of Youssef et al. [53] demonstrated how to estimate reference evapotranspiration with an accuracy close to 99%.

C. Crop Management

One of the most significant parts of agriculture has always been crop production management. In order to feed both cattle and humans, crop production is crucial. Throughout human agrarian history, one of the key objectives is to upsurge the economic efficacy of farming. To ensure consistently high-quality output, agricultural production sites should undergo routine inspections and implement all required crop production strategies. Because farmers invest time and energy into each visit, the crop's price tag reflects that. As a result of farmers' obsession with crop monitoring and evaluation, smart agriculture has emerged as a critical tool. Although digitalization will have a greater effect on wide-area communication networks that include rapid data transmission, it permeates most areas of engineering [54]. Cultivating field crops, producing vegetables, and fruit are all part of crop production, which is a subset of agriculture [55]. "Smart farming" refers to a new paradigm that maximizes agricultural output with the help of cutting-edge information technology [56] with advancements in AI, automation, and connectivity, farmers can effectively monitor different procedures and provide targeted treatments for cultivation using robots that are superhumanly efficient.

These tasks only require a set of guidelines based on mathematics or logic because to derive valuable correlations from data, machine learning makes use of learning rules like supervised learning, unsupervised learning, hybrid learning and reinforced learning [57].

These features allow deep learning networks to potentially uncover hidden structures in data that is neither labeled nor structured. A major improvement over previous methods, deep learning networks are able to extract features with little to no human intervention. The proliferation of high-speed wireless transmission networks led to dramatic increase in consumer demand for such services [58]. When comparing Deep Anomaly to region-based convolution neural networks (RCNN), the former is superior for human detection at 45–90 meters [59]. This method can detect anomalies and generate uniform field characteristics. In this article, learned about the DL classification of land cover and crop kinds using remote sensing data [60]. Traditional fully linked MLPs and random forests were compared to CNN. We talk about how to use

visual sensor data to train self-learning CNN to identify diverse types of plants [61]. Offers automatic weed detection in UAV photos of line crops using deep learning with unsupervised data labeling [62]. Use of convolutional neural networks (CNNs) on unsupervised training datasets will provide fully autonomous weed detection. Incorporating a deep residual neural network onto a mobile capturing equipment allowed for the introduction of a crop disease classification system. Thorough testing enhanced the precision of the balancing process. 0.78 to 0.8 [63] is the range.

To diagnose mildew disease on millet crop photos, a deep neural network with transfer learning is employed [64]. The f1-score was 91.75%, recall was 94.50%, precision was 90.0%, and accuracy was 95% in the experiments. A deep convolutional neural network was employed to estimate agricultural yields using NDVI and RGB data acquired by UAVs [65]. In terms of CNN performance, RGB images beat NDVI images. In terms of critical characteristics, low-altitude remote sensing-based images and CNN architecture for rice grain production were considered [66]. During the ripening stage, Deep CNN performed significantly improved and was stable. Researchers have looked at a deep learning-based multi-temporal crop classification system [67]. DL models LSTM and Conv1D were compared to XGBoost, SVM, and RF parameters. The development of a new crop vision collection that makes use of deep learning classification and accurate agricultural recognition has also been accomplished [68]. On agricultural datasets, his proposed algorithm achieved a 99.81% accuracy rate, surpassing VGG, DenseNet, ResNet, SqueezNet, and Inception. Recognizing and differentiating crops in soil is made possible by deep learning technology [69]. Information is derived from a digital surface model with a high level of resolution. For the purpose of crop pest classification, automatic feature extraction is used in conjunction with transfer learning approaches involving convolutional neural networks [70]. The most accurate datasets are Xie1, NBAUR, and Xie2, with respective accuracies of 94.47%, 96.75%, and 95.9%.

D. Soil Management

For the vast majority of creatures, the soil is the food web, providing them with the mineral resources they need to survive. When soils are well-managed, plants do not suffer from mineral element deficiencies or toxicities, and the right minerals make it into the food chain. Crop yield, ecological stability, and human well-being are all impacted by poor soil management in some way.

According to Dickson et al. [72] and Bhaskar et al. [71] soil categorization opens up numerous sectors including soil improvement, crop management, land consolidation and more. Physiological factors assessed from real-time field models are the most important criteria for soil identification. Root development, plant emergence rate, water penetration, and crop production are all affected by physical variables such as temperature and moisture, which affect the formation of particles and pores. Chemical features including pH, organic carbon, and the nitrogen, phosphorus, potassium (NPK) parameters dictate the accessibility of nutrients, the existence of other species, and the motility of pollutants. The various components that make up soil include clay, sand, peat, silt, and loam. Soil

particles in the target zone consist mostly of sand, clay, and silt, with very little peat and loam.

It is considerably more challenging to keep these soils suitable for farming. Soils like laterite, which are mostly composed of rock deposits from hot climates, are abundant in iron and aluminum. Soils with large concentrations of iron oxides, such as laterite, have a reddish hue [72]. Almost all laterites have a rusty-red hue because of the significant iron oxide content. Soil surface formation is guaranteed by periodic rainfall and sunny seasons. The crops are adequately nourished by this soil type. The southern Indian subcontinent is a significant producer of the rice variety *Oryza sativa*. Rice is a staple crop and a source of income for many farmers in the area surrounding the exploration location. Milling, visual, culinary, and nutritional qualities are all part of what makes rice grain quality. It is widely recognized that the root's balanced qualities are closely related to grain quality in rice. Root morphological and physiological features impact rice vegetative growth and grain satisfying, which in turn affects grain quality. The features that were researched and described and are applicable to the exploration site, which consists primarily of clay and laterite soil.

A methodology for digital soil mapping was created by Behrens et al. [73] using Artificial Neural Networks (ANNs). This methodology is able to predict soil units in a test area in Rhineland, Germany, Palatinate. Grinand et al. [74] developed a classification tree-based method for predicting soil distribution at an unexplored location by using a soil-landscape pattern obtained from a soil map. Soil datasets and exploration site data should be collected as part of the proper method for soil classification at the exploration site. After that, the datasets should be pre-processed. Finally, models should be trained using Deep Neural Network and Machine Learning techniques, and the soil should be classified into four distinct groups. They rely on accurate soil detection to help with nutrient supply to the field, which in turn increases crop production. It's also crucial for their livelihood to determine what kind of weeds will grow from the soil so that they can eradicate them.

E. Weed Identification

One of the most important things that can influence crop yield is weed control. Khan et al. [75] found that weeds can reduce crop output and production quality by competing with crops for water, fertilizer, light, growing space and other nutrients. Insects and diseases that harm crops could also call this place home. A study found that weed suppression resulted in a 13.2% annual loss of crop production enough to feed one billion people for a year Yuan et al. [76]. A key component of crop management and ensuring food security is weed control. Manual weeding, chemical weeding, biological weeding, mechanical weeding, etc. are all common weed management strategies Stepanovic et al. [78], Marx et al. [77], Morin [80], Kunz et al. [79], Andert [81].

The best method for controlling weeds in the field is to do it by hand. The high cost and labor intensity, however, make it impractical for cultivation on a broad scale. Because it doesn't harm non-target organisms much, biological weeding is eco-friendly and safe, but it takes a lengthy time to restore ecosystem afterward. The majority of weeds are eliminated

with chemical weed killers, which is the most popular method of weed control. However, other problems, including chemical residues, weed resistance, and environmental contamination, have resulted from the excessive use of herbicides. The study found that in different farmland systems, 513 biotypes of 267 weed species have become resistant to 21 different herbicides Heap, [82]. Therefore, it will be crucial to use technologies like detailed spraying or mechanical weed management on individual weeds in order to prevent the over-application of herbicide.

Automatic mechanical weeding is becoming more popular as a result of the organic farming movement Cordill and Grift [83]. It prevented needless tillage, which saved gasoline, and allowed for weed management without chemical input. Nevertheless, intelligent mechanical weeding has faced significant challenges because to the low accuracy of weed detection and the resulting unforeseen harm to the plant-soil system Gašparović et al. [85], Swain et al. [84]. So, it's critical to make weed detection more precise in the fields.

So using AI models like SVM, decision tree, a random forest algorithm, and KNN classifiers are some of traditional AI methods that have been utilized in weed identification research. It is expected that these algorithms will employ intricate manual craftsmanship to extract weed image color, texture, form spectrum, and other attributes. As a result, the weed image extraction was lacking or features were obscured, it would be impossible to differentiate between weed species that are otherwise comparable. Image processing technology was utilized by traditional weed detection algorithms to extract characteristics of weeds, crops, and backgrounds from images. A model that uses wavelet texture information to differentiate sugar beets from weeds was presented by Bakhshipour et al. [86]. A total of fourteen of the fifty-two texture features were chosen using principal component analysis. Despite numerous occlusions and overlapping leaves, it proved wavelet texture features might accurately differentiate among crops and weeds. Only crops and weeds with clearly distinct pixel values in the RGB matrix or other parameter matrices derived from it could be identified by the color feature-based models. In most cases, the color feature was utilized in conjunction with other features; for instance, Kazmi et al. [87] suggested a technique that combined surface color with edge form to detect leaves and integrate vegetation indices. With a precision of 99.07%, the vegetation index was combined with regional characteristics. It was challenging to differentiate between weed species using traditional image processing approaches, even if same methods could differentiate between crops and weeds.

To improve weed detection, deep learning networks can generate abstract high-level properties instead of the low-level attributes used by traditional machine vision networks, such as color, shape, and texture. The present target identification models have, as is well-known, benefited from deep learning's increased accuracy and generalizability. A few examples of popular target detection networks are the YOLO model, Faster R-CNN, and Single Shot Detector Redmon et al. [88], Ren et al. [89], Quan et al. [90]. Using a total of 10,413 pictures, Dyrmann et al. [91] employed CNN to distinguish 22 distinct plant species. The weed species with the most picture resources had the highest classification accuracy, according to the results.

Therefore, there needs to be enough datasets for deep learning-based weed identification.

The author Hinton et al. [92] proposal highlighted the deep and highly-connected topology of DL networks, which led to the idea of deep learning being introduced. The dataset is trained by Deep learning has been demonstrating strong accuracy and resilience in image identification as of late. To be more specific, ImageNet a massive multi-variety dataset with 3.2 million images demonstrated the significance of large-scale datasets in enhancing the identification accurateness of the models trained with DL methods by Russakovsky et al. [93]. Unfortunately, dataset for training deep learning weed identification models have very tiny scales in both the number of images and the type of weeds.

F. Seed Categorization

Farmers and food processors alike are understandably worried about seed segregation in mixed cropping. Farmers and agro-industries also have the difficult challenge of classifying and packing seeds according to their quality. Additionally, the conventional methods of seed separation after threshing like sieving, hand-picking, etc. are laborious and time-consuming. Therefore, seed segregation must be automated.

For that AI methods play an important role in yield prediction [94], improvement of image contrast [95], illness categorization [96], etc. inspired each of the study [97] in order to broaden the scope in seed categorization according to variety, size, they are of high quality. Using SVM, the authors of [98] were able to categorization of normal and broken maize kernels [99]. The SVM classifier achieved a 95.6% success rate for healthy and an 80.6% success rate for the process of identifying damaged or defective seeds, an error rate of about 19% was noted. Researchers [100], [101], and [102] continued this line of inquiry by classifying four different types of maize seeds using models based on SVM, K-means and DCNN. They used the DCNN and claimed a perfect training accuracy rate based strategy. However, when measuring the model's efficacy on the testing dataset, a significant amount of incorrect classifications was found in relation to a single corn category.

In addition, the researchers [103] automated the process of inspecting maize kernels by utilizing ML and DL models' capabilities. For kernel separation, they utilized k-means clustering. In order to distinguish between kernels that were flawed and those that were not, they used a number of models, including ResNet, VGGNet, and AlexNet. Outperforming VGGNet and AlexNet, the ResNet model achieved an accurateness of 98.2%. The writers in [104] also distinguished between healthy and malformed corn seeds using SVM, AlexNet, VGG-19, and GoogleNet. The GoogleNet model had the highest accuracy rate of 95% out of all of these models. In order to classify and test seeds, the following works [105] employ ML and DL algorithms effectively. In order to distinguish between haploid and diploid seeds, detect seed coating, distinguish between common maize seed and silage seed for animal feed, and identify defective from non-defective seeds, they utilized Convolutional Neural Network (CNN) classifiers. Sunflower seed identification was accomplished by the authors [106] using DL models. By utilizing optimization procedures, they

successfully circumvent the issue of overfitting. It was asserted by the authors that the optimized GoogleNet model attained a 95% accuracy rate. Unlike a large lot, however, the model calls for human involvement to arrange the seeds.

When training the model, the authors also took into account just one perspective on seeds. Consequently, by training the model on numerous perspectives of seeds, there is a chance to enhance its robustness and reliability. In order to incredulous the obstacles stated in the previous study, the authors in [107] took into account the entire soybean seed surface. They achieved a 98.87% success rate by using a circumrotating method for full surface detection. When applied to the dataset that included defective seeds, the MobileNet model enhanced the classification accuracy. In addition, the technique for identifying soybean seeds was suggested by the authors in [108]. To demonstrate the effect of transfer learning, they used pre-trained CNN models such as AlexNet, Xception, ResNet18, Inception-v3, DenseNet201, and NASNetLarge. With a reported accuracy of 97.2%, the authors asserted that NASNetLarge was the most accurate model. Using morphological and textural characteristics of seeds, the authors of [109] extended the use of ML models for weed detection by applying the naïve Bayes algorithm [110]. The model's accuracy on the grayscale and monochrome photos was 98%, according to the research. Colored images, however, show a marked decline in accuracy.

G. Yield Forecasting

Predicting how much food will be harvested from a specific plot of land is known as Crop Yield Prediction (CYP). Businesses, governments, and farmers all rely on it to help them make educated decisions on agricultural output. The varied temperature, topography, temporal dependencies inherent in yields and farming techniques across India make accurate crop yield forecast a difficult undertaking. Nonetheless, one can anticipate crop production based on a number of criteria, such as: Outside conditions: When it comes to determining harvest success, the weather is a major player. When it comes to plant growth, factors like rainfall, temperature and humidity are important. Crop yield is also prejudiced by the soil's type and fertility. To account for these aspects and anticipate crop yield, one might utilize crop yield prediction models. These models can use machine learning, statistical methods, or a mix of the two.

Consequently, better approaches for assessing and modelling agricultural data are required to enhance crop yield forecast and management. Using ML algorithms and proximate sensing, Farhat Abbas et al. [111] established a CYP system. In order to conduct training, four datasets that are available to the public were gathered: PE-2017, PE-2018, NB-2017, and NB-2018. In order to forecast agricultural output, the gathered data were fed into machine learning models such k-nearest neighbor (KNN), support vector regression (SVR), linear regression (LR), and elastic net (EN). With a smaller Root Mean Square Error (RMSE) than competing techniques, the SVR outperformed them on all four datasets. Martin Kuradusenge et al. [112], introduced many ML models in order to improve the system's performance, the Irish potato and maize datasets were first collected and pre-processing activities, such as removing null values and determining association, were

executed. Afterwards, three ML models SVM, Random Forest (RF) and Polynomial Regression (PR) were used to classify the pre-processed data for CYP. When it came to forecasting potato and maize crop yields, the RF model outperformed the SVM and PR models, with RMSEs of 510.8 and 129.9, respectively, on the datasets that were examined.

Recurrent neural networks and temporal convolutional networks are examples of the hybrid DL techniques that Liyun Gong et al. [113] suggested for CYP. The data was gathered from many actual tomato-growing greenhouses. Before feeding the standardized data to the RNN for processing, gathered data was pre-processed using data normalization. Lastly, TCN was instructed to process tomato CYP using the RNN's output. For the datasets that were collected, the technique outperformed the similar methods with reduced RMSE. For CYP with agrarian characteristics, Dhivya Elavarasan and P. M. Durai Raj Vincent [114] introduced a hybrid method known as reinforced RF. At first, the system retrieved crop data from the agricultural dataset and input it into the reinforced RF hybrid DL model. The relevance of the input data was determined by the reinforced RF using the reinforcement learning approach in every internal node. After that, the RF classified crop yield using the most important variables found by the reinforcement model. Outperforming state-of-the-art ML models for CYP including SVM, LR, and KNN, the hybrid technique produced superior results.

To optimize CYP, Aghila Rajagopal et al. [115] created a deep-learning approach. After the data was pre-processed, principal component analysis was used to extract the important features from the pre-processed dataset. After that, an updated chicken swarm technique was used to further optimize the characteristics that were chosen in order to boost the classifier's performance. Lastly, a discrete DBN-VGGNet classifier was used for classification. Outperforming the prior state-of-the-art models, the system attained a 97% accuracy rate with a 0.01% MSE. For large-scale CYP, Dilli Paudel et al. [116] proposed a set of machine-learning models. Data on agricultural yields, including results from crop growth simulations, weather measurements, and yield statistics, were first gathered by the system from a variety of sources. Preparation for categorization procedures involved cleaning the acquired data. The classifier was then fed samples of input data that had undergone feature design. As for CYP, it made use of ML classifiers such as SVM, Ridge regression, KNN, and gradient-boosted decision trees.

H. Disease Detection

Plant diseases are a worldwide threat to food security and can also have serious personal consequences. The economy and the security of our food supply depend critically on healthy crops. A crop's health can only be gauged by its growth and leaf condition.

Therefore, by analyzing symptoms seen in leaf images can learn about many plant illnesses. Every year, farmers can lose a substantial amount of money due to several plant diseases that impact vegetables like potatoes, tomatoes, and peppers. Early blight and late blight are the two varieties of blight. Though a particular bacterium causes late blight, a fungus causes early blight. By promptly detecting and efficiently treating these

diseases, farmers can save both time and money. In the next twenty-five years, the human population is projected to surpass 9 billion. A 70% upsurge in food production is necessary to keep up with the continuously increasing demand for food. Many nations, particularly those with a strong agricultural economy, face the devastating threat of crop disease.

By extracting data from real time image processing with ML and DL become prominent tool for plant disease identification because it will effectively diagnose plant illness by exploring with computer vision, machine learning approaches have shown promise by extracting data from real-time image processing. There has been extensive use of classic ML methods for plant disease detection, including feature extraction and classification. Color, texture, and form are some of the visual attributes that may be extracted using these methods to train a classifier to distinguish between healthy and sick plants. Diseases like leaf blotch, powdery mildew and rust as well as symptoms of diseases caused by abiotic stresses like drought and nutrient deficiency, have been extensively detected using these methods Anjna et al. [118], Mohanty et al. [117], Genaev et al. [119]. However, these methods do not accurately identify subtle symptoms of diseases or detect diseases in their early stages. They also have trouble management complicated and high-resolution images.

By using DL technology like CNNs and DBNs to detect pests and irregularities in plants. The use of these technologies to detect and identify lesions from digital pictures has been yielding encouraging results by Kaur and Sharma [120], Siddiqua et al. [121], Wang [122]. Deep learning models have the ability to automatically learn image attributes, allowing them to detect subtle disease symptoms that could otherwise go undetected by typical image processing approaches. However, not all applications can accommodate Deep Learning models due to their high processing requirements and large amounts of labelled training data. In order to locate and identify certain areas of interest in images, like disease symptoms or plant leaves, CV methods like object detection and semantic segmentation can be employed Kurmi and Gangwar [123]. By combining these techniques with ML or DL algorithms, images can be automatically transformed into patterns or features that can be used for disease identification and categorization. To train their models, CV algorithms require massive amounts of labelled picture data, which means they might not be able to handle previously discovered diseases.

Image, sensor, and meteorological data, among other massive datasets, have been subjected to ML and DL-based analysis in order to uncover patterns and generate forecasts. Cedric et al. [125], Yoosefzadeh-Najafabadi et al. [124] and Domingues et al. [126] are just a few examples of ML algorithms that are actually used to forecast crop yields, detect plant diseases and pests and optimize plant growth. Sladojevic et al. [127], Alzubaidi et al. [128], and Dhaka et al. [129] all found that DL models, including CNNs and DBNs, outperformed standard image processing approaches when it came to plant lesion diagnosis using image analysis and classification. Compared to more conventional ways, ML and DL-based methodologies provide many benefits in the fields of agriculture and botany. These techniques can evaluate massive amounts of data, automate activities, and improve accuracy and efficiency.

1. Harvesting of Yields

Rising food demand due to population growth is the biggest threat to food security. In order to increase supply, farmers will need to enhance yields while utilizing the same amount of land. Technology can help farmers increase production through agricultural output prediction. For better crop selection and management during the growing season, decision-makers can employ CYP a decision-support tool powered by ML and DL. During the growing season, it may choose which crops to harvest and how to tend to them.

With the use of agricultural yield estimation, farmers may increase output when weather is good and reduce output loss when weather is bad. Positive predictions of agricultural output are affected by a great deal of variables, including farmer practices, decisions, pesticides, fertilizers, weather, and market pricing. Climate, area wise production, rainfall, and historical yield statistics can all be used to make educated guesses about future crop yields. AI methods has been making strides in many sectors, including farming, as of late.

In order to predict the harvest used the dataset that includes the entire cultivated area, the length of the canals, the average highest temperature and irrigation water sources like wells and tanks. The researcher created computational model outperformed alternatives built with Regression Tree, Lasso, Deep Neural Network and Shallow Neural Network techniques. The RMSE for dataset validation using forecasted weather data is 12% of the average yield and 50% of the standard deviation [130]. Using the following parameters: minimum/maximum/average temperatures, rainfall, area, production and yield, the accuracy was 97.5% from 1998 to 2002 for the Kharif season [131]. Crop production estimates during the Kharif season in Andhra Pradesh's Vishakhapatnam district were the primary focus of the study. Because rainfall has such a large impact on the yield of Kharif crops, researchers first employed modular artificial neural networks to predict when it would rain, and then they used SVR to estimate the yield of crops based on both area and rainfall. These two methods were used to increase the harvest productivity.

The research aimed to accomplish four things: first, study how well the ANN model predicted corn and soybean yields when weather was bad; second, compare the evolved ANN model to other multivariate linear regression models; and last, test how well the model estimated yields at the regional, state, and local levels. Researchers in India's Maharashtra state employed artificial neural networks to compare rice harvests in different urban areas. They used the Indian government's accessible records to compile data for Maharashtra's 27 districts.

This study estimates higher crop yields utilizing ML methods like KNN, SVR, RF and ANN. The research's data set consists of 745 examples; 70% of those cases were randomly assigned to train the model, while 30% were used for testing and performance evaluation. Random Forest is found to obtain the highest level of accuracy in the final analysis of maya gopal P.S [132]. The study proposes a novel model for soybean yield prediction using Long-Short Term Memory (LSTM) satellite data collected in southern Brazil [133]. The main objective of the study is to evaluate LSTM neural networks, random forest, and multivariate OLS linear regression for

their effectiveness [134]. The first stage in using rainfall, land surface temperature, and vegetation indices as self-determining variables to forecast soybean data is to find out how soon the model can reliably expect the yield. All algorithms are outperformed by Long Short Term Memory for all forecasts except DOY 16. According to [135], when it comes to DOY 16, multivariate OLS linear regression is the best algorithm. This study discusses the outcomes of applying a Sequential Minimal Optimization Classifier. Data from 27 districts in Maharashtra, India, and the WEKA tool were used to conduct the experiment. Other strategies perform better than Sequential minimum optimization, according to the results of the experiment on the same dataset. While Multilayer Perceptron and BayesNet showed the greatest accuracy and enhanced quality, sequential minimum optimization showed the worst accurateness and poor quality [136]. One method that has been suggested for estimating crop productivity is the use of Parallel Layer Regression (PLR) and Deep Belief Networks (DBNs). Pulses, ragi, rice, and cassava are five of Karnataka's most important crops that are being studied using a DBN technique. Each entry in the applicable database is forecasted by the proposed methodology to produce one of the five crops. Finally, the experimental results show that the method has great promise for real-time data and human interaction validated accurate prediction of agricultural efficiency in terms of specificity, sensitivity and accuracy [137].

By utilizing a KNN algorithm, a CYP System (CYPS) is put into place. Yield projections, on the other hand, need to take into account a number of variables that can affect the quantity and quality of a farmer's harvest. In order to forecast yield production, authors employ precise fields such as year, crop, area, region, and season. These factors, along with crop type and production area, have a significant impact on yield production. Accurate understanding of crop yield history is necessary for decisions linked to agricultural risk management [138]. Rao et al. [139] used two separate metrics, entropy and GINI, to compare Random Forest, Decision Tree Classifier, and KNN. RF has produced the most precise outcomes, according to the findings. Based on feature vectors, VGG_19 achieved a good performance of 91.35% and VGG_16 achieved a good performance of 91.17% [140]. Because of its great efficiency, hydroponics has been suggested by Vanipriya et al. [141] as a solution to the problem of low agricultural production in India. Furthermore, it provides a more environmentally friendly option for soil cultivation. The economy and agricultural output are two factors that determine food production [142].

III. LIMITATIONS AND FUTURE STUDY

Based on the study, the findings of agriculture based on image processing has a lot of potential to automate and improve farming tasks with different agriculture farming. For future comprehensive insights there is a need to enhanced farming by combining image, IoT, data fusion techniques, transfer learning and domain adaption and computing techniques. The techniques like DeepLab [143] can be used to classify plants, detect pests, and analyze soil because it is a semantic segmentation model for classifying every pixel. The other method is efficient net [144] which is used to detect disease, fruit counting and yield prediction effectively because it is designed to optimize the model and computation. While

our study mainly highlighted the CNN model mostly because it can understand the decision making process in farmers to balance perspective for further refinement. While this review has different datasets for different farming which highlights class imbalance, performance scenario for model evaluation, and comparability. In farming, sustainability is considered as a main factor for long term viability of the environment through energy consumption, and electronic waste associated with cost environment.

However, it isn't perfect for all jobs because of some problems. Here's a look at how these limits affect different farming tasks:

A. Fruit Counting

1) *Occlusions and crossing over*: Fruits that are hidden by leaves or that intersect with other fruits can make it hard to count or identify them.

2) *Changes in lighting*: Sunlight or artificial lighting can cast shadows and create effects that make it harder to see fruits and vegetables.

3) *Challenges unique to each species*: Because fruits come in many shapes, sizes, and colors, they require very specific formulas.

4) *Environments that change*: Moving wind or changes in the shape of the tree can make it harder to locate the fruit.

B. Water Management

1) *Problems with surface reflection*: High reflection from bodies of water or irrigation systems can make it hard to determine how much water is in an area or how it is distributed.

2) *Limitations of resolution*: Images from satellites or drones might not have enough detail for micro-irrigation and other small-scale water management tasks.

3) *Estimating the soil moisture*: Indirect methods, like NIR imaging, might not provide an accurate reading of soil moisture because dryness on the top can mask the conditions below.

C. Crops Management

1) *Changes in growth stages*: The appearance of crops changes significantly over time, so adaptive programs are needed to monitor them continuously.

2) *Problems with the environment*: When taking images outdoors, weather conditions like rain or fog can make it difficult to see crops.

3) *Difference between weeds and crops*: It's challenging to differentiate between crops and weeds that are grown closely together because they appear similar.

D. Soil Management

1) *Data at the surface level*: Image-based methods usually only show what's on the surface and don't reveal things like nutrient levels or soil compaction.

2) *Dependence on indirect indicators*: The color and texture of soil that are inferred from images might not always be a reliable indicator of its fertility or organic content.

3) *Environmental factors*: Changes in light, moisture, or plant debris can complicate the assessment of soil condition.

E. Weed Identification

1) *How they are like crops*: Weeds that look like crops in terms of leaf structure or color can be difficult to tell apart.

2) *Lots of plants*: It's hard to tell the difference between weeds and crops in areas with a lot of crops.

3) *Changes with the seasons*: Weed growth trends change with the seasons, requiring models to be retrained frequently.

F. Seed Categorization

1) *Changes in size and shape*: Seeds from the same species can naturally vary in size, shape, and texture, complicating classification.

2) *Waste and impurities*: Misclassification can occur when images contain trash or damaged seeds.

3) *Problems with lighting and contrast*: Uneven lighting can obscure crucial features of a seed necessary for identification.

G. Yield Forecasting

1) *Complex networks of dependencies*: Yield depends on many factors that are difficult to discern from images alone, such as weather, soil health, and pest presence.

2) *Lack of data*: Model accuracy suffers from the absence of historical image data for certain crops or regions.

3) *Problems with spatial resolution*: Low-resolution images might not capture important crop features essential for accurate predictions.

H. Disease Detection

1) *Signs of an early stage*: In the early stages of a disease, subtle changes in the texture or color of leaves might be too faint for standard image processing to detect.

2) *Nutrient deficiencies and other problems*: Some diseases exhibit symptoms that are very similar to those caused by nutrient deficiencies, which can lead to incorrect diagnoses.

3) *Noise in the environment*: Dust, water droplets, and other impurities on plant surfaces can make accurate disease identification challenging.

I. Harvesting

1) *Conditions of the dynamic field*: Changing field conditions, such as uneven terrain or variable lighting, complicate the task for robots to harvest crops using image processing.

2) *Produce that is Covered or hidden*: Fruits and vegetables that are fully or partially hidden are difficult to locate and harvest accurately.

3) *Risk of damage*: During automated picking, damage can occur if items are not properly positioned or identified.

J. General Cons Across Applications

1) *Dependence on data quality*: Models perform less reliably when images are of poor quality, resolutions are not uniform, and diverse datasets are lacking.

2) *Problems with scalability*: Real-time processing for large-scale systems (like entire farms) requires substantial computational resources.

3) *Issues with adaptability*: Image processing algorithms need to be retrained for new crops, regions, or weather conditions.

4) *Hardware limitations*: Small-scale farmers may not be able to afford as many drones, cameras, and other imaging tools, increasing costs and reducing accessibility.

K. Pros that Apply to All Situations

1) *Real-Time monitoring*: Imaging and sensors provide up-to-the-minute information, allowing immediate responses to changes in the field.

2) *Scalability*: Data analysis tools can handle large datasets, making them suitable for both small farms and large-scale operations.

3) *Cost savings*: Optimizing resource use reduces expenses on water, chemicals, and labor.

4) *Sustainability*: Promotes environmentally friendly practices by minimizing the use of excessive energy, water, and chemicals.

5) *Precision agriculture*: Delivers precise, relevant information that increases output and reduces waste.

6) *Risk reduction*: Predictive models identify potential risks such as drought, pests, or diseases.

7) *Enhanced Decision-Making*: Provides valuable insights based on historical trends, current conditions, and predictive algorithms.

8) *Accessibility*: Data analysis tools are accessible to farmers worldwide, even in remote locations, via mobile apps and cloud-based platforms.

L. Potential Author Bias

- Potential bias in evaluation and model selection
- Limitations related to Dataset: like data quality, imbalance, insufficient data
- Overfitting and generalization of model for different contexts.
- Uncertainty in discussion of model prediction.

IV. CONCLUSION

This study details the newest developments in AI research aimed at digitizing farming to increase food yields. Modern agriculture has been changed by AI technologies that help with things like counting fruits and vegetables, managing water and soil, keeping an eye on crops, identifying weeds, sorting seeds into groups, predicting yields, finding diseases

and gathering crops automatically. The results show that AI has the potential to make farming more accurate, efficient and environmentally friendly. These new technologies help farmers make the best use of their resources, do less work by hand and make decisions based on data, which leads to better productivity and greater resilience against environmental problems for 9 billion people living on the planet. By the end of 2050, using new tools in farming is no longer a choice but a must because it put a lot of stress on farming systems that try to meet rising food needs in a way that doesn't harm the environment. So it is most important to keep researching and developing the integration of AI with agriculture for helping farmers to deal with problems that makes agriculture resilient and build a healthy future.

REFERENCES

- [1] T. Saranya et al., "Engineering applications of artificial intelligence survey paper: A comparative study of deep learning and internet of things for precision agriculture," *Eng. Appl. Artif. Intell.*, 2023.
- [2] M. Altalak et al., "Smart agriculture applications using deep learning technologies: A survey," *Appl. Sci.*, 2022.
- [3] P. Bharman et al., "Deep learning in agriculture: A review," *Asian J. Res. Comput. Sci.*, 2022.
- [4] Pathan SK, Rehman MA, "A review on development of sensor node for precision agriculture," *Proced Eng*, vol. 64, pp. 96–104, 2013.
- [5] Hunt ER, Doraiswamy PC, "Assessing sensor performance and potential of an ultrasonic anemometer–temperature–humidity sensor network," *Comput Electron Agric*, vol. 70, no. 1, pp. 15–26, 2010.
- [6] Wang, Q.; Nuske, S.; Bergerman, M.; Singh, S., "Automated crop yield estimation for apple orchards," *Experimental Robotics*, Springer: Berlin, Germany, pp. 745–758, 2013.
- [7] Li, Y.; Cao, Z.; Lu, H.; Xiao, Y.; Zhu, Y.; Cremers, A.B., "In-field cotton detection via region-based semantic image segmentation," *Comput. Electron. Agric.*, 127, 475–486, 2016.
- [8] Lu, H.; Cao, Z.; Xiao, Y.; Li, Y.; Zhu, Y., "Region-based colour modelling for joint crop and maize tassel segmentation," *Biosyst. Eng.*, 147, 139–150, 2016.
- [9] Schillaci, G.; Pennisi, A.; Franco, F.; Longo, D., "Detecting tomato crops in greenhouses using a vision-based method," *Proceedings of the International Conference Ragusa SHWA2012*, Ragusa Ibla, Italy, 3–6 September 2012; pp. 252–258.
- [10] Linker, R.; Cohen, O.; Naor, A., "Determination of the number of green apples in rgb images recorded in orchards," *Comput. Electron. Agric.*, 81, 45–57, 2012.
- [11] Tabb, A.L.; Peterson, D.L.; Park, J., "Segmentation of apple fruit from video via background modeling," *Proceedings of the 2006 ASABE Annual Meeting*, Oregon, Portland, 2006.
- [12] Laliberte, A.S.; Ripple, W.J., "Automated wildlife counts from remotely sensed imagery," *Wildl. Soc. Bull.*, 31, 362–371, 2003.
- [13] Del Río, J.; Aguzzi, J.; Costa, C.; Menesatti, P.; Sbragaglia, V.; Noguera, M.; Sarda, F.; Manuèl, A., "A new colorimetrically-calibrated automated video-imaging protocol for day-night fish counting at the obsea coastal cabled observatory," *Sensors*, 13, 14740–14753, 2013.
- [14] Ryan, D.; Denman, S.; Fookes, C.; Sridharan, S., "Crowd counting using multiple local features," *Proceedings of the Digital Image Computing: Techniques and Applications*, Melbourne, Australia; pp. 81–88, 2009.
- [15] Kim, J.-W.; Choi, K.-S.; Choi, B.-D.; Ko, S.-J., "Real-time vision-based people counting system for the security door," *Proceedings of the International Technical Conference on Circuits/Systems Computers and Communications*, Phuket Arcadia, Thailand, pp. 1416–1419, 2002.
- [16] Lempitsky, V.; Zisserman, A., "Learning to count objects in images," *Proceedings of the Advances in Neural Information Processing Systems*, Vancouver, BC, Canada, pp. 1324–1332, 2010.
- [17] Giuffrida, M.V.; Minervini, M.; Tsafaris, S.A., "Learning to count leaves in rosette plants," *Proceedings of the British Machine Vision Conference (BMVC)*, Swansea, UK, 2015.
- [18] Koirala A, Zhang Q, "Applications of computer vision for assessing quality of fruits: A review," *Comput Electron Agric*, 153:123–134, 2018.
- [19] Sa I, Ge Z, Dayoub F, Upcroft B, Perez T, McCool C, "Deep-fruits: A fruit detection system using deep neural networks," *Sensors*, 16(8):1222, 2016.
- [20] Anand R, Sahni RK, Kumar SP, Thorat DS, Kumar AK, "Advancement in agricultural practices with use of drones in the context of precision farming," *Glob J Eng Sci*, 11(2):1–7, 2023.
- [21] Bargoti, S.; Underwood, J., "Image segmentation for fruit detection and yield estimation in apple orchards," *arXiv*, 2016, arXiv:1610.08120.
- [22] Xie, W.; Noble, J.A.; Zisserman, A., "Microscopy cell counting with fully convolutional regression networks," *Proceedings of the MICCAI 1st Workshop on Deep Learning in Medical Image Analysis*, Munich, Germany, 5–9 October 2015.
- [23] Zhang, C.; Li, H.; Wang, X.; Yang, X., "Cross-scene crowd counting via deep convolutional neural networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 7–12 June 2015; pp. 833–841.
- [24] Wang, L.; Liu, S.; Lu, W.; Gu, B.; Zhu, R.; Zhu, H., "Laser detection method for cotton orientation in robotic cotton picking," *Trans. Chin. Soc. Agric. Eng*, 30, 42–48, 2014.
- [25] Marston, L. & Cai, X., "An overview of water reallocation and the barriers to its implementation," *Wiley Interdisciplinary Reviews-Water*, 3, 658–677, 2016.
- [26] Salmoral, G., Willaarts, B. A., Garrido, A. & Guse, B., "Fostering integrated land and water management approaches: evaluating the water footprint of a Mediterranean basin under different agricultural land use scenarios," *Land Use Policy*, 61, 24–39, 2017.
- [27] Aquastat (2019). Water use [WWW Document]. Available at: <http://www.fao.org/aquastat/en/overview/methodology/water-use> (accessed 12 December 2019).
- [28] Koscielniak, J., Janowiak, F. & Kurczyk, Z., "Increase in photosynthesis of maize hybrids (Zea mays L.) at suboptimal temperature (15 degrees C) by selection of parental lines on the basis of chlorophyll alpha fluorescence measurements," *Photosynthetica*, 43, 125–134, 2005.
- [29] Nazari, B., Liaghat, A., Akbari, M. R. & Keshavarz, M., "Irrigation water management in Iran: implications for water use efficiency improvement," *Agricultural Water Management*, 208, 7–18, 2018. <https://doi.org/10.1016/j.agwat.2018.06.003>.
- [30] Castanedo, V., Saucedo, H. & Fuentes, C., "Comparison between a hydrodynamic full model and a hydrologic model in border irrigation," *Agrociencia*, 47, 209–223, 2013.
- [31] Lee, T., Kim, S. & Shin, Y., "Development of landsat-based down-scaling algorithm for SMAP soil moisture footprints," *Journal of the Korean Society of Agricultural Engineers*, 60, 49–54, 2018.
- [32] Kim, D.-J., Han, K.-H., Zhang, Y.-S., Cho, H.-R., Hwang, S.-A., Ok, J.-H., Choi, K.-S. & Choi, J.-S., "Evaluation of evapotranspiration in different paddy soils using weighable lysimeter before flooding stage," *Korean Journal of Soil Science & Fertilizer*, 51, 510–521, 2018.
- [33] Deng, S., Yin, Q., Zhang, S., Shi, K., Jia, Z. & Ma, L., "Drip irrigation affects the morphology and distribution of olive roots," *Hortscience*, 52, 1298–1306, 2017.
- [34] Kumar, R., "Ecohydrologic modeling of crop evapotranspiration in wheat (Triticum-aestivum) at sub-temperate and subhumidregion of India," *International Journal of Agricultural and Biological Engineering*, 6, 19–26, 2013.
- [35] Roth, G., Harris, G., Gillies, M., Montgomery, J. & Wigginton, D., "Water-use efficiency and productivity trends in Australian irrigated cotton: a review," *Crop & Pasture Science*, 64, 1033–1048, 2013.
- [36] van Steenberg, F., Basharat, M. & Lashari, B. K., "Key challenges and opportunities for conjunctive management of surface and groundwater in mega-irrigation systems: lower Indus. Pakistan," *Resources-Basel*, 4, 831–856, 2015.

- [37] Githinji, L. J. M., Dane, J. H. & Walker, R. H., "Water-use patterns of tall fescue and hybrid bluegrass cultivars subjected to ET-based irrigation scheduling," *Irrigation Science*, 27, 377–391, 2009.
- [38] Preite, L., Solari, F., & Vignali, G., "Technologies to Optimize the Water Consumption in Agriculture: A Systematic Review," *Sustainability (Switzerland)*, 15, 7, 2023b, MDPI. doi: 10.3390/su15075975.
- [39] Mazzei, D., & Ramjattan, R., "Machine Learning for Industry 4.0: A Systematic Review Using Deep Learning-Based Topic Modelling," *Sensors*, 22, 22, 2022, MDPI. doi: 10.3390/s22228641.
- [40] Liakos, K., Busato, P., Moshou, D., Pearson, S., Bochtis, D., "Machine Learning in Agriculture: A Review," *Sensors*, 18 (8), 2674, 2018. <https://doi.org/10.3390/s18082674>.
- [41] Zhou, Z., Majeed, Y., Diverres Naranjo, G., Gambacorta, E.M.T., "Assessment for crop water stress with infrared thermal imagery in precision agriculture: A review and future prospects for deep learning applications," *Comput. Electron. Agric.*, 182, 106019, 2021. <https://doi.org/10.1016/J.COMPAG.2021.106019>.
- [42] Corell, M., P´erez-L´opez, D., Mart´ın-Palomo, M.J., Centeno, A., Gir´on, I., Galindo, A., Moreno, M.M., Moreno, C., Memmi, H., Torrecillas, A., Moreno, F., Moriana, A., "Comparison of the water potential baseline in different locations. Usefulness for irrigation scheduling of olive orchards," *Agric Water Manag.*, 177, 308–316, 2016.
- [43] Giusti, E., Marsili-Libelli, S., "A Fuzzy Decision Support System for irrigation and water conservation in agriculture," *Environ. Model. Softw.*, 63, 73–86, 2015. <https://doi.org/10.1016/J.ENVSOFT.2014.09.020>.
- [44] Navarro-Hell´ın, H., Mart´ınez-del-Rincon, J., Domingo-Miguel, R., Soto-Valles, F., Torres- S´anchez, R., "A decision support system for managing irrigation in agriculture," *Comput. Electron. Agric.*, 124, 121–131, 2016. <https://doi.org/10.1016/J.COMPAG.2016.04.003>.
- [45] Chandrappa, V.Y., Ray, B., Ashwatha, N., Shrestha, P., "Spatiotemporal modeling to predict soil moisture for sustainable smart irrigation," *Internet of Things*, 21, 100671, 2022.
- [46] Gu, Z., Zhu, T., Jiao, X., Xu, J., Qi, Z., "Neural network soil moisture model for irrigation scheduling," *Comput. Electron. Agric.*, 180, 105801, 2021.
- [47] Kavya, M., Mathew, A., Shekar, P.R., P, S., "Short term water demand forecast modelling using artificial intelligence for smart water management," *Sustain. Cities Soc.*, 95, 104610, 2023. <https://doi.org/10.1016/J.SCS.2023.104610>.
- [48] Srivastava, S., Kumar, N., Malakar, A., Sruti, Choudhury, D., Chit-taranjan Ray, & Roy, T., "A Machine Learning-Based Probabilistic Approach for Irrigation Scheduling," *Water Resource Management*, doi: 10.1007/s11269-024-03746-7, 2024.
- [49] Aly, M. S., Saad, Darwish, M., & Aly, A. A., "High performance machine learning approach for reference evapotranspiration estimation," *Stochastic Environmental Research and Risk Management*, doi: 10.1007/s00477-023-02594-y, 2024.
- [50] Yong, S.L.S., Ng, J.L., Huang, Y.F., Ang, C.K., Ahmad Kamal, N., Mirzaei, M., Najah Ahmed, A., "Enhanced Daily Reference Evapotranspiration Estimation Using Optimized Hybrid Support Vector Regression Models," *Water Resour. Manag.*, 2024.
- [51] Adnan, R.M., Mostafa, R.R., Reza, A., Islam, M.T., Kisi, O., Kuriqi, A., Heddam, S., "Estimating reference evapotranspiration using hybrid adaptive fuzzy inferencing coupled with heuristic algorithms," *Comput. Electron. Agric.*, 191, 106541, 2021.
- [52] Povařanov´a, B., Cistý, M., Bajtek, Z., "Using feature engineering and machine learning in FAO reference evapotranspiration estimation," *J. Hydrol. Hydromech.*, 71, 425–438, 2023. <https://doi.org/10.2478/johh-2023-0032>.
- [53] Youssef, M.A., Peters, R.T., El-Shirbeny, M., Abd-ElGawad, A.M., Rashad, Y.M., Hafez, M., Arafa, Y., "Enhancing irrigation water management based on ET_o prediction using machine learning to mitigate climate change," *Cogent Food and Agriculture*, 10 (1), 2348697, 2024. <https://doi.org/10.1080/23311932.2024.2348697>.
- [54] S. Dhanasekar, T. Jothy Stella, Mani Jayakumar, "Study of Polymer Matrix Composites for Electronics Applications," *Journal of Nanomaterials*, vol. 2022, Article ID 8605099, 7 pages, 2022. <https://doi.org/10.1155/2022/8605099>.
- [55] Dhanasekar S, Malin Bruntha P, Martin Sagayam K, "An Improved Area Efficient 16-QAM Transceiver Design using Vedic Multiplier for Wireless Applications," *International Journal of Recent Technology and Engineering*, vol.8, no. 3, pp.4419-4425, doi: 10.35940/ijrte.C5535.098319, 2019.
- [56] Dhanasekar, S. Bruntha, P.M. Neebha, T.M. Arunkumar, N. Senathipathi, N. Priya, C., "An Area Effective OFDM Transceiver System with Multi-Radix FFT/IFFT Algorithm for Wireless Applications," *In Proceedings of the 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Coimbatore, India, 19–20 March 2021, pp. 551–556.
- [57] Zeynep Unal, "Smart Farming Becomes Even Smarter with Deep Learning-A Bibliographical Analysis," *IEEE Access*, 8, 2020, doi:10.1109/ACCESS.2020.3000175.
- [58] Oksana Mamai, Velta Parsova, Natalya Lipatova, Julia Gazizyanova, and Igor Mamai, "The system of effective management of crop production in modern conditions," *BIO Web of Conferences* 17, 00027, doi:<https://doi.org/10.1051/bioconf/20201700027>, 2020.
- [59] Dhanasekar S, Ramesh J, "VLSI Implementation of Variable Bit Rate OFDM Transceiver System with Multi-Radix FFT/IFFT Processor for wireless applications," *Journal of Electrical Engineering*, vol. 18, 2018- Edition: 1 – Article 18.1.22. ISSN:1582-4594, 2018.
- [60] Dhanasekar S, Ramesh J, "FPGA Implementation of Variable Bit Rate 16 QAM Transceiver System," *International Journal of Applied Engineering Research*, vol.10, pp.26479-26507, 2015.
- [61] Dhanasekar S, Suriavel Rao R S and Victor Du John H, "A Fast and Compact multiplier for Digital Signal Processors in sensor driven smart vehicles," *International Journal of Mechanical Engineering and Technology*, vol. 9, no.10, pp. 157–167, 2018.
- [62] P. Christiansen, L. N. Nielsen, K. A. Steen, R. N. Jørgensen, and H. Karstoft, "DeepAnomaly: Combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field," *Sensors*, vol. 16, no. 11, pp. 2-21, 2016.
- [63] N. Kussul, M. Lavreniuk, S. Skakun, and A. Shelestov, "Deep learning classification of land cover and crop types using remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 778 - 782, 2017.
- [64] F. J. Knoll, V. Czymmek, S. Poczihoski, T. Holtorf, and S. Hussmann, "Improving efficiency of organic farming by using a deep learning classification approach," *Comput. Electron. Agricult.*, vol. 153, pp. 347-356, 2018.
- [65] M. Bah, A. Haane, and R. Canals, "Deep learning with unsupervised data labeling for weed detection in line crops in UAV images," *Remote Sens.*, vol. 10, no. 11, p. 1690, 2018.
- [66] A. Picon, A. Alvarez-Gila, M. Seitz, A. Ortiz-Barredo, J. Echazarra, and A. Johannes, "Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild," *Comput. Electron. Agricult.*, vol. 161, pp. 280290, 2019.
- [67] S. Coulibaly, B. Kamsu-Foguem, D. Kamissoko, and D. Traore, "Deep Neural networks with transfer learning in millet crop images," *Comput. Ind.*, vol. 108, pp. 115 - 120, 2019.
- [68] P. Neavuori, N. Narra, and T. Lipping, "Crop yield prediction with deep convolutional neural networks," *Comput. Electron. Agricult.*, vol. 163, Art. no. 104859, 2019.
- [69] Q. Yang, L. Shi, J. Han, Y. Zha, and P. Zhu, "Deep convolutional neural networks for Rice grain yield estimation at the ripening stage using UAV based remotely sensed images," *Field Crops Res.*, vol. 235, pp. 142 -153, 2019.
- [70] L. Zhong, L. Hu, and H. Zhou, "Deep learning based multi-temporal crop classification," *Remote Sens. Environ.*, vol. 221, pp. 430-443, 2019.
- [71] Bhaskar B.P. et al., "Soil informatics for agricultural land suitability assessment in Seoni district, Madhya Pradesh, India," *Indian J. Agric. Res.* 49(4), 315-320, 2015. 10.5958/0976-058X.2015.00057.8.
- [72] Dickson A.A. et al., "Fertility capability classification based land evaluation in relation to socio-economic conditions of small holder farmers in bayelsa state of Nigeria," *Indian J. Agric. Res.* 36(1), 10-16, 2002..
- [73] Behrens T. et al., "Digital soil mapping using artificial neural networks," *J. Plant Nutr. Soil Sci.* 2005. <https://doi.org/10.1002/jpln.200421414>.

- [74] Grinand C. et al., "Extrapolating regional soil landscapes from an existing soil map: sampling intensity, validation procedures, and integration of spatial context," *Geoderma*.
- [75] Khan, S., Tufail, M., Khan, M. T., Khan, Z. A., and Anwar, S., "Deep learning based identification system of weeds and crops in strawberry and pea fields for a precision agriculture sprayer," *Precis. Agric.*, 22, 1711–1727, doi: 10.1007/s11119-021-09808-9, 2021.
- [76] Yuan, H., Zhao, N., and Chen, M., "Research progress and prospect of field weed recognition based on image processing," *J. Agric. Machinery*, 51, 323–334, 2020.
- [77] Marx, C., Barcikowski, S., Hustedt, M., Haferkamp, H., and Rath, T., "Design and application of a weed damage model for laser-based weed control," *Biosyst. Eng.*, 113, 148–157, doi: 10.1016/j.biosystemseng.2012.07.002, 2012.
- [78] Stepanovic, S., Datta, A., Neilson, B., Bruening, C., Shapiro, C., Gogos, G., et al., "The effectiveness of flame weeding and cultivation on weed control, yield and yield components of organic soybean as influenced by manure application," *Renew. Agric. Food Syst.*, 31, 288–299, doi: 10.1017/S1742170515000216, 2016.
- [79] Kunz, C., Weber, J. F., Peteinatos, G. G., Sökefeld, M., and Gerhards, R., "Camera steered mechanical weed control in sugar beet, maize and soybean," *Precis. Agric.*, 19, 708–720, doi: 10.1007/s11119-017-9551-4, 2018.
- [80] Morin, L., "Progress in biological control of weeds with plant pathogens," *Annu. Rev. Phytopathol.*, 58, 201–223, doi: 10.1146/annurev-phyto-010820-012823, 2020.
- [81] Andert, S., "The method and timing of weed control affect the productivity of intercropped maize (*Zea mays* L.) and bean (*Phaseolus vulgaris* L.)," *Agriculture*, 11, 380, 2021.
- [82] Heap, I., "The international herbicide-resistant weed database," Available at: www.weedscience.org, 2022.
- [83] Cordill, C., and Grift, T. E., "Design and testing of an intra-row mechanical weeding machine for corn," *Biosyst. Eng.*, 110, 247–252, 2011.
- [84] Swain, K. C., Nørremark, M., Jørgensen, R. N., Midtby, H. S., and Green, O., "Weed identification using an automated active shape matching (AASM) technique," *Biosyst. Eng.*, 110, 450–457, doi: 10.1016/j.biosystemseng.2011.09.011, 2011.
- [85] Gasparovic, M., Zrinjski, M., Barkovic, D., and Radocaj, D., "An automatic method for weed mapping in oat fields based on UAV imagery," *Comput. Electron. Agric.*, 173, 105385, doi: 10.1016/j.compag.2020.105385, 2020.
- [86] Bakhshipour, A., and Zareiforush, H., "Development of a fuzzy model for differentiating peanut plant from broadleaf weeds using image features," *Plant Methods*, 16, 153, doi: 10.1186/s13007-020-00695-1, 2020.
- [87] Kazmi, W., Garcia-Ruiz, F., Nielsen, J., Rasmussen, J., and Andersen, H. J., "Exploiting affine invariant regions and leaf edge shapes for weed detection," *Comput. Electron. Agric.*, 118, 290–299, doi: 10.1016/j.compag.2015.08.023, 2015.
- [88] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A., "You only look once: Unified, real-time object detection," in *2016 IEEE conference on computer vision and pattern recognition (CVPR)*, Las Vegas, NV, USA: IEEE, 779–788, 2016.
- [89] Ren, S., He, K., Girshick, R., and Sun, J., "Faster r-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, 39, 1137–1149, doi: 10.1109/TPAMI.2016.2577031, 2017.
- [90] Quan, L., Jiang, W., Li, H., Li, H., Wang, Q., and Chen, L., "Intelligent intra-row robotic weeding system combining deep learning technology with a targeted weeding mode," *Biosyst. Eng.*, 216, 13–31, doi: 10.1016/j.biosystemseng.2022.01.019, 2022.
- [91] Dyrmann, M., Karstoft, H., and Midtby, H. S., "Plant species classification using deep convolutional neural network," *Biosyst. Eng.*, 151, 72–80, doi: 10.1016/j.biosystemseng.2016.08.024, 2016.
- [92] Peteinatos, G. G., Reichel, P., Karouta, J., Andujar, D., and Gerhards, R., "Weed identification in maize, sunflower, and potatoes with the aid of convolutional neural networks," *Remote Sens.*, 12, 4185, doi: 10.3390/rs12244185, 2020.
- [93] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., et al., "ImageNet Large scale visual recognition challenge," *Int. J. Comput. Vis.*, 115, 211–252, 2015.
- [94] Deepak Sinwar, Vijaypal Singh Dhaka, Manoj Kumar Sharma, and Geeta Rani. 2020. AI-Based Yield Prediction and Smart Irrigation. 2, (2020), 155–180.
- [95] Monika Agarwal, Geeta Rani, and Vijaypal Singh Dhaka. 2020. Optimized contrast enhancement for tumor detection. *Int. J. Imaging Syst. Technol.* 30, 3 (2020), 687–703. DOI:<https://doi.org/10.1002/ima.22408>.
- [96] Nidhi Kundu, Geeta Rani, and Vijaypal Singh Dhaka. 2020. Machine Learning and IoT based Disease Predictor and Alert Generator System. *Proc. 4th Int. Conf. Comput. Methodol. Commun. ICCMC 2020 Iccmc* (2020), 764–769.
- [97] Tianwei Ren, Zhe Liu, Lin Zhang, Diyou Liu, Xiaojie Xi, Yanghui Kang, Seeds Classification and Quality Testing Using Deep Learning and YOLO v5, *DSMLAI'21, August 9-12, 2021*
- [98] Kantip Kiratiratanapruk and Wasin Sintupinyo. 2011. Color and texture for corn seed classification by machine vision. *2011 Int. Symp. Intell. Signal Process. Commun. Syst. "The Decad. Intell. Green Signal Process. Commun. ISPACS 2011* (2011), 7–11.
- [99] Tony Vaiciulis. SVM. Retrieved from <https://www-cdf.fnal.gov/physics/statistics/recommendations/svm/svm.html>
- [100] Jun Zhang, Limin Dai, and Fang Cheng. 2021. Corn seed variety classification based on hyperspectral reflectance imaging and deep convolutional neural network. *J. Food Meas. Charact.* 15, 1 (2021), 484–494.
- [101] Te Ma, Satoru Tsuchikawa, and Tetsuya Inagaki. 2020. Rapid and non-destructive seed viability prediction using near-infrared hyperspectral imaging coupled with a deep learning approach. *Comput. Electron. Agric.* 177, April (2020).
- [102] Pengcheng Nie, Jinnuo Zhang, Xuping Feng, Chenliang Yu, and Yong He. 2019. Classification of hybrid seeds using near-infrared hyperspectral imaging technology combined with deep learning. *Sensors Actuators, B Chem.* 296, May (2019), 126630.
- [103] Chao Ni, Dongyi Wang, Robert Vinson, Maxwell Holmes, and Yang Tao. (2019) Automatic inspection machine for maize kernels based on deep convolutional neural networks. *Biosyst. Eng.* 178, 131–144.
- [104] Sheng Huang, Xiaofei Fan, Lei Sun, Yanlu Shen, and Xuesong Suo. (2019). Research on Classification Method of Maize Seed Defect Based on Machine Vision. *J. Sensors* 2019, 1 DOI:<https://doi.org/10.1155/2019/2716975>
- [105] Lei Pang, Sen Men, Lei Yan, and Jiang Xiao. 2020. Rapid Vitality Estimation and Prediction of Corn Seeds Based on Spectra and Images Using Deep Learning and Hyperspectral Imaging Techniques. *IEEE Access* 8, (2020), 123026–123036.
- [106] Ferhat Kurtulmuş. 2020. Identification of sunflower seeds with deep convolutional neural networks. *J. Food Meas. Charact.* 0123456789 (2020).
- [107] Guoyang Zhao, Longzhe Quan, Hailong Li, Huaiqu Feng, Songwei Li, Shuhan Zhang, and Ruiqi Liu. 2021. Real-time recognition system of soybean seed full-surface defects based on deep learning. *Comput. Electron. Agric.* 187, May (2021), 106230. DOI:<https://doi.org/10.1016/j.compag.2021.106230>.
- [108] Shaolong Zhu, Jinyu Zhang, Maoni Chao, Xinjuan Xu, Puwen Song, Jinlong Zhang, and Zhongwen Huang. 2020. A rapid and highly efficient method for the identification of soybean seed varieties: Hyperspectral images combined with transfer learning. *Molecules* 25, 1 (2020).
- [109] Pablo M. Granitto, Pablo F. Verdes, and H. Alejandro Ceccatto. (2005). "Large-scale investigation of weed seed identification by machine vision", *Comput. Electron. Agric.* 47, 1 (2005), 15–24.
- [110] Tom M. Mitchell. 2019. Machine Learning. DOI:<https://doi.org/10.1109/ICDAR.2019.00014>
- [111] F. Abbas, H. Afzaal, A.A. Farooque, S. Tang (2020) "Crop yield prediction through proximal sensing and machine learning algorithms", *Agronomy*, 10 (7) (2020), p. 1046
- [112] M. Kuradusenge, E. Hitimana, D. Hanyurwimfura, P. Rukundo, K. Mtonga, A. Mukasine, C. Uwitonze, J. Ngabonziza, A. Uwamahoro (2023) "Crop yield prediction using machine learning models: case of Irish Potato and Maize", *Agriculture*, 13 (1) (2023), p. 225

- [113] K. Gavahi, P. Abbaszadeh, H. Moradkhani (2021) DeepYield: a combined convolutional neural network with long short-term memory for crop yield forecasting *Expert Syst. Appl.*, 184 (2021), Article 115511.
- [114] D. Elavarasan, P.D.R. Vincent (2021) "A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters", *J. Ambient Intell. Humaniz. Comput.*, 12 (2021), pp. 10009-10022
- [115] A. Rajagopal, S. Jha, M. Khari, S. Ahmad, B. Alouffi, A. Alharbi (2021) "A novel approach in prediction of crop production using recurrent cuckoo search optimization neural networks", *Appl. Sci.*, 11 (21) (2021), p. 9816
- [116] Palaniappan, M. & Annamalai, M. (2019). Advances in signal and image processing in biomedical applications. 10.5772/intechopen.88759.
- [117] Mohanty, S. P., Hughes, D. P., Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Front. Plant Sci.* 7 (September).
- [118] Anjna, Sood, M., Singh, P. K. (2020). Hybrid system for detection and classification of plant disease using qualitative texture features analysis. *Proc. Comput. Sci.* 167 (2019), 1056–1065. doi: 10.1016/j.procs.2020.03.404
- [119] Genaev, M. A., Skolotneva, E. S., Gulyaeva, E. I., Orlova, E. A., Bechtold, N. P., Afonnikov, D. A. (2021). Image-based wheat fungi diseases identification by deep learning. *Plants* 10 (8), 1–21. doi: 10.3390/plants10081500
- [120] Kaur, L., Sharma, S. G. (2021). Identification of plant diseases and distinct approaches for their management. *Bull. Natl. Res. Centre* 45 (1), 1–10. doi: 10.1186/s42269-021-00627-6.
- [121] Siddiqua, A., Kabir, M. A., Ferdous, T., Ali, I. B., and Weston, L. A. (2022).Evaluating plant disease detection mobile applications: Quality and limitations.*Agronomy* 12 (8), 1869. doi: 10.3390/agronomy12081869
- [122] Wang, B. (2022). Identification of crop diseases and insect pests based on deeplearning. *Sci. Program.* 2022, 1–10. doi: 10.1155/2022/9179998
- [123] Kurmi, Y., and Gangwar, S. (2022). A leaf image localization based algorithm fordifferent crops disease classification. *Inf. Process. Agric.* 9 (3), 456–474. doi: 10.1016/j.inpa.2021.03.001
- [124] Yoosfzadeh-Najafabadi, M., Earl, H. J., Tulpan, D., Sulik, J., and Eskandari, M (2021). Application of machine learning algorithms in plant breeding: Predicting yieldfrom hyperspectral reflectance in soybean. *Front. Plant Sci.* 11 (January). doi: 10.3389/fpls.2020.624273
- [125] Cedric, L. S., Adoni, W. Y. H., Aworka, R., Zoueu, J. T., Mutombo, F. K., Krichen, M.,et al. (2022). Crops yield prediction based on machine learning models: Case of WestAfrican countries. *Smart Agric. Technol.* 2 (March), 100049. doi: 10.1016/j.atech.2022.100049
- [126] Domingues, T., Brandão, T., and Ferreira, J. C. (2022). Machine learning fordetection and prediction of crop diseases and pests: A comprehensive survey.*Agriculture* 12 (9), 1350. doi: 10.3390/agriculture12091350;
- [127] Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., and Stefanovic, D. (2016).Deep neural networks based recognition of plant diseases by leaf image classification.*Comput. Intell. Neurosci.* 2016, 1–12. doi: 10.1155/2016/3289801
- [128] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O.,et al. (2021). Review of deep learning: concepts, CNN architectures, challenges,applications, future directions. *J. Big Data* 8, 1–74. doi: 10.1186/s40537-021-00444-8
- [129] Dhaka, V. S., Meena, S. V., Rani, G., Sinwar, D., Ijaz, M. F., and Wozniak, M. (2021).A survey of deep convolutional neural networks applied for prediction of plant leafdiseases. *Sensors* 21 (14), 4749. doi: 10.3390/s21144749
- [130] Saeed Khaki and Lizhi Wang. (2019) Crop yield predictionusing deep neural networks. *Frontiers in plant science*,10:621.
- [131] Ekaansh Khosla, Ramesh Dharavath, and RashmiPriya. (2020) Crop yield prediction using aggregated rainfall basedmodular artificial neural networks and supportvector regression. *Environment, Development andSustainability*, 22:5687–5708.
- [132] Maya Gopal PS (2019) Performance evaluation of best featuresubsets for crop yield prediction using machine learningalgorithms. *Applied Artificial Intelligence*, 33(7):621–642.
- [133] The Food and Agriculture Organization (FAO).FAOSTAT. https://www.fao.org/faostat/en/#rankings/countries%20_by_com%20modity_exports, 2023.[Online; accessed Jan. 20, 2023].
- [134] Francisco Raimundo, Andre Gloria, and PedroSebastiao. (2021) Prediction of weather forecast for smartagriculture supported by machine learning. In*2021 IEEE World AI IoT Congress (AIoT)*, pages0160–0164. IEEE.
- [135] A Suruliandi, G Mariammal, and SP Raja. (2021) Cropprediction based on soil and environmentalcharacteristics using feature selection techniques.*Mathematical and Computer Modelling of DynamicalSystems*, 27(1):117–140.
- [136] Pushpa Mohan and Kiran Kumari Patil. (2017) Cropproduction rate estimation using parallel layerregression with deep belief network. In *2017International Conference on Electrical, Electronics,Communication, Computer, and OptimizationTechniques (ICEEC-COT)*, pages 168–173. IEEE.
- [137] Ms Kavita and Pratistha Mathur. (2020) Crop yieldestimation in india using machine learning. In *2020IEEE 5th International Conference on ComputingCommunication and Automation (ICCCA)*, pages 220–224. IEEE.
- [138] Saiteja Kunchakuri, S Pallerla, Sathya Kande, andNageswara Rao Sirisala. (2021) An efficient crop yieldprediction system using machine learning algorithm. In*4th Smart Cities Symposium (SCS 2021)*, volume 2021, pages 120–125. IET.
- [139] Madhuri Shripathi Rao, Arushi Singh, NV SubbaReddy, and Dinesh U Acharya. (2022) Crop prediction usingmachine learning. In *Journal of Physics: ConferenceSeries*, volume 2161, page 012033. IOP Publishing.
- [140] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and JianSun. (2016) Deep residual learning for image recognition. In*Proceedings of the IEEE conference on computer visionand pattern recognition*, pages 770–778.
- [141] CH Vanipriya, Subhash Malladi, Gaurav Gupta, et al. (2021)Artificial intelligence enabled plant emotion xpresser inthe development hydroponics system. *Materials Today:Proceedings*, 45:5034–5040.
- [142] Abhishek Tomar, Gaurav Gupta, Waleed Salehi,CH Vanipriya, Nagesh Kumar, and BrijbhushanSharma. (2022) A review on leaf-based plant disease detectionsystems using machine learning. *Recent Innovations in Computing: Proceedings of ICRIC 2021, Volume 1*, pages 297–303.
- [143] [143]Song, Z., Zou, S., Zhou, W.et al.Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning.*Nat Commun* 11, 4294 2020. <https://doi.org/10.1038/s41467-020-18147-8>.
- [144] [144]Kabir, H., Wu, J., Dahal, S.et al.Automated estimation of cementitious sorptivity via computer vision. *Nat Commun* 15, 9935 2024. <https://doi.org/10.1038/s41467-024-53993-w>

LMS-YOLO11n: A Lightweight Multi-Scale Weed Detection Model

YaJun Zhang¹, Yu Xu², Jie Hou³, YanHai Song⁴,

Department School of Software, Xinjiang University, Urumqi, China^{1,2,4}
Guangzhou Xinhe Information Technology Corporation, Guangzhou, China³

Abstract—With the advancement of precision agriculture, efficient and accurate weed detection has emerged as a pivotal task in modern crop management. Current weed detection methods face dual challenges: inadequate extraction of detailed features and edge information, coupled with the necessity for real-time performance. To address these issues, this paper proposes a lightweight multi-scale weed detection model based on YOLOv11n (You-only-look-once-11). Our approach incorporates three innovative components: (1) A fast-gated lightweight unit combined with C3K2 to enhance local and global interaction capabilities of weed features. (2) An adaptive hierarchical feature fusion network based on HSFPN, which improves the extraction of weed edge information. (3) A lightweight group convolution detection head module that captures multi-scale feature details while maintaining a lightweight structure. Experimental validation on two public datasets, CottonWeedDet3 and CottonWeed2, demonstrates that our model achieves an mAP50 improvement of 2.5% on CottonWeedDet3 and 1.9% on CottonWeed2 compared to YOLOv11n, with a 37% reduction in parameters and a 26% decrease in computational effort.

Keywords—You-only-look-once-11; weed; lightweight; group convolution

I. INTRODUCTION

Modern agriculture faces numerous challenges that hinder productivity and sustainable development. Weeds are a major threat, directly impacting crop yield and food security. Weeds compete with crops for light, water, and soil nutrients, spreading diseases and pests, significantly reducing crop yield, and causing economic losses [1-2]. Selective herbicides and manual weeding are the two major weed management techniques used today; the former entails evenly applying herbicides throughout fields. This method results in significant waste, as most herbicides are sprayed on crops or bare soil, rather than directly on the weeds. Additionally, excessive herbicide use harms the ecosystem. Manual weeding, on the other hand, is costly and difficult to scale for large-scale agricultural operations.

With the development of AI technology, precision agriculture offers a solution to these problems [3-4], with the key first step being the accurate and rapid detection of weed locations [5]. Therefore, in-depth research on weed detection technology is crucial for the development of precision agriculture, contributing to the future efficiency, precision, and sustainability of farming.

Early weed detection methods were mostly based on machine learning, such as Kumar and Prema's [6] Wrapped Curve Transform Angle Texture Pattern extraction method, which improved weed identification accuracy in fields. Sujaritha et al.

[7] proposed a circular leaf pattern extraction method based on morphological operations, combined with rotational invariance and wavelet decomposition, enabling automatic weed and crop recognition and efficient removal in sugarcane fields. However, these methods struggle to handle challenges such as the complexity of field environments, weed species diversity, and lighting changes, resulting in poor detection performance and instability, which limits their application in diverse environments. In contrast, deep learning uses convolutional neural networks to extract both global and local features of weeds, compensating for the shortcomings of machine learning in feature extraction. As a result, deep learning-based weed detection has become the mainstream method.

While deep learning-based methods outperform machine learning in terms of accuracy, they struggle to meet the lightweight requirements of edge devices, making weed detection on edge devices a new direction in object detection. This study faces challenges such as high similarity between weed species, occlusion issues affecting detection accuracy, and the deployment limitations of edge devices. To address these challenges, this paper proposes the LMS-YOLO11n weed detection method based on YOLO11n. The main contributions of the LMS-YOLO11n model are as follows:

- 1) To meet the demands of detail extraction and real-time performance in weed detection, this paper proposes the lightweight multi-scale feature extraction module FastGLU, combined with CGLU's convolutional gating mechanism and FasterNet's lightweight characteristics. It extracts key channel information through partial convolution (PConv) and uses CGLU to enhance the interaction between local and global features, reducing computational costs while achieving efficient and diverse feature extraction.
- 2) To address the challenge of weed edge information extraction, this paper designs the adaptive hierarchical feature fusion network (AHFPN). By combining the ideas of HSFPN and PAN, the feature fusion mechanism is improved to enhance sensitivity and capability in edge information extraction, optimize the interaction and fusion of multi-scale features, and improve adaptability to weed diversity and various growth stages, while reducing computational burden.
- 3) To meet the real-time requirements of weed detection, this paper introduces the lightweight group convolution detection head (LGCD) module. By incorporating group convolution into the position regression branch, the computational load and parameter count are significantly reduced, and kernel size optimization improves the ability

to capture multi-level feature details, balancing feature extraction richness with model efficiency to meet the deployment requirements of edge devices.

This paper has the following structure: Section II examines the state of domestic and international weed detection research; Section III provides further details about the LMS-YOLO11n approach; experiments in Section IV confirm the model's generalization performance; and a summary of the work and recommendations for future research are provided in Section V.

II. RELATED WORK

Object detection methods can be broadly categorized into single-stage and two-stage models. Two-stage detection models generate candidate regions quickly and refine them in a second processing stage. Typical models include RCNN [8] and Fast-RCNN [9]. Zhang [10] and colleagues successfully detected weeds and soybeans in complex backdrops by optimizing the Faster R-CNN method with VGG19-CBAM as the backbone network, achieving successful detection of soybeans and weeds in complex backgrounds. Ozcan et al. [11] compared the performance of single-stage and two-stage CNN models in precision agriculture and found that Faster R-CNN Inception v2 offers higher accuracy. However, when training and inference time are critical, the SSD MobileNet v2 model significantly improves accuracy with increased training data. Li et al. [12] proposed an improved Faster R-CNN model for automatic detection of hydroponic lettuce seedlings, achieving an accuracy of 86.2% through enhancement techniques, outperforming models like RetinaNet, SSD, Cascade RCNN, and FCOS. Although two-stage detection methods are generally more accurate than single-stage methods, they require significant computational resources, making them difficult to deploy on mobile devices. Moreover, the longer detection time limits their ability to meet detection in real-time requirements.

Due to the limitations of two-stage detection, there is growing interest in single-stage detection methods, exemplified by SSD [13] and YOLO [14-18]. Unlike two-stage methods, single-stage detection integrates region proposal, classification, and regression into a single network, significantly improving speed and efficiency. Chen et al. [19] proposed the YOLO-sesame model, an improved YOLOv4 variant that incorporates Local Importance Pooling (LIP) and SE modules to enhance feature extraction. The model also uses an Adaptive Spatial Feature Fusion (ASFF) structure to optimize the detection of objects of varying sizes, improving both real-time performance and accuracy in sesame field weed detection. Hong et al. [20] presented an enhanced YOLOv5 algorithm for effective asparagus identification in intricate settings. The model incorporates Coordinate Attention (CA) in the backbone network to emphasize growth features of asparagus and replaces PANet with BiFPN to enhance feature propagation and reuse, significantly improving support for intelligent mechanical harvesting under various weather conditions. A network of convolutional neuron models called RIC-Net, which combines residual structures with Inception, was proposed by Zhao et al. [21]. The model replaces MLP layers with 1D convolutions for optimized feature detection and integrates CBAM modules with weighted operations to highlight diseased areas, improving classification accuracy for leaf diseases in maize, potatoes, and tomatoes.

Song et al. [22] developed an improved YOLOv5 algorithm by replacing the backbone with MobileNetv2 to reduce model complexity. ECANet attention mechanisms were introduced to enhance detailed feature extraction for soybean leaves, and CIOU_Loss + DIOU_NMS was used to improve accuracy and robustness, particularly for dense occlusion and small object detection in precision agriculture spraying. Zhang et al. [23] proposed CCCS-YOLO, an improved YOLOv5-based algorithm. The model integrates Faster_Block into YOLOv5s's C3 module to create C3_Faster, simplifying the network structure and enhancing detection. It improves the convolutional block in the head for better target-background differentiation, replaces the neck's upsampling module with the lightweight CARAFE module for small object detection and contextual information fusion, and uses Soft-NMS-EIoU to enhance detection accuracy in dense scenarios. Guo et al. [24] proposed LW-YOLOv8n, a lightweight weed detection model. The model integrates SERMAttention with SE and SRM modules to capture global information, incorporates lightweight Context Guided Blocks in C2f layers to enhance local and contextual feature learning, and introduces an improved BiFPN network in the neck for weighted multi-scale feature fusion. This method is appropriate for edge devices with limited resources as it lowers parameters and complexity while preserving excellent detection accuracy. Fan et al. [25] presented YOLO-WDNet, a model for lightweight weed identification. It replaces CSP-Darknet53 with ShuffleNet v2 as the backbone to reduce parameters and complexity, designs a Parallel Hybrid Attention Mechanism (PHAM) to focus on regions of interest, improves BiFPN in the neck for multi-scale and overlapping plant feature recognition, and proposes an EIOU loss function to enhance detection accuracy in dense scenarios.

Despite significant advancements, a gap persists in the development of single-stage detection models that combine high accuracy with sufficient lightness and efficiency for deployment on resource-limited edge devices. The work presented in this paper endeavors to address this gap by introducing a novel, lightweight single-stage detection model. This model integrates advanced convolutional gating mechanisms, optimized feature fusion strategies, and a streamlined detection head leveraging grouped convolutions, all tailored to elevate detection accuracy and efficiency specifically within agricultural applications.

III. METHODOLOGY

A. YOLO11 Principle

YOLO11, the latest model in the YOLO series, was released by Ultralytics in 2024. Based on structural complexity and size, YOLO11 is available in five versions: YOLO11n, YOLO11s, YOLO11m, YOLO11l, and YOLO11x. YOLO11n, the version with the smallest computational and parameter requirements, is designed for weed detection scenarios that demand real-time performance and limited computational resources. Therefore, this study selects YOLO11n as the baseline model. YOLO11n consists of three main components: Backbone, Neck, and Head. The Backbone replaces the C2f module from YOLOv8 with the latest C3K2 module, significantly improving feature extraction efficiency. The Neck continues to use the FPN+PAN structure for feature fusion, while the Head incorporates depthwise separable convolutions, significantly reducing computation and parameter requirements. Although

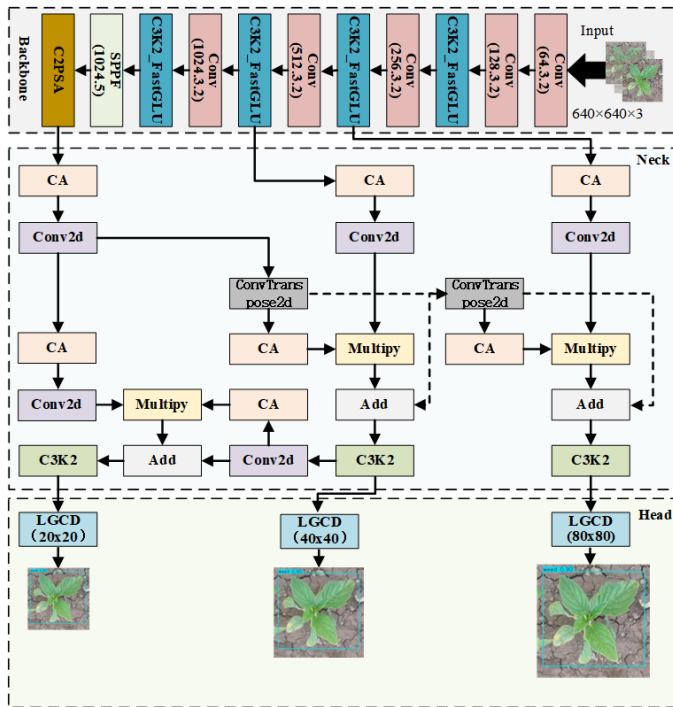


Fig. 1. Structure of LMS-YOLO11n.

YOLO11n is the most lightweight version, its computational complexity remains 6.3GFLOPs, and its parameter size is 2.58MB, which still poses challenges for real-time detection and edge computing deployment.

B. Lightweight Multi-Scale Weed Detection Model

In deep learning-based weed detection tasks, the scale and complexity of the model directly determine its practical effectiveness. Although YOLO11 surpasses many mainstream object detection models in speed, it contains significant redundant features. These redundant features are primarily generated by convolutional computations in the backbone network, consuming substantial computational resources, increasing complexity, and reducing inference speed. Additionally, the similar textures of crop seedlings and weeds, coupled with multi-scale features, make YOLO11 less effective at extracting features under complex lighting conditions. To address the need for lightweight models and real-time detection, while enhancing the extraction of fine-grained weed features and edge information, this study proposes the LMS-YOLO11n model based on YOLO11n. LMS-YOLO11n integrates the C3K2_FCGLU module into the YOLO11n framework to replace the original C3K2 module, enabling more efficient weed feature extraction. Furthermore, by introducing the AHFPN designed based on HSFPN [26], the neck network is optimized to improve the recognition and fusion of multi-scale overlapping plant features. Finally, the LGCD module, based on grouped convolution [27], is used to refine the Head, enhancing multi-scale information capture while reducing parameters and computation. Fig. 1 displays the LMS-YOLO11n structure with an input picture size of 640 x 640 x 3.

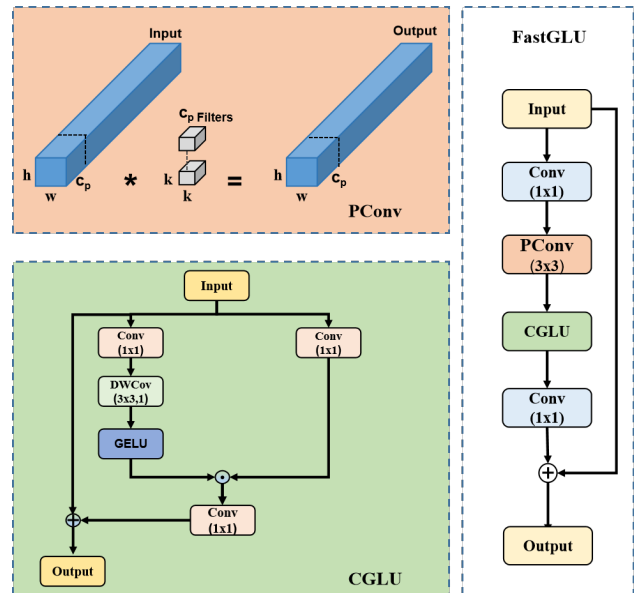


Fig. 2. Structure of CGLU, PConv and FastGLU.

C. C2f_FastGLU

To address the requirements for detail extraction and real-time processing in weed detection, this paper proposes a Fast Gated Lightweight Unit (FastGLU), as shown in Fig. 2. FastGLU captures fine-grained feature information, enhancing the model's ability to perceive image details, expand the receptive field, and extract local features. Additionally, it excels in optimizing multi-channel information usage, reducing parameters and computational costs, maintaining gradient flow, and enhancing spatial feature extraction. This allows the model to efficiently handle weeds of varying sizes and shapes. FasterNet [28] introduced the concept of Partial Convolution. PConv is a convolutional method designed to improve data processing efficiency and reduce memory overhead. It applies standard convolutions to a subset of input channels to effectively extract spatial features while omitting convolutions on other channels, thereby reducing computational and memory demands. Specifically, it selects the first and last consecutive c_p channels as representatives of the input feature map, assuming the input and output feature maps have the same number of channels. This design not only simplifies computation but also optimizes memory access efficiency, enabling effective feature representation. By applying convolutions only to a subset of input channels to extract spatial features while ignoring others, its computational complexity is defined in Eq. (1).

$$\begin{aligned}
 F_{Conv} &= h \times w \times k^2 \times c^2 \\
 F_{PConv} &= h \times w \times k^2 \times c_p^2
 \end{aligned}
 \quad (1)$$

In this equation, h and w represent the height and width of the feature map, k denotes the kernel size, c is the number of input feature map channels, and c_p represents the selected input channels used for spatial feature extraction in the PConv operation. In this study, c_p is set to 1/4 of c , reducing the computational cost of PConv to just 1/16 that of a standard convolution.

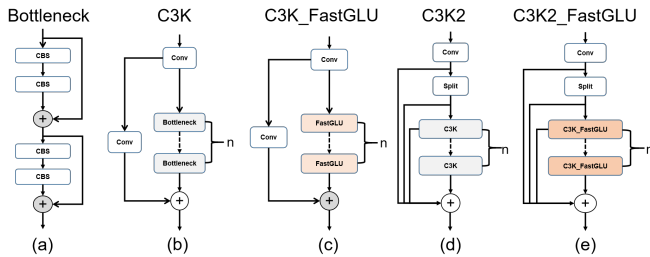


Fig. 3. (a) Bottleneck; (b) C3K; (c) C3K_FastGLU; (d) C3K2; (e) C3K2_FastGLU structure.

The Faster Block accelerates network processing by reducing computational and memory access demands. Its structure consists of a PConv layer followed by two pointwise convolution (PWConv) layers. However, the Faster Block has a limited receptive field, and PConv processes only part of the channel information, which hinders fine-grained weed feature extraction. The Gated Linear Unit (GLU) is an activation mechanism designed to enhance the extraction of complex features, initially used in language processing and sequence modeling tasks. Currently, GLU [29] has evolved into several variants, including the Gated Recurrent Unit, Depthwise Separable GLU, FFN with SE module, and Convolutional Gated Linear Unit [30]. In this study, CGLU is integrated into PConv to enhance fine-grained local feature extraction and optimize the interaction between local and global feature information. CGLU first employs two parallel 1x1 convolutions for per-channel control, with one feature map further processed by a 3x3 depthwise separable convolution to capture local features. These features are then fed into the Gated Linear Unit. A portion of the features is activated by the GELU function to serve as a gating signal, which multiplies with another feature set to enable channel attention control, enhancing feature selection and emphasis.

The primary structure of the C3K2 module in YOLO11 Fig. 3(d) is based on the C3K module. The C3K module Fig. 3(b) consists of standard convolutions and Bottleneck units. The Bottleneck unit Fig. 3(a) is composed of CBS modules. CBS is a fundamental convolutional unit comprising convolution operations, batch normalization, and an activation function. The proposed C3K2_FastGLU module Fig. 3(e) replaces the C3K module in C3K2 with C3K_FastGLU. The backbone network faces several bottleneck issues. Introducing C3K2_FastGLU effectively reduces computational cost, significantly improves multi-scale feature extraction, alleviates information transmission bottlenecks, and maintains efficient feature representation and generalization in lightweight designs. Therefore, in the YOLO11n backbone, the C3K2 modules in the P2, P3, P4, and P5 layers are replaced with C3K2_FastGLU.

D. AHFPN

To meet the deployment requirements of weed detection on edge devices, this study uses HSFPN to fuse extracted features. HSFPN consists of a Channel Attention (CA) module and a Semantic Feature Fusion (SFF) module, as illustrated in Fig. 4.

The Channel Attention (CA) module applies average pooling and max pooling to each channel's features, extracting the most relevant and average information for each channel. The pooled average and maximum results are combined, and the Sigmoid function calculates the weight for each channel. Finally, the weights are multiplied by the corresponding feature maps to filter redundant data. Additionally, a 1x1 convolution is used to adjust channel dimensions to 256, ensuring compatibility across different scales.

The Semantic Feature Fusion (SFF) module employs weights from higher-level characteristics to selectively integrate key semantic data derived from lower-level characteristics. The process includes: 1) applying a 3x3 transposed convolution with a stride of 2 to process higher-level features; 2) aligning the transposed higher-level features' dimensions with those of the lower-level features by the use of bilinear interpolation; 3) employing the CA module to convert higher-level features into weights; and 4) combining the optimized lower-level features with higher-level characteristics to improve the depiction of aspects. The specific definition is given in Eq. (2):

$$\begin{aligned} f_{att} &= BL(T - Conv(f_{high})) \\ f_{out} &= f_{low} * CA(f_{att}) + f_{att} \end{aligned} \quad (2)$$

Include among these $f_{high} \in R^{C \times H \times W}$, $f_{low} \in R^{C \times H_1 \times W_1}$. C is the number of channels, H and W are the height and width of the feature map, BL is the bilinear interpolation, and T is the transposed convolution.

However, HSFPN has limited capability in perceiving edge information, making it difficult to distinguish between early-stage weed growth and crops. To address this, we propose an Adaptive Hierarchical Feature Fusion Network (AHFPN), which significantly enhances the model's ability to handle multi-scale weed targets and enriches feature representations to improve detection accuracy across different weed growth stages. This module combines the concepts of HSFPN and PAN [31], with improvements tailored to different weed targets. The main process includes:

- 1) Adding a Conv2d layer to the P4 output to enhance the extraction of high-level semantic information.
- 2) The output of the P5 layer undergoes CA and a 1x1 Conv2d operation to extract key channel information and adjust weights.
- 3) The processed high-level features are weighted, multiplied by previously fused features, and then added together.
- 4) Finally, the fused features pass through the C3k2 module to further enhance feature extraction, resulting in more refined high-level semantic features.

E. LGCD

The detection head in YOLO11 identifies object locations and categories from the feature map. The process is as follows: In the position regression branch, two standard convolutions are used for feature fusion, followed by a convolution layer for location prediction; In the classification branch, depthwise separable convolutions [32] are used for feature fusion, followed by pointwise convolutions for channel-wise information

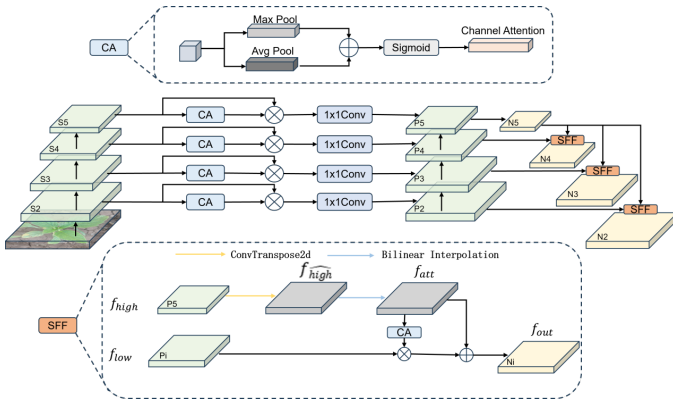


Fig. 4. HSPFN structure and diagram of CA and SF modules.

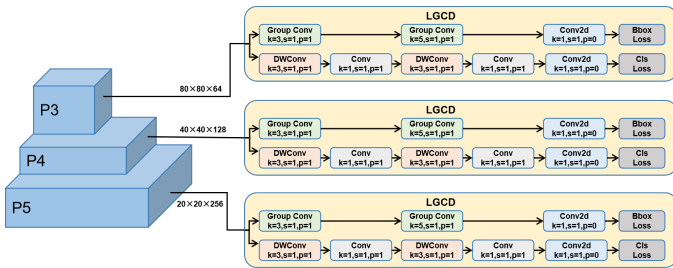


Fig. 5. LGCD structure.

interaction. Finally, a convolution layer performs classification prediction, with a Softmax activation function generating category probabilities. Although YOLO11 is significantly lighter compared to previous YOLO models, it still does not fully meet the real-time and multi-scale requirements for weed detection. Therefore, we propose the lightweight Grouped Convolution Detection Head module, as shown in Fig. 5.

The LGCD module is based on the concept of grouped convolutions, which improves the YOLO11 detection head. In this study, we replace the first two standard convolutions in the position regression branch of the YOLO11n detection head with grouped convolutions to reduce computation and parameter count, achieving the model. To minimize feature loss, we modify the kernel of the second grouped convolution to 5x5, ensuring lightweight while extracting multi-scale weed features and improving detection accuracy.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Datasets

CottonWeedDet3 [33] contains 848 RGB images captured in cotton fields in the southern United States, covering three common weed categories: Carpetweed, Morningglory, and Palmer Amaranth. The images capture various angles and natural lighting conditions, ensuring the dataset's diversity and relevance for application. The dataset construction process includes image acquisition, preprocessing, bounding box annotation, data cleaning, format conversion, and data augmentation. Experts manually annotated the images using the SuperAnnotate platform and converted them to the VIA format for further use. To improve data quality, low-quality annotations and out-of-focus areas smaller than 200x200 pixels

were cleaned, and erroneous labels were corrected. The final dataset contains 848 images and 1,532 bounding boxes, split into training, validation, and test sets in a 7:1:2 ratio. Fig. 6 shows an example image of the CottonWeedDet3 dataset.



Fig. 6. Example plot of CottonWeedDet3 dataset.

The CottonWeed2 [34] dataset contains 570 images, labeled into two categories: weeds and cotton. The weed category includes various plants, such as Wormwood, Common Sunflower, Chicory, Caltrop, Ginkgo, Castor Bean, Crabgrass, False Sea Purslane, and Amaranthus, reflecting the diversity of weed species. This diversity provides rich and complex samples for model training. Using digital cameras or cell-phones, the photos were taken from actual cotton fields in India and stored in.jpg format, which makes them extremely useful. To standardize data processing and adapt to model input, all images were resized to 416 x 416 pixels. This processing method facilitates model handling while reducing computational complexity. The dataset is divided into training, validation, and test sets in a 6:2:2 ratio. Fig. 7 shows an example image of the CottonWeed2 dataset.



Fig. 7. Example plot of CottonWeed2 dataset.

The CottonWeedDet12 dataset [35], collected at Mississippi State University Research Farm, includes 5,648 images and 9,370 bounding box annotations of 12 cotton weed species. Captured under natural light between February and October 2021 using smartphones or handheld cameras, the dataset spans diverse growth stages, lighting, weather, and field conditions to ensure complexity and diversity. The dataset was re-divided using a custom script into training, validation, and test sets at a 7:1:2 ratio. Fig. 8 presents example images from the public CottonWeedDet12 dataset.

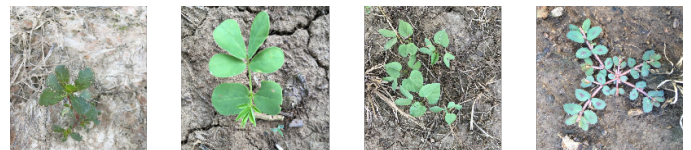


Fig. 8. Example plot of CottonWeedDet12 dataset.

B. Experimental Configuration

The experiment was conducted on a Windows 11 operating system. The model architecture includes Python 3.8.19,

PyTorch 2.1.1, and TorchVision 0.16.1, with PyCharm as the integrated development environment. The CPU is an Intel i5-12400F, and the GPU is an Nvidia GeForce RTX 4060Ti (16GB) with 4352 CUDA cores, running CUDA version 12.1.

C. Experimental Parameter Setting and Evaluation Indicators

The model was trained for 300 epochs with a batch size of 8. The AdamW optimizer was used, with an initial learning rate of 0.01 and a momentum of 0.937. The input image size was 640×640. Multiple evaluation metrics were used to assess the effectiveness of this study, including Precision (P), Recall (R), Mean Average Precision (mAP), Parameters (Params), and Giga Floating Point Operations per Second (GFLOPs). The model's recognition performance was measured using IOU thresholds of 0.50 and 0.50:0.95. Params were used to measure the model's parameter count, and GFLOPs were used to measure its computational complexity. The specific definition is given in Eq. (3)-(8):

$$Precision = TP / (TP + FP) \quad (3)$$

$$Recall = TP / (TP + FN) \quad (4)$$

$$AP = \int_0^1 Precision(Recall) dR \quad (5)$$

$$mAP = \sum_{i=1}^N AP_i / N \quad (6)$$

$$GFlops = O \left(\sum_{i=1}^n K_i^2 * C_{i-1}^2 * C_i + \sum_{i=1}^n m^2 * C_i \right) \quad (7)$$

$$Params = O \left(\sum_{i=1}^n M_i^2 * K_i^2 * C_{i-1} * C_i \right) \quad (8)$$

TP represents genuine positives, FP represents false positives, and FN represents false negatives. $Precision$ and $Recall$ refer to the Precision-Recall curve. N represents the number of defects. O represents the constant order, K represents the kernel size, C represents the number of channels, M represents the input image size, and i represents the number of iterations.

D. Ablation Experiments

1) *CottonWeedDet3 ablation experiments*: To evaluate the performance of the LMS-YOLO11n model in weed detection, ablation experiments were conducted on the CottonWeedDet3 dataset, testing the C3K2_FastGLU, AHFPN, and LGCD modules separately. The results of the ablation experiments are shown in Table I. Compared to the baseline model YOLO11n, LMS-YOLO11n improves mAP50 by 2.5%, while reducing computational load and parameter count by 26% and 37%, respectively.

First, the optimization of C3K2 is discussed. By integrating the FastGLU designed in this study with C3K2, mAP50 increased by 0.9, while both computational load and parameter count were reduced by 6%. This suggests that the C3K2_FastGLU module enhances the model's ability to extract both local and global features of weeds.

Next, the improvement in the NECK section is discussed. After applying the designed AHFPN feature fusion module, mAP50 increased by 0.4%, while computational load decreased by 11% and parameter count by 26%. This demonstrates the effectiveness of the AHFPN module in enhancing weed edge features.

In the detection head, after applying the group convolution-based LGCD module, mAP50 increased by 2.3%, while computational load decreased by 17% and parameter count by 11%. This demonstrates that the LGCD module improves the detailed capture of multi-scale feature information, achieving an optimized balance between feature extraction diversity and model computational efficiency.

Finally, the effect of the cumulative modules was demonstrated. First, combining C3K2_FastGLU with AHFPN resulted in a 0.1% increase in mAP50, with computational load and parameter count decreasing by 0.14% and 32%, respectively. Adding the LGCD module further improved accuracy by 2.5%, while reducing computational load by 26% and parameter count by 37% compared to the baseline model.

2) *CottonWeed2 ablation experiments*: To validate the model's robustness, ablation experiments were conducted on the CottonWeed2 dataset. The results of the ablation experiments are shown in Table II.

Incorporating C3K2_FastGLU optimized the model's feature extraction capabilities. Compared to the baseline model, mAP50 increased to 75%, while computation (GFLOPs reduced from 6.3 to 5.9) and parameter size (reduced from 2.58 MB to 2.41 MB) decreased. This demonstrates that C3K2_FastGLU enhanced the model's ability to perceive fine-grained weed features and overall contextual information.

The inclusion of AHFPN significantly enhanced feature fusion capabilities, particularly for detecting multi-scale targets. Experimental results showed a mAP50 increase to 75.2%, a 0.7% improvement over the baseline model, with computation and parameter size reduced to 5.6 GFLOPs and 1.89 MB, respectively. This indicates that AHFPN optimized the model structure for selecting and fusing multi-scale features while maintaining computational efficiency.

LGCD focused on improving the fine-grained modeling of multi-scale feature information. Although incorporating this module slightly reduced mAP50 by 0.6%, it decreased computation and parameter size to 5.4 GFLOPs and 2.28 MB, respectively. This highlights its efficiency in reducing redundant computations and enhancing contextual feature fusion, making it well-suited for lightweight and edge device applications.

When C3K2_FastGLU, AHFPN, and LGCD modules were progressively combined, the model exhibited significant synergistic performance improvements. With all three modules combined, the model achieved 79.2% precision, a recall rate of

TABLE I. COTTONWEEDDET3 ABLATION EXPERIMENT TABLE

YOLO11n	C3K2_FastGLU	AHFPN	LGCD	P	R	mAP50	mAP50-95	GFLOPs	Params(MB)
✓				85.8	62.4	73.6	58.4	6.3	2.58
✓	✓			74.7	68	74.5	58.1	5.9	2.41
✓		✓		77.8	67	74	57.8	5.6	1.9
✓			✓	75.1	70.7	75.9	62	5.4	2.28
✓	✓	✓		78.6	68.5	74.6	57.8	5.2	1.73
✓		✓	✓	74	69.4	73.3	58.3	5	1.78
✓	✓	✓	✓	78.2	69.5	76.1	60	4.6	1.61

72%, a mAP50 of 76.4%, and a mAP50-95 of 51.2%, showing comprehensive improvements over the baseline model. Additionally, computation decreased to 4.6 GFLOPs and parameter size to 1.61 MB, demonstrating excellent lightweight performance and resource optimization.

Therefore, by comparing the ablation data from these two datasets, The LMS-YOLO11n model put out in this work may effectively extract the edges and fine-grained characteristics of weeds and meet the needs of deploying in various embedded weed detection devices with real-time requirements.

E. Comparison Experiments

1) *CottonWeedDet3 Comparison Experiments with the Latest Models:* To comprehensively evaluate the advantages of the proposed LMS-YOLO11n model, several state-of-the-art models, including YOLOv3-tiny, YOLOv5n, YOLOv6n, YOLOv7-tiny, YOLOv8n, YOLOv10n, and YOLO11n, were selected for comparison. The detection performance on the CottonWeedDet3 dataset is compared in Table III, where the bolded text indicates the greatest outcomes. Table III demonstrates that the improved LMS-YOLO11n model outperforms YOLOv3-tiny, YOLOv5n, YOLOv6n, YOLOv7-tiny, YOLOv8n, YOLOv10n, and YOLO11n. The LMS-YOLO11n model achieved a mAP50 of 76.1% and a mAP50-95 of 0.60% on the CottonWeedDet3 dataset, with a computational load of 4.6 GFLOPs and a parameter size of 1.61 MB. While improving accuracy, the model significantly reduced parameter size and computational load. Although the accuracy(P) slightly decreased, mAP50 increased by 2.5%, computational load decreased by 26% and parameter size reduced by 37%.

2) *CottonWeed2 Comparison Experiments with the Latest Models:* To further verify the generalization capability of LMS-YOLO11n, a comparative experiment was conducted on the CottonWeedDet2 dataset. Table IV presents the comparison results of the latest models on the CottonWeed2 dataset. The LMS-YOLO11n model achieved an mAP50 of 76.4%, and an mAP50-95 of 51.2% on the CottonWeed2 dataset, with a computational load of 4.6 GFLOPs and a parameter size of 1.61 MB. Compared to YOLOv3-tiny, YOLOv5n, YOLOv6n, YOLOv7-tiny, YOLOv8n, YOLOv10n, and YOLO11n, LMS-YOLO11n achieved mAP50 improvements of 8%, 3.1%, 3.0%, 1.3%, 3.1%, 8.4%, and 1.9%, respectively. Additionally, the computational load decreased by 67%, 35%, 60%, 65%, 43%, 28%, and 26%, while the parameter size reduced by 83%, 35%, 61%, 86%, 46%, 21%, and 37%, respectively. These results demonstrate that LMS-YOLO11n achieved the best performance in terms of mAP50, computational load, and parameter size.

3) *CottonWeedDet12 Comparison Experiments with the Latest Models:* In this paper, comparison experiments are also conducted on the CottonWeedDet12 dataset to further validate the robustness of the LMS-YOLO11n model. The relevant comparison data are shown in Table V. Table V demonstrates the performance comparison of multiple models on the CottonWeed12 dataset, and LMS-YOLO11n stands out in terms of comprehensive performance. YOLOv3-tiny, although having a mAP50 of 91.7%, has a computational and parametric count of 14.3 GFLOPs and 9.52 MB, respectively, which is the model with the highest consumption of computational resources in the table, restricting its application on resource-constrained devices. YOLOv5n and YOLOv6n are optimized in terms of computation volume of 7.1 GFLOPs and 11.5 GFLOPs and number of parameters of 2.5 MB and 4.15 MB, respectively, but their mAP50 values of 92.5% and 90.7% are slightly lower than that of the LMS-YOLO11n. YOLOv7-tiny's mAP50 of 92.7% is still high, but its 13.3 GFLOPs of computation and 12.3 MB of parameter count are similar to that of YOLOv3-tiny. YOLOv8n further optimizes the parameter count and computation with a mAP50 of 92.3%, with values of 3 MB and 8.1 GFLOPs, respectively, but is still not as light as that of LMS-YOLO11n. YOLOv10n and YOLO11n, as more advanced models, exhibit mAP50s of 93% and 93.6%, with their computational and parametric quantities reduced to 6.4 GFLOPs and 6.3 GFLOPs and 2.04 MB and 2.58 MB, respectively. The LMS-YOLO11n, with the minimum computational quantity of 4.6 GFLOPs, and the LMS-YOLO11n achieve the same mAP50 as the YOLO11n with a minimum number of parameters of 1.62 MB. Taken together, the LMS-YOLO11n achieves an optimal balance between performance, efficiency, and lightness with a mAP50 of 93.6%, several parameters of 1.62 MB, and a computation volume of 4.6 GFLOPs, making it suitable for complex field environments and resource-constrained edge device scenarios.

F. Model Detection Effect and Visualization Analysis

The improved LMS-YOLO11n demonstrates excellent detection performance, providing accurate and comprehensive weed recognition under various environmental conditions. HiResCAM [36] was used to perform visualization analysis on the CottonWeedDet3 and CottonWeed2 datasets. In the images, darker colors indicate higher attention, while lighter colors represent lower attention, as shown in Fig. 9 and Fig. 10.

Fig. 9 and Fig. 10 show that both LMS-YOLO11n and YOLO11n can identify and locate target areas dominated by weed structures. However, compared to YOLO11n, LMS-YOLO11n reduces false detections and more accurately focuses on the actual weed shapes. Specifically, YOLO11n

TABLE II. COTTONWEED2 ABLATION EXPERIMENT TABLE

YOLO11n	C3K2_FastGLU	AHFPN	LGCD	P	R	mAP50	mAP50-95	GFLOPs	Params(MB)
✓				80.9	71.7	74.5	51.8	6.3	2.58
✓	✓			81.4	67.4	75	50.5	5.9	2.41
✓		✓		86.3	66.7	75.2	51.1	5.6	1.89
✓			✓	76.6	67	73.9	51.1	5.4	2.28
✓	✓	✓		84.8	64.3	75	50.1	5.2	1.73
✓		✓	✓	91.4	67	75.1	52.7	5	1.78
✓	✓	✓	✓	79.2	72	76.4	51.2	4.6	1.61

TABLE III. COMPARISON OF EXPERIMENTAL RESULTS OF DIFFERENT MODELS ON COTTONWEEDDET3 DATASET

	P	R	mAP50	mAP50-95	GFLOPs	Params (MB)
YOLOv3-tiny	78.7	67.4	71.4	52.3	14.3	9.52
YOLOv5n	75.8	64.6	70.6	54.9	7.1	2.5
YOLOv6n	76.6	70.7	74.2	58.6	11.5	4.15
YOLOv7-tiny	85	64.5	73.8	58	13.3	12.3
YOLOv8n	84.2	65.9	74.3	58.4	8.1	3
YOLOv10n	74.7	65.2	70.5	56.1	6.4	2.04
YOLO11n	85.8	62.4	73.6	58.4	6.3	2.58
LMS-YOLO11n	78.2	69.5	76.1	60.6	4.6	1.61

TABLE IV. COMPARISON OF EXPERIMENTAL RESULTS OF DIFFERENT MODELS ON COTTONWEED2 DATASET

	P	R	mAP50	mAP50-95	GFLOPs	Params (MB)
YOLOv3-tiny	64.3	71.3	68.1	43.0	14.3	9.52
YOLOv5n	83.9	66.6	72.7	51.1	7.1	2.5
YOLOv6n	84.5	65.7	72.6	51.2	11.5	4.15
YOLOv7-tiny	77.6	72.1	74.9	45.2	13.3	12.3
YOLOv8n	78.9	70.0	73.3	53.1	8.1	3.0
YOLOv10n	83.9	59.5	68.0	43.6	6.4	2.04
YOLO11n	80.9	71.7	74.5	51.8	6.3	2.58
LMS-YOLO11n	79.2	72.0	76.4	51.2	4.6	1.61

TABLE V. COMPARISON OF EXPERIMENTAL RESULTS OF DIFFERENT MODELS ON COTTONWEEDDET12 DATASET

	P	R	mAP50	mAP50-95	GFLOPs	Params(MB)
YOLOv3-tiny	88.8	86.4	91.7	80.3	14.3	9.52
YOLOv5n	90.6	86.4	92.5	85.3	7.1	2.5
YOLOv6n	90.5	84.8	90.7	84.1	11.5	4.15
YOLOv7-tiny	92.3	86	92.7	82.1	13.3	12.3
YOLOv8n	92.2	85.8	92.3	85.6	8.1	3
YOLOv10n	91.3	87.6	93	88.2	6.4	2.04
YOLO11n	92.4	86	93.6	87.2	6.3	2.58
LMS-YOLO11n	89.6	88.9	93.6	86.1	4.6	1.62

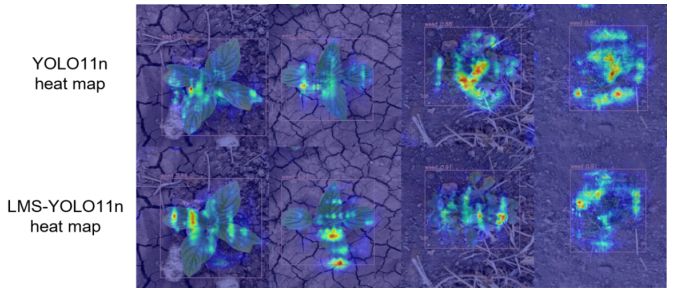


Fig. 10. Contrasting thermal diagrams before and after CottonWeed2 model improvement.

may be affected by morphological similarities and lighting variations in complex field environments, leading to scattered feature capture and reduced target localization accuracy. In contrast, LMS-YOLO11n, with its lightweight design and optimized feature extraction modules, effectively suppresses environmental noise and significantly enhances target feature extraction accuracy. It shows a stronger ability to differentiate when weeds and crops have similar morphologies. This demonstrates the superiority and reliability of LMS-YOLO11n for efficient weed detection in complex field scenarios.

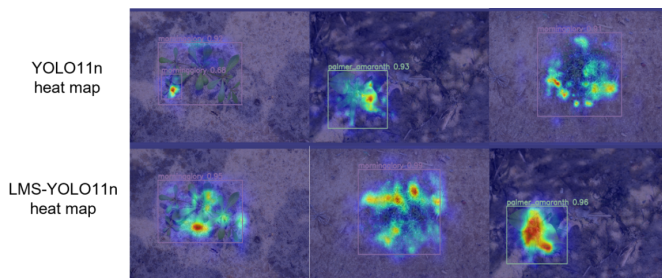


Fig. 9. Contrasting thermal diagrams before and after CottonWeedDet3 model improvement.

V. CONCLUSION AND FUTURE WORK

A. Conclusion

This study introduces LMS-YOLO11n, a novel lightweight weed detection network designed for precision agriculture on edge devices and in complex field environments. The network leverages innovative structural designs to enhance detection accuracy and computational efficiency, making it especially suitable for resource-constrained edge devices.

In the feature extraction phase, LMS-YOLO11n replaces the traditional C3K2 module with the C3K2_FastGLU module, integrating partial convolution and CGLU mechanisms to extract fine-grained weed features more effectively. Compared to traditional convolution methods, FastGLU uses channel-level weighting to enhance sensitivity to fine details, enabling more precise differentiation between weeds and crops in complex field environments. For feature fusion, the study introduces the Adaptive Hierarchical Feature Pyramid Network (AHFPN), which optimizes feature selection and fusion to improve multi-scale weed detection capabilities. AHFPN effectively integrates multi-scale feature maps, enhancing weed detection and preventing the loss of small-scale target information, thereby improving detection accuracy. To boost model efficiency,

LMS-YOLO11n replaces the traditional detection head with the lightweight LGCD module. LGCD, designed with group convolutions, reduces parameter and computation requirements while maintaining high detection accuracy, making it ideal for low-power edge devices capable of efficient real-time weed detection.

LMS-YOLO11n demonstrates superior performance across three datasets. On the CottonWeedDet3 dataset, the model achieved a mAP50 of 76.1%. On CottonWeed2, it reached 76.4%, while on CottonWeedDet12, it achieved 93.6%, reducing computation and parameter sizes by 26% and 37%, in contrast to the baseline model, correspondingly. These results demonstrate that LMS-YOLO11n achieves high-precision detection in complex environments and is deployable on edge devices, providing accurate real-time agricultural monitoring solutions.

B. Future Work

Despite the significant results, the proposed model has some limitations:

- 1) Lack of experiments and deployments in real-world agricultural scenarios. Real agricultural environments involve variations in weed types, densities, and growth states, requiring further validation of the model's performance under these conditions.
- 2) Detection accuracy for young weeds needs improvement. The simple morphology and texture of young weeds often confuse with the background or crops, leading to limited detection accuracy.

To address these issues, this paper will explore multimodal fusion techniques in future research to solve the challenges of young weed detection. By fusing different types of data sources, it can provide richer feature information for the model and help it recognize young weeds more accurately. Meanwhile, more field experiments are planned to be conducted in combination with practical agricultural application scenarios to test the performance of the model in different environments, to further improve its adaptability and accuracy, and to provide more effective solutions for precision agriculture.

ACKNOWLEDGMENT

This work was supported by the Programs for Natural Science Foundation of Xinjiang Uygur Autonomous Region, Grant number 2022D01C54.

REFERENCES

- [1] Bah, M. D., Hafiane, A., Canals, R., & Emile, B. (2019, November). Deep features and One-class classification with unsupervised data for weed detection in UAV images. In *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)* (pp. 1-5). IEEE.
- [2] Donayre, D. K. M., Santiago, S. E., Martin, E. C., Lee, J. T., Corales, R. G., Janiya, J. D., & Kumar, V. (2019). Weeds of Vegetables and other Cash Crops in the Philippines. *Philippine Rice Research Institute, Malingaya, Science City of Muñoz, Nueva Ecija*. 141pp.
- [3] Zhang, W., Miao, Z., Li, N., He, C., & Sun, T. (2022). Review of current robotic approaches for precision weed management. *Current robotics reports*, 3(3), 139-151.

- [4] Jin, X., Liu, T., Chen, Y., & Yu, J. (2022). Deep learning-based weed detection in turf: a review. *Agronomy*, 12(12), 3051.
- [5] Wang, A., Peng, T., Cao, H., Xu, Y., Wei, X., & Cui, B. (2022). TIA-YOLOv5: An improved YOLOv5 network for real-time detection of crop and weed in the field. *Frontiers in Plant Science*, 13, 1091655.
- [6] Kumar, D. A., & Prema, P. (2016). A NOVEL WRAPPING CURVELET TRANSFORMATION BASED ANGULAR TEXTURE PATTERN (WCTATP) EXTRACTION METHOD FOR WEED IDENTIFICATION. *ICTACT Journal on Image & Video Processing*, 6(3).
- [7] Sujaritha, M., Annadurai, S., Satheshkumar, J., Sharan, S. K., & Mahesh, L. (2017). Weed detecting robot in sugarcane fields using fuzzy real time classifier. *Computers and electronics in agriculture*, 134, 160-171.
- [8] Chen, X., & Gupta, A. (2017). An implementation of faster rcnn with study for region sampling. *arxiv preprint arxiv:1702.02138*.
- [9] Ren, S., He, K., Girshick, R., & Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE transactions on pattern analysis and machine intelligence*, 39(6), 1137-1149.
- [10] Zhang, X., Cui, J., Liu, H., Han, Y., Ai, H., Dong, C., ... & Chu, Y. (2023). Weed identification in soybean seedling stage based on optimized Faster R-CNN algorithm. *Agriculture*, 13(1), 175.
- [11] Özcan, R., Tütüncü, K., & Karaca, M. (2022). Comparison of Plant Detection Performance of CNN-based Single-stage and Two-stage Models for Precision Agriculture. *Selcuk Journal of Agriculture and Food Sciences*, 36(4), 53-58.
- [12] Li, Z., Li, Y., Yang, Y., Guo, R., Yang, J., Yue, J., & Wang, Y. (2021). A high-precision detection method of hydroponic lettuce seedlings status based on improved Faster RCNN. *Computers and electronics in agriculture*, 182, 106054.
- [13] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). Ssd: Single shot multibox detector. In *Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14* (pp. 21-37). Springer International Publishing.
- [14] Redmon, J. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- [15] Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., ... & Wei, X. (2022). YOLOv6: A single-stage object detection framework for industrial applications. *arxiv preprint arxiv:2209.02976*.
- [16] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 7464-7475).
- [17] Wang, C. Y., Yeh, I. H., & Mark Liao, H. Y. (2025). Yolov9: Learning what you want to learn using programmable gradient information. In *European Conference on Computer Vision* (pp. 1-21). Springer, Cham.
- [18] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). Yolov10: Real-time end-to-end object detection. *arxiv preprint arxiv:2405.14458*.
- [19] Chen, J., Wang, H., Zhang, H., Luo, T., Wei, D., Long, T., & Wang, Z. (2022). Weed detection in sesame fields using a YOLO model with an enhanced attention mechanism and feature fusion. *Computers and Electronics in Agriculture*, 202, 107412.
- [20] Zhao, Y., Sun, C., Xu, X., & Chen, J. (2022). RIC-Net: A plant disease classification model based on the fusion of Inception and residual structure and embedded attention mechanism. *computers and Electronics in Agriculture*, 193, 106644.
- [21] Hong, W., Ma, Z., Ye, B., Yu, G., Tang, T., & Zheng, M. (2023). Detection of green asparagus in complex environments based on the improved YOLOv5 algorithm. *Sensors*, 23(3), 1562.
- [22] Liu, L., Liang, J., Wang, J., Hu, P., Wan, L., & Zheng, Q. (2023). An improved YOLOv5-based approach to soybean phenotype information perception. *Computers and Electrical Engineering*, 106, 108582.
- [23] Zhang, C., Liu, J., Li, H., Chen, H., Xu, Z., & Ou, Z. (2023). Weed Detection Method Based on Lightweight and Contextual Information Fusion. *Applied Sciences*, 13(24), 13074.
- [24] Guo, A., Jia, Z., Wang, J., Zhou, G., Ge, B., & Chen, W. (2024). A lightweight weed detection model with global contextual joint features. *Engineering Applications of Artificial Intelligence*, 136, 108903.

- [25] Fan, X., Sun, T., Chai, X., & Zhou, J. (2024). YOLO-WDNet: A lightweight and accurate model for weeds detection in cotton field. *Computers and Electronics in Agriculture*, 225, 109317.
- [26] Chen, Y., Zhang, C., Chen, B., Huang, Y., Sun, Y., Wang, C., ... & Gao, Y. (2024). Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases. *Computers in Biology and Medicine*, 170, 107917.
- [27] Chen, J., Kao, S. H., He, H., Zhuo, W., Wen, S., Lee, C. H., & Chan, S. H. G. (2023). Run, don't walk: chasing higher FLOPS for faster neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 12021-12031).
- [28] Shazeer, N. (2020). Glu variants improve transformer. *arxiv preprint arxiv:2002.05202*.
- [29] Shi, D. TransNeXt: Robust Foveal Visual Perception for Vision Transformers. *arxiv 2023. arxiv preprint arxiv:2311.17132*.
- [30] Wang, W., **, E., Song, X., Zang, Y., Wang, W., Lu, T., ... & Shen, C. (2019). Efficient and accurate arbitrary-shaped text detection with pixel aggregation network. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 8440-8449).
- [31] Howard, A. G. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arxiv preprint arxiv:1704.04861*.
- [32] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1492-1500).
- [33] Rahman, A., Lu, Y., & Wang, H. (2022). Deep neural networks for weed detections towards precision weeding. In *2022 ASABE Annual International Meeting* (p.1). American Society of Agricultural and Biological Engineers.
- [34] Kumaran, D.T. Cotton-Weed Dataset. (2021). Available online: <https://universe.roboflow.com/deepak-kumaran-t/cotton-weed>.
- [35] Dang, F., Chen, D., Lu, Y., & Li, Z. (2023). YOLOWeeds: A novel benchmark of YOLO object detectors for multi-class weed detection in cotton production systems. *Computers and Electronics in Agriculture*, 205, 107655.
- [36] Draelos, R. L., & Carin, L. (2020). Use HiResCAM instead of Grad-CAM for faithful explanations of convolutional neural networks. *arxiv preprint arxiv:2011.08891*.

DBYOLOv8: Dual-Branch YOLOv8 Network for Small Object Detection on Drone Image

Yawei Tan¹, Bingxin Xu², Jiangsheng Sun³, Cheng Xu⁴, Weiguo Pan⁵, Songyin Dai⁶, Hongzhe Liu⁷
Beijing Key Laboratory of Information Service Engineering, Beijing Union University, China^{1,2,4,5,6,7}
Science and Technology Innovation Research Center, Army Research Academy³

Abstract—Object detection based on drone platforms is a valuable yet challenging research field. Although general object detection networks based on deep learning have achieved breakthroughs in natural scenes, drone images in urban environments often exhibit characteristics such as a high proportion of small objects, dense distribution, and significant scale variations, posing significant challenges for accurate detection. To address these issues, this paper proposes a dual-branch object detection algorithm based on YOLOv8 improvements. Firstly, an auxiliary branch is constructed by extending the YOLOv8 backbone to aggregate high-level semantic information within the network, enhancing the feature extraction capability. Secondly, a Multi-Branch Feature Enhancement (MBFE) module is designed to enrich the feature representation of small objects and enhance the correlation of local features. Third, Spatial-to-Depth Convolution (SPDConv) is utilized to mitigate the loss of small object information during downsampling, preserving more small object feature information. Finally, a dual-branch feature pyramid is designed for feature fusion to accommodate the dual-branch input. Experimental results on the VisDrone benchmark dataset demonstrate that DBYOLOv8 outperforms state-of-the-art object detection methods. Our proposed DBYOLOv8s achieve mAP@0.5 of 49.3% and mAP@0.5:0.95 of 30.4%, which are 2.8% and 1.5% higher than YOLOv9e, respectively.

Keywords—Drone images; dual-branch; small object detection; YOLOv8

I. INTRODUCTION

With the development of hardware and artificial intelligence, drones have been gradually applied to intelligent transportation, agricultural monitoring, fire rescue and other fields. In urban traffic monitoring and urban combat missions, UAVs (unmanned aerial vehicle) play an important role by virtue of their advantages such as fast flight speed, high degree of freedom, broad vision and strong adaptability. However, the streets in the city scene have the characteristics of traffic congestion, dense people, and a wide variety of targets. In addition, due to the high-altitude flight of UAV, objects in UAV images are often too small in size and contain limited feature information, which makes it difficult for the network to extract effective features and easy to be lost in the propagation process across the feature layer [1]. In addition, the size of similar objects varies so much that it is difficult for universal object detection methods to effectively locate and identify these objects [2].

Uav object detection is one of the branches of general target detection. According to the processing flow, the target detection algorithm can be divided into two stages and one stage. The two-stage algorithm is characterized by generating a series

of regions of interest, and then classifying and regressing these regions. Its advantage is that the two-stage detection algorithm is more detailed, resulting in higher detection accuracy. The disadvantage is that the inference speed is slower than that of single-stage algorithms. The two-stage representative algorithms include Faster R-CNN [3] and Mask R-CNN [4]. The single-stage algorithm extracts the feature information of the target by convolutional neural network, generates the candidate frame, and classifies and locates the target. This detection method consumes less computer resources during inference, and UAV target detection is usually improved based on single-stage algorithm. Single-stage representative algorithms include SSD [5] and YOLO series [6]. YOLOv8 is a commonly used single-phase detection framework, which is often used for various object detection tasks [7]. Its advantage is that the framework is mature, the externally adapted function library is more common, and a variety of inference accuracy improvement tools can be used directly. However, the objects in UAV images often have problems such as small size, complex background environment, and dense area overlap, which limits the ability of the frame to detect small objects. Secondly, the framework is still weak in detecting similar objects at different scales. Therefore, it is necessary to improve the YOLOv8 algorithm to make it suitable for UAV small object detection.

This paper presents a dual-branch small object detection algorithm based on YOLOv8. Firstly, a composite strategy is used to construct auxiliary branches, and the multi-layer semantic information is comprehensively utilized to improve the feature extraction capability of the framework. Second, a multi-branch feature enhancement module is designed, which uses convolution check of different sizes for parallel processing of small object feature information to improve the representation ability of object feature information. In addition, SPDConv can effectively reduce the loss of feature information in the transmission process, which is very effective for small object detection. When it is embedded in the shallow detection branch of the network, the false detection problem can be well improved. Finally, a dual-branch feature pyramid is constructed to deal with the multi-scale change of the target. Experimental results show that the proposed algorithm greatly improves the performance of object detection and can better cope with the requirements of different tasks on model size. Our main improvements and advantages are as follows:

- C2f module is used to construct auxiliary branch, which aggregate multi-high-level semantic information and enhance the feature extraction ability of small objects. SPDConv [13] is used to alleviate the loss of

feature information in the downsampling process.

- Multi-branch Feature Enhancement Module (MBFE) is designed to extract small object feature information by using parallel branches of different convolution kernel sizes, which can realize the diversification of small objects feature information expression.
- Dual-branch feature pyramid network (DBFPN) is established for cross-layer connection with YOLOv8 backbone to compensate for information loss caused by feature information transformation.

The structure of this paper is as follows: In Section II, we will briefly introduce our related work to improve the thinking. In Section III, we take a detailed look at the dual-branch YOLOv8 framework. In Section IV, we conduct experiments on a classical drone dataset and provide a detailed analysis of the results. In Section V, we analyze the existing shortcomings and the continued exploration of future work.

II. RELATED WORK

In object detection, the size of the object can be divided into absolute scale and relative scale according to the definition. In the definition of relative scale [8], usually the relative area of all object instances in the same category, that is, the median ratio of the boundary box area to the image area is between 0.08% and 0.58%. However, the way to define small objects based on absolute scale is more widely used, and the MS COCO dataset [9] defines small targets as those with a resolution less than 32 pixels by 32 pixels. The existing methods to solve the small object detection of UAVs can be classified into three categories: (1) By enhancing the feature information of small objects, the network can locate the objects more clearly. (2) Improve the detection accuracy of small objects by improving the ability of network feature extraction. (3) Adopt multi-scale detection strategies to deal with small objects of different sizes.

To improve the ability of network feature extraction, Liang et al. [10]. proposed CBNetV2 network, which uses shallow network to aggregate different high-level semantic information, aiming to enhance the comprehensive application of feature texture features by the model. In addition, they demonstrated experimentally that small objects can be detected more efficiently when shallow features are aggregated only with feature layers higher than this one. This work pioneered the concept of composite backbone networks. Wang et al [11]. proposed Yolov9, the representative of YOLO series, and designed a Programmable Gradient Information (PGI), which uses the characteristics of reversible architecture to retain more input information, thereby reducing the loss of small target feature information. In this framework, the composite backbone network is also constructed. Yan et al. [12]. propose an HCB network that includes a detail extraction backbone (DEB) designed with a smaller acceptance field to better capture details of small objects. This design enhances feature representation without compromising spatial information. However, the above method only uses a single strategy, and because more parameters are often introduced in order to obtain more gradient information, the computational complexity increases and the practical application is limited.

For the enhancement of small object feature information, the detection accuracy of the network can be improved by improving the feature representation of the network for small object and reducing the problem of the loss of small object feature information. Zhang et al. [13]. developed a feature enhancement module specifically for aerial image detection, using the improved FFM module to further capture the context information of small objects, thereby improving the detection accuracy. Raja et al. [14]. designed a step-free convolution to solve the problem of information loss caused by different interpolation calculations for small objects through this lossless downsampling method, thus improving the network's ability to perceive small objects. However, the improved method has been proved by experiments that the network pays too much attention to texture information when applied equally in each feature layer, which leads to the decrease of detection accuracy.

In order to cope with targets of different sizes, Tsung et al. [15]. proposed the concept of feature pyramid. By connecting shallow texture information and high semantic information from top to bottom, the object feature information of each detection layer is enriched. On this basis, Liu et al. [16]. added a new bottom-up path that preserves more detailed information, which is also effective for multi-scale small objects. Tan et al. [17]. used a weighted feature fusion mechanism to give each input feature path a learnable weight, allowing the network to automatically adjust the importance of each path, thus making more efficient use of feature information. By removing invalid nodes and reusing features, the computational cost is reduced, making it suitable for resource-constrained devices. However, the processing method of feature pyramid is only suitable for a single backbone, and for multi-branch networks, the conventional feature pyramid will dilute the target feature information in the concatenation operation.

Although the existing improvement methods have continuously improved the UAV target detection performance, the existing network architecture is still difficult to achieve high precision and multi-task adaptation, especially for the dense area detection problem in the urban scene, and it is urgent to further improve the detection accuracy. Therefore, a variety of improvement methods should be comprehensively used to enhance the detection ability of the detection network for UAV images.

III. DBYOLOv8 ALGORITHM

A. Overview of YOLOv8

YOLOv8 is an object detection framework based on single-stage deep learning. Compared with the existing version of YOLO series, YOLOv8 can adjust the size of the model by adjusting the scale factor to adapt to different task requirements. Compared to YOLOv10 [18] and YOLOv11, YOLOv8 framework is more mature and has many existing tools to assist reasoning. YOLOv8 network structure is mainly divided into three parts: (1) Backbone network for extracting object feature information. (2) Processing multi-scale features of the feature pyramid pool layer. (3) The detection head of the classification object type information. Fig. 1 shows the schematic diagram of the YOLOv8 algorithm framework. The backbone network extracts the feature information of the object by using the convolution layer of step by step downsampling. The excellent

feature extraction ability is the basis of realizing the high-precision object detection algorithm. Path Aggregation-FPN (PAFPN) structure is introduced into the neck structure, and the feature mapping of different scales is combined to enhance the ability of the algorithm to recognize objects of different sizes. The head layer is the main decoupling head structure, which separates the classification and detection head, and becomes the Anchor-Free detection scheme. The YOLOv8 algorithm is widely used in many fields (for example, agricultural inspection, UAV object detection and autonomous driving). However,

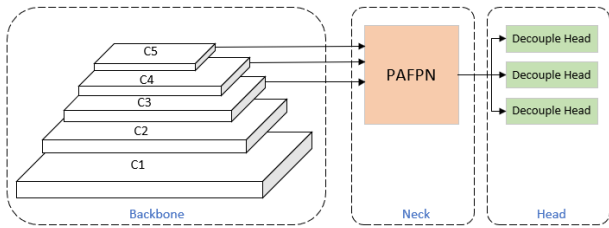


Fig. 1. Simplified diagram of YOLOv8 network structure.

the performance of the baseline YOLOv8 is not optimal, and there is no targeted design for small objects. In addition, YOLOv8 does not fully combine shallow features and deep features, so that the feature information of small targets is seriously lost in the process of feature transmission. Therefore, the general object detection algorithm framework YOLOv8 is not suitable for small object detection tasks of UAVs. In order to meet the higher task requirements of UAVs for object detection algorithms, it is necessary to improve the existing algorithms by task driving.

B. Overall Structure of the Optimized DBYOLOv8 Network

In the improved dual-branch YOLOv8 object detection algorithm, taking YOLOv8 as the benchmark model, three aspects of backbone network structure, feature enhancement module and multi-scale feature fusion are optimized and improved. Fig. 2 shows the optimized two-branch YOLOv8 backbone network structure. In the feature extraction stage, this paper designed an auxiliary branch to aggregate the target

features of different feature layers. Inspired by the auxiliary branch constructed by CNet and YOLOv9, the auxiliary branch based on YOLOv8 structure was constructed by using C2F module, which can adjust the size of the model. Based on this method, the constructed DBYOLOv8 model can adapt to the model size requirements of different tasks, and the DBYOLOv8 network model is smaller at the same level of detection accuracy. Feature layer scales the feature map to a fixed size by interpolation, and the small object feature information will be lost in the process of transferring between feature layers. By introducing SPDConv in the shallow layer of the trunk and branches, the problem of small object information loss is alleviated by splitting and reassembling. In addition, EMA [19] module with parallel structure and CBLinear structure are combined to extract small object feature information through different receptive fields of parallel branches. This combination forms MBFE module, does not introduce additional parameters, diversifies the small target feature information, and enhances the generalization ability of the model. The double branch feature pyramid is improved based on BiFPN. By fusing the two-branch feature input with feature weighting, the structure can fuse multi-scale features in the neck network and enhance the model's ability to recognize targets of various sizes and shapes.

C. Auxiliary Branch

In the process of feature extraction, the feature information of small objects is lost or offset to a certain extent with the reduction of the feature map size and the calculation method of interpolation. It constitutes a unique phenomenon, shallow feature is close to the input layer and contains richer texture information, while the deep feature has a larger receptive field and contains more semantic information after multiple convolution. The integrated use of shallow and deep feature can effectively improve the network detection performance [20].

Inspired by CNetV2, the PGI proposed by YOLOv9 framework builds its auxiliary branch by combining multilevel high-level feature information with shallow feature, hoping to enhance the feature representation capability of the backbone. However, the RepNSCPELAN module is designed to capture

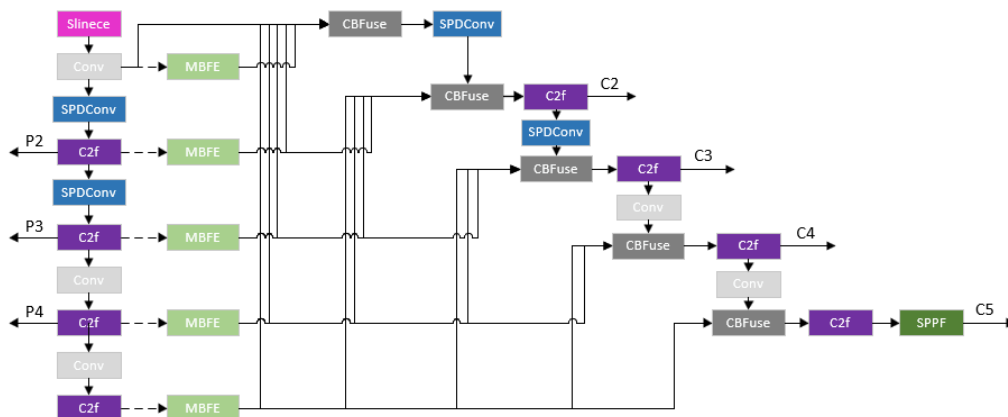


Fig. 2. Our improved DBYOLOv8 feature extraction structure.

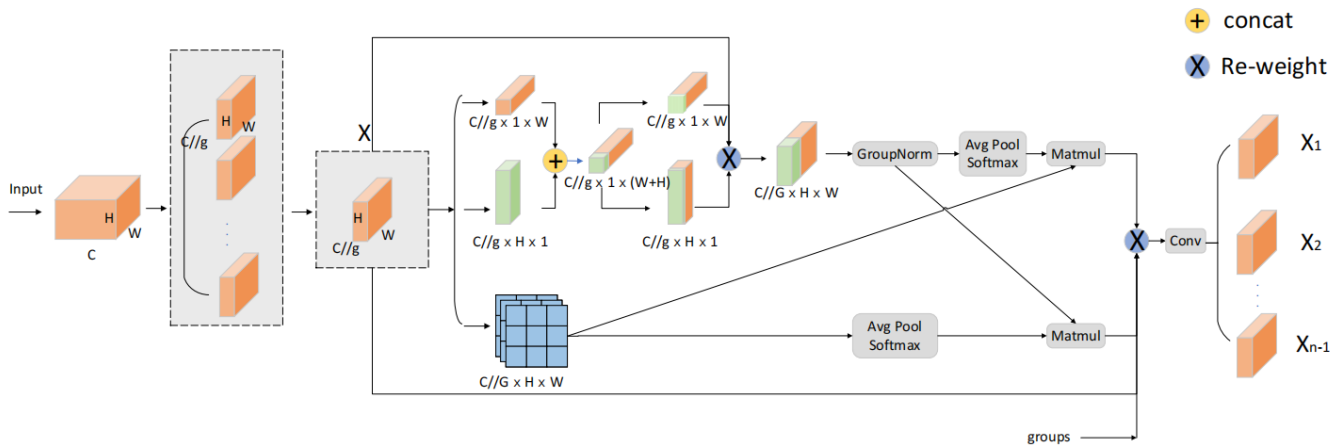


Fig. 3. Multi branch feature enhancement module.

a richer flow of gradient information, which in turn greatly increases the number of model parameters. And it can not adjust the size of its model through scaling factors, making it difficult to adapt to the needs of multiple tasks. In order to improve the feature extraction ability of the framework for small objects without increasing the model size, we built a similar auxiliary branch based on YOLOv8 framework. We use the C2f module to obtain the feature gradient information and the scaling factor to adjust the size of the model to adapt to the task requirements of different platforms.

D. Multi-branch Feature Enhancement Module

The feature information of small objects in UAV images in urban scenes is less, but the background information is complex. Background will seriously affect the extraction of object feature information, which leads to confusion in the transfer process of feature information, thus affecting the performance of the detector. To alleviate this problem, YOLOv9 uses the CBLinear module to process the feature information extracted from the backbone. However, CBLinear module uses the full connection layer to process feature information, which is not friendly for small objects. And it is easy to dilute the feature information of small objects in the process of feature flow, resulting in a certain degree of feature extraction ability loss [21]. Based on the above problems, we design the MBFE module to diversify the small object feature representation.

Channel or spatial attention have been shown to be remarkably effective in producing more recognizable feature representations in various computer vision tasks. In order to diversify the feature information of small objects, we introduced EMA attention mechanism. This mechanism employs multi-branch feature extraction operations, where feature maps are processed in parallel through 1×1 and 3×3 branches. By leveraging different receptive fields, it captures the feature information of small objects and further aggregates the output features of these parallel branches through cross-dimensional interactions to capture pixel-level pairwise relationships. Subsequently, operations such as channel number adjustment and segmentation are performed to obtain a list of multiple output feature maps, which are suitable for subsequent feature aggregation

requirements. The designed MBFE module is illustrated in Fig. 3.

E. SPDConv Module

SPDConv proposed by Sunkara et al. is a lossless downsampling method specifically designed for low-resolution images and small objects. Traditional downsampling techniques, such as strided convolutions and pooling layers, often result in the loss of fine-grained information when dealing with low-resolution images or small objects. To address this issue, SPDConv introduces a lossless downsampling method to segment the input image and splicing the input image in the channel dimension so as to retain the input image feature information. After this operation, the convolution layer with stride=0 is used for feature extraction, and the fine-grained details of the image are retained because the size of the feature map is not changed. This approach significantly mitigates the problem of small object loss during the feature extraction phase. We only applied SPDConv during the initial downsampling stages, as deeper features, after multiple convolution operations, have already become highly ambiguous in terms of small object location information. Using SPDConv for downsampling at these deeper stages could negatively impact the detection network [22]. The specific addition location is illustrated in Fig. 1.

F. Dual-branch Feature Pyramid Module

BiFPN removes connection layers that are not intended for fusion and employs a bidirectional weighted strategy to update gradients. Our designed DBFPN is based on BiFPN and is adapted for a dual-branch backbone design. Although the auxiliary branch is constructed based on the yolov8 backbone, after multiple convolutional operations, there is a slight deviation in the mapping of small object feature information and the original image object information. Therefore, incorporating the main branch feature layer information during the construction of the feature pyramid is beneficial for comprehensively utilizing the feature extraction capabilities of both branches. For small object feature information, we have added a P2 detection layer and, considering the need to control the number of

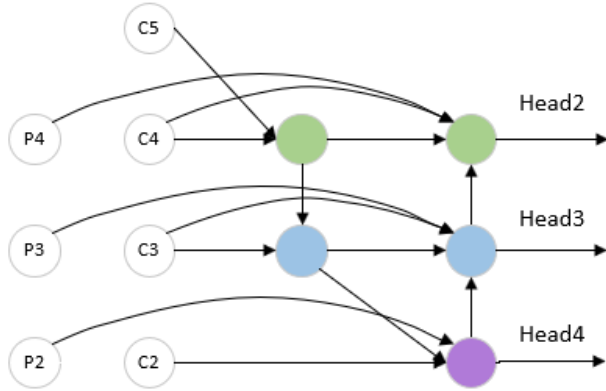


Fig. 4. Dual branch feature pyramid module.

parameters, removed the P5 detection layer [23]. Our designed DBFPN is illustrated in Fig. 4.

IV. EXPERIMENTAL VALIDATION AND ANALYSIS

A. Dataset Analysis

In this study, the VisDrone2019 [24] drone object detection dataset, which can represent urban scenes, was used to test the detection performance of DBYOLOv8 on various types of small objects in drone images. The dataset comprises 6,471 images for training and 548 images for validation, with annotations for 10 types of objects, including pedestrians, cars, motorcycles, and others. An analysis of the training set revealed that small objects constitute approximately 60% of the dataset based on their relative scale. Specifically, the dataset categorizes objects as follows: extremely small (es) objects with a length*width in the range [0, 144], relatively small (rs) objects with dimensions in the range [144, 400] pixels, and generally small (gs) objects with sizes in the range [400, 1024] pixels [25]. Given that the dataset was collected using a drone platform, it is particularly well-suited for assessing the performance of the DBYOLOv8 model in detecting small objects from a drone's perspective. Examples of target statistics are shown in Fig. 5.

In order to verify the generalization of our proposed method, a comparative test was also performed on our AI-TOD dataset [26]. AI-TOD dataset is a representative dataset for the detection of tiny objects in aerial images. The dataset contains 28,036 images labeled with a total of 700,621 instances across eight categories (aircraft, Bridges, tanks, ships, swimming pools, vehicles, people, windmills). Compared with other aerial image datasets of the same type, the average size of the object in this dataset is 12.8 pixels, which is much smaller than the object instances in other datasets. Therefore, it is a good way to evaluate the model's perception of small scale objects.

B. Experimental Condition and Assessment Metrics

The experiment was trained and verified on the research group server. The hardware system consists of the following parts: Intel i9 series 13th generation processor I9-13900KF, RTX4090 (24G) graphics card, 64G memory. The software

system uses Ubuntu22.04 operating system, and uses Pytorch framework to realize all the algorithms running and improving. For comparison with other algorithms, the input image is set to 640 × 640 pixels and the epoch is set to 200 rounds. The other Settings are the default Settings for the YOLOv8 project provided by the ultralytics team.

To evaluate the algorithm's detection performance on objects in drone images, precision (P), recall (R), mAP@50 and mAP@50:95 were used as evaluation indexes. True positive (TP), false negative (FN), false positive (FP) and true negative (TN) were used as anchor frame positioning quality evaluation. Precision Indicates the percentage of the predicted positive samples that are actually positive. The calculation formula is:

$$precision = \frac{TP}{TP + FP} \quad (1)$$

Recall indicates the proportion of the actual number of positive samples in the total positive samples in which the prediction result is positive. The calculation formula is:

$$recall = \frac{TP}{TP + FN} \quad (2)$$

AP is the Average Precision, which is simply to average the precision value on the PR curve. For the pr curve, we use the integral to calculate. The calculation formula is:

$$AP = \int_0^1 p(r)dr \quad (3)$$

mAP is an evaluation index associated with Intersection over Union (IoU), which averages the detection accuracy of all categories. When IoU is set to 0.5, it is usually used as an evaluation index of the detection accuracy of the universal target. mAP@50:95 indicates the mAP with the IoU threshold ranging from 0.5 to 0.95 and the step size of 0.05. Then the average value is obtained. It can also reflect the performance difference of detection algorithms for objects at different scales.

C. Ablation Study

To validate the detection capability of our proposed model for small objects, we constructed auxiliary branches on the basis of the YOLOv8s model, incorporating the SPDconv module, MBFE module, and DBFPN module to build the DBYOLOv8s model. We set the image size to 1280x1280, which is close to the original image size and better reflects the object detection performance of our model on this dataset. The ablation experiment results are shown in Table I. Under the same parameter settings, our method significantly improves the object detection capability for drone images.

1) *Effect of auxiliary branches*: Small objects occupy a high proportion in drone images and contain limited feature information. To enhance the backbone's ability to extract features from small targets, we constructed an auxiliary branch using the SPDConv module, CBFuse module, and C2f module. By aggregating high-level semantic information from layers not lower than the current one, we enriched the feature information of small targets. Compared to the baseline model, the results for mAP@0.5 and mAP@0.5:0.95 improved by 4.7% and 3.5%, respectively.

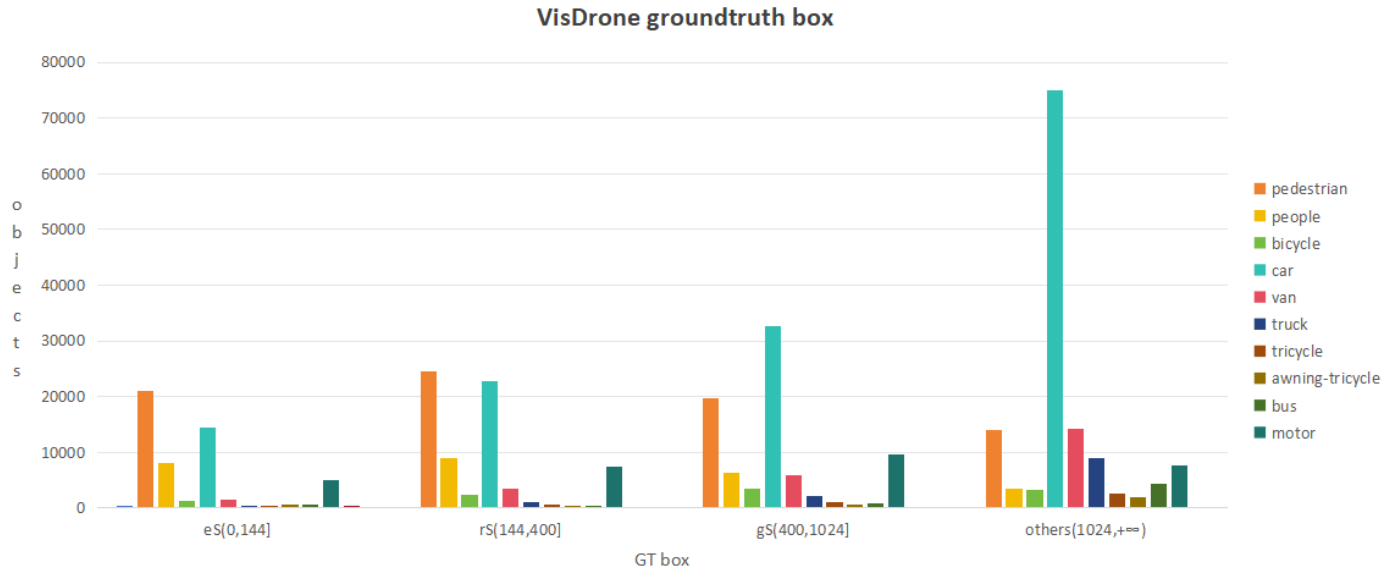


Fig. 5. VisDrone train dataset.

TABLE I. ABLATION STUDY

Baseline	Auxiliary Branch	DBFPN	SPDConv	MBFE	mAP@50(%)	mAP@50:95(%)	Params
✓					56.3	35.4	10.6
✓	✓				61.0	38.9	20.2
✓	✓	✓			61.0	39.0	23.3
✓	✓	✓	✓		61.7	39.5	24.0
✓	✓	✓	✓	✓	62.1	39.9	24.0

2) *Effect of DBFPN module:* We understand that the feature information after multiple convolutions differs from the original feature information. Moreover, the flow of feature information across layers can result in some information loss. Therefore, our proposed DBFPN integrates dual-branch feature information through skip connections, which improves mAP@0.5:0.95 by 0.1%.

3) *Effect of SPDConv module:* Small objects may experience varying degrees of feature information loss during the downsampling process due to differences in interpolation methods. As mentioned above, SPDConv can alleviate the issue of feature information loss caused by downsampling in low-resolution images. However, if SPDConv is uniformly applied to replace every downsampling step, it can negatively impact detection performance. This is because, in deeper layers of the network, small objects have less feature information, and the feature information of larger objects is diluted by SPDConv, leading to missed detections. We conducted three sets of experiments: one with SPDConv added to all layers, one with SPDConv added only in the shallow layers, and one with SPDConv added only in the deep layers. The detector achieved the best performance when SPDConv was added only in the shallow layers, as verified by the experiments shown in Table II.

4) *Effect of MBFE module:* The feature information of small objects is processed through parallel branches, allowing

the extraction of target information using convolution kernels of different sizes. This method of enhancing small object features effectively diversifies the representation of small object feature information, thereby enhancing the network's feature extraction capabilities. Compared to the CBLinear module that solely employs fully connected layers, this approach improves mAP@0.5 and mAP@0.5:0.95 by 0.4% without introducing additional parameters.

TABLE II. COMPARISON RESULT OF DIFFERENT SPDConv ADDITION POSITIONS ON THEVISDRONE2019 VALIDATION DATASETS. THE BEST RESULT IS HIGHLIGHTED IN BOLD

Method	mAP@50(%)	mAP@50:95(%)
P1 - > P3	61.1	39.2
P3 - > P5	60.1	38.4
P1 - > P5	60.9	39.0

D. Comparison with State-of-the-Arts

Due to the varying size constraints for tasks across different platforms, we designed two DBYOLOv8 models of different sizes based on the YOLOv8s and L models. The scaling factors for the DBYOLOv8s model are [0.35, 0.50], while those for the DBYOLOv8L model are [1.00, 1.00]. We compared DBYOLOv8 with other widely used object detection algorithms (primarily the s and l models of various object detection

TABLE III. COMPARISON RESULTS OF DIFFERENT OBJECT DETECTORS ON THEVISDRONE2019 VALIDATION DATASETS. THE BEST RESULT IS HIGHLIGHTED IN BOLD

Method	Inputsize	mAP@50(%)	mAP@50:95(%)	Params(M)	FLOPs(G)
RetinaNet[27]	1333*800	39.3	21.8	-	524.95
Faster-RCNN	1333*800	43.6	24.8	-	322.25
YOLOv5-s[28]	640*640	32.2	17.5	7.2	16.5
TPHYOLOv5-s[29]	640*640	37.4	21.7	-	-
YOLOv8-s	640*640	37.3	22.1	11.1	28.5
Drone-YOLO[30]	640*640	44.3	-	10.9	-
yolov10s	640*640	41.2	24.8	8.0	24.5
YOLOv11s	640*640	41.6	25.2	9.4	21.3
HIC-YOLO[31]	640*640	44.3	26.0	-	-
YOLOv5-l	640*640	42.9	26.3	46.5	109.1
YOLOv8l	640*640	43.7	26.7	43.6	165.4
TPHYOLOv5-l	640*640	41.8	24.0	-	-
YOLOv8-x	640*640	44.3	27.2	68.2	258.5
YOLOv9e	640*640	46.5	28.9	57.3	189.0
DBYOLOv8-s	640*640	49.3	30.4	24.0	119.8
DBYOLOv8-l	640*640	54.4	34.3	175.7	877.0

frameworks). The results, as shown in Table III, indicate that DBYOLOv8 achieved the best and second-best results in terms of mAP. Compared to YOLOv8l and YOLOv9e, DBYOLOv8s achieved higher mAP@50:95 values by 3.7% and 1.5%, respectively, but with significantly fewer parameters. DBYOLOv8 demonstrated superior performance in small object detection compared to other methods, and the experimental results confirm the competitive advantage and effectiveness of this approach.

To verify the effectiveness of the proposed method in identifying complex backgrounds, significant scale differences, and densely packed small objects, we provide visual examples of DBYOLOv8s and YOLOv8l in Fig. 6. In the first line of the image, the vehicles on the far side of the street are extremely small in size. Carefully examining the red box, it is clear that YOLOv8l cannot fully recognize these extremely small objects, while our proposed method also has certain detection accuracy for extremely small objects. The second line of images taken by the drone from a low altitude Angle shows that the green box shows that YOLOv8l missed the target, and the blue box shows that the method incorrectly identified the person as a motorcycle. In contrast, our proposed approach is not affected by these factors. The observations show that our proposed method shows significant advantages over other methods in processing images of this nature.

In order to verify the detection ability of the method for small targets and its generalization on other datasets, we conducted training and inference experiments on AI-TOD datasets using the same parameters. Compared with the mainstream YOLO improved algorithm and DERT improved algorithm, our proposed DBYOLOv8s has undoubtedly obtained the best detection results. Experimental results are shown in Table IV. Compared with YOLOv8l, the proposed algorithm at mAP@50:95 improves by 1.2%. Compared with other algorithms, our method also has obvious advantages.

Compared with the VisDrone dataset, the AI-TOD dataset has more small object instances, which indicates that our method will improve the detection accuracy if there is more sufficient data support. These results across different datasets underscore the robustness and effectiveness of the proposed method.

TABLE IV. COMPARISON RESULTS OF DIFFERENT OBJECT DETECTORS ON AI-TOD VALIDATION DATASETS. THE BEST RESULT IS HIGHLIGHTED IN BOLD

Method	mAP@50(%)	mAP@50:95(%)
YOLOv6[32]	42.2	18.4
YOLOv7[33]	49.5	19.7
YOLOv8l	48.4	22.0
RT-DETR	48.9	22.7
YOLOv10b	46.7	21.6
YOLOv9c	45.8	20.3
DBYOLOv8-s	55.2	23.2

V. CONCLUSION

In order to meet the requirements of existing algorithm frameworks for UAV small object detection, we propose a dual-branch YOLOv8 small object detection algorithm. Firstly, we construct auxiliary branches with compound strategy, combine shallow feature information and higher level semantic information, and increase the feature extraction capability of detection network for small objects. Second, in order to enhance the feature representation of small objects, a multi-branch feature enhancement module is designed to extract the feature information of small objects in parallel through features of different convolution kernel sizes. This module can effectively diversify the representation of small object feature information and counter the problem of the loss of feature information in the process of transmission. Third,

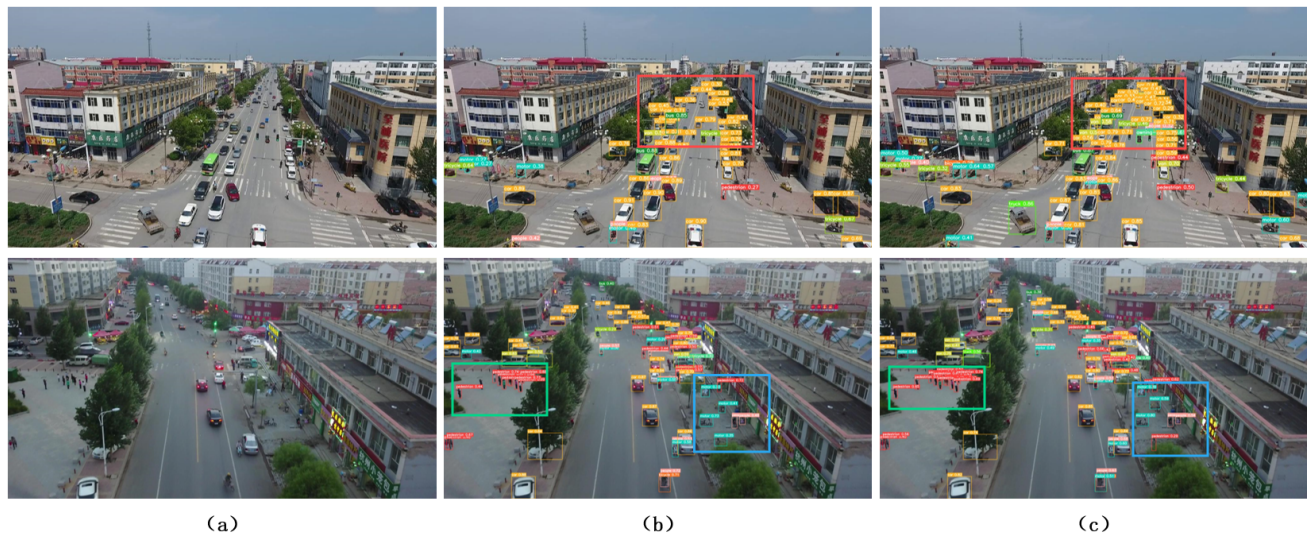


Fig. 6. Comparison of testing results. (a) Original image. (b) YOLOv8l detection results. (c) Our DBYOLOv8s detection results.

we replace the original subsampling with SPDConv in the shallow layer of the network, and maximize the retention of object feature information through recombination and splicing operations, reducing the missing problem caused by the loss of small and medium-sized object feature information during the subsampling process. Secondly, in order to deal with the contact deviation between the feature information and the original image information caused by multiple convolution, we construct a dual-branch feature pyramid to comprehensively use the double-branch feature information to solve the problem of object scale change in the UAV image. Finally, in addition to using the VisDrone dataset, we also used the AI-TOD dataset to evaluate our proposed approach. The effectiveness of our proposed method is verified by experiments. Compared with the basic YOLOv8s, the DBYOLOv8s algorithm proposed in this paper has increased mAP@50 by 12% and mAP@50:95 by 8.3% on the VisDrone dataset, demonstrating excellent performance compared with other object detection algorithms. On AI-TOD dataset, experimental results validate the generalization of our proposed algorithm, and further prove that our proposed algorithm has higher detection accuracy for small objects if there is more sufficient data support. In addition, the DBYOLOv8l built by us based on YOLOv8l has higher detection accuracy, but the model is larger, which is suitable for tasks with higher detection accuracy supported by high-performance computers. Combined with the existing algorithm foundation and research direction, our future research will focus on the following aspects to tackle difficulties: 1. Explore lightweight technology, reduce model parameters by replacing lightweight backbone or model pruning technology, so that the algorithm can be deployed on embedded devices with low power consumption in the future. 2. Research on small object loss function positioning technology, so that the model can improve the positioning accuracy of dense small objects under complex background. 3. Explore the feature description of different architectures for small objects, and combine the dual-branch idea with CNN architecture and Transformer architecture to further improve the detection accuracy of small objects.

REFERENCES

- [1] Jiang, Huiwei and Peng, Min and Zhong, Yuanjun and Xie, Haofeng and Hao, Zemin and Lin, Jingming and Ma, Xiaoli and Hu, Xiangyun, "A survey on deep learning-based change detection from high-resolution remote sensing images," *Remote Sensing*, vol.14(7), p.1552, 2022.
- [2] P. Mittal, R. Singh, and A. Sharma, "Deep learning-based object detection in low-altitude UAV datasets: A survey," *Image and Vision Computing*, vol. 104, p. 104046, 2020.
- [3] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1137–1149, Jun. 2017
- [4] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988
- [5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, Springer, 2016, pp. 21–37.
- [6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016, , pp. 779–788
- [7] G. Jocher, A. Chaurasia, and J. Qiu. (2023). *Ultralytics YOLO (Version 8.0.0)*. [Online]. Available: <https://github.com/ultralytics/ultralytics>
- [8] Chen, Chenyi, Liu, Ming-Yu, Tuzel, Oncel, and Xiao, Jianxiong. "R-CNN for Small Object Detection", in *Computer Vision – ACCV 2016*, pages 214–230
- [9] Lin, Tsung-Yi, Maire, Michael, Belongie, Serge, Hays, James, Perona, Pietro, Ramanan, Deva, Dollár, Piotr, and Zitnick, C. Lawrence. "Microsoft COCO: Common Objects in Context". In *Computer Vision – ECCV 2014, Lecture Notes in Computer Science*, pages 740–755, 2014.
- [10] T. Liang, X. Chu, Y. Liu, Y. Wang, Z. Tang, W.-T. Chu, J. Chen, and H. Ling, "CBNetV2: A Composite Backbone Network Architecture for Object Detection," *Cornell University - arXiv*, Jul. 2021.
- [11] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "Yolov9: Learning what you want to learn using programmable gradient information," *arXiv preprint arXiv:2402.13616*, 2024.
- [12] Z. Yan, H. Zheng, and Y. Li, "Detail injection with heterogeneous composite backbone network for object detection," *Multimedia Tools and Applications*, vol. 81, no. 8, pp. 11621–11637, 2022.
- [13] Y. Zhang, M. Ye, G. Zhu, Y. Liu, P. Guo, and J. Yan, "FFCA-YOLO for small object detection in remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

- [14] R. Sunkara and T. Luo, "No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects," in *Joint European conference on machine learning and knowledge discovery in databases*, Springer, 2022, pp. 443–459.
- [15] Lin, Tsung-Yi, Dollár, Piotr, Girshick, Ross, He, Kaiming, Hariharan, Bharath, and Belongie, Serge. "Feature Pyramid Networks for Object Detection". In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [16] Liu, Shu, Qi, Lu, Qin, Haifang, Shi, Jianping, and Jia, Jiaya. "Path Aggregation Network for Instance Segmentation". In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.
- [17] M. V. Reddy, K. A. Reddy, M. S. S. Goud, G. Hemanth, and K. Lohith, "Efficient Det: Scalable and Efficient Object Detection," *NeuroQuantology*, vol. 20, no. 19, p. 5559, 2022.
- [18] Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, and Guiguang Ding, "Yolov10: Real-time end-to-end object detection," *arXiv.2405.14458*, 2024. 1, 3.
- [19] D. Ouyang, S. He, G. Zhang, M. Luo, H. Guo, J. Zhan, and Z. Huang, "Efficient multi-scale attention module with cross-spatial learning," in *International Conference on Acoustics, Speech and Signal Processing*, 2023, pp. 1–5.
- [20] Yangyang Li, Qin Huang, Xuan Pei, Yanqiao Chen, Licheng Jiao, and Ronghua Shang, "Cross-layer attention network for small object detection in remote sensing imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pages 2148–2161, 2020.
- [21] Jiangfan Zhang, Yan Zhang, Zhiguang Shi, Yu Zhang, and Ruobin Gao, "Unmanned Aerial Vehicle Object Detection Based on Information-Preserving and Fine-Grained Feature Aggregation", *Remote Sensing*, vol. 16, no. 14, 2024.
- [22] Rui Zhong, Ende Peng, Ziqiang Li, Qing Ai, Tao Han, and Yong Tang, "SPD-YOLOv8: an small-size object detection model of UAV imagery in complex scene", *The Journal of Supercomputing*, Springer, 2024, pages 1–21.
- [23] Lingjie Jiang, Baoxi Yuan, Jiawei Du, Boyu Chen, Hanfei Xie, Juan Tian, and Ziqi Yuan, "MFFSODNet: Multi-Scale Feature Fusion Small Object Detection Network for UAV Aerial Images", *IEEE Transactions on Instrumentation and Measurement*, 2024.
- [24] Dawei Du, Pengfei Zhu, Longyin Wen, Xiao Bian, Haibin Lin, and Qinghua Hu et al. "VisDrone-DET2019: The Vision Meets Drone Object Detection in Image Challenge Results", in *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 213–226, 2019.
- [25] Yuan, Xiang, Cheng, Gong, Yan, Kebing, Zeng, Qinghua, and Han, Junwei. "Small Object Detection via Coarse-to-fine Proposal Generation and Imitation Learning". In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6294–6304, 2023.
- [26] Wang, Jinwang, Yang, Wen, Guo, Haowen, Zhang, Ruixiang, and Xia, Gui-Song. "Tiny Object Detection in Aerial Images". In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 3791–3798, 2021.
- [27] Lin, Tsung-Yi, Goyal, Priya, Girshick, Ross, He, Kaiming, and Dollár, Piotr. "Focal Loss for Dense Object Detection". In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2999–3007, 2017.
- [28]] G. Jocher. (2020). *YOLOv5 By Ultralytics*. [Online]. Available: <https://github.com/ultralytics/yolov5>
- [29] Xingkui Zhu, Shuchang Lyu, Xu Wang, and Qi Zhao, "TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios", in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, Oct. 2021.
- [30] Z. Zhang, "Drone-YOLO: an efficient neural network method for target detection in drone images," *Drones*, vol. 7, no. 8, p. 526, 2023.
- [31] Tang, Shiyi, Zhang, Shu, and Fang, Yini. "HIC-YOLOv5: Improved YOLOv5 For Small Object Detection". In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6614–6619, 2024.
- [32] Chuyi Li, Lulu Li, Yifei Geng, Hongliang Jiang, Meng Cheng, Bo Zhang, Zaidan Ke, Xiaoming Xu, Xiangxiang Chu, "YOLOv6 v3.0: A Full-Scale Reloading," *arXiv*, preprint arXiv:2301.05586
- [33] Wang, Chien-Yao, Bochkovskiy, Alexey, and Liao, Hong-Yuan Mark. "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors". In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7464–7475, 2023.

Eagle Framework: An Automatic Parallelism Tuning Architecture for Semantic Reasoners

Haifa Ali Al-Hebshi¹, Muhammad Ahtisham Aslam², Kawther Saeedi³

Information Systems Department-Faculty of Computing and Information Technology, King Abdulaziz University
Jeddah, 21589, Saudi Arabia^{1,3}

Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, Berlin, 10589, Germany²

Abstract—Parallel semantic reasoners use parallel architectures to improve the efficiency of reasoning tasks. Studies in semantic reasoning rely on manual tuning to configure the degree of parallelism. However, manual tuning becomes increasingly challenging as ontologies become massive and complex. Studies in related fields have developed automatic tuning frameworks using optimization search methods. Although these methods offer performance gains, reducing search time and space size is still an open problem. This study aims to bridge the gap in semantic reasoning and the problem in existing search methods. To achieve these aims, we propose Eagle Framework (EF), an innovative automatic tuning framework designed to improve the performance of parallel semantic reasoners. EF automatically configures the degree of parallelism and calculates the performance data. It incorporates a novel search space and algorithm, inspired by the AVL tree, that efficiently identifies the optimal degree of parallelism. In a case study, EF completed the tuning processes in seconds to a few minutes, achieving performance gains up to 65 times faster than common search methods. The reliability findings, with ICC scores ranging from 0.90 to 0.99, confirmed the consistency of the performance data calculated by EF. The regression analysis revealed the effectiveness of EF in identifying the factors that affect reasoning scalability, with the conclusion that the size of the ontology is the dominant factor. The study underscores the need for adaptive approaches to tune the degree of parallelism based on the size of the ontology.

Keywords—Automatic tuning; parallel semantic reasoning; performance optimization; ontology; high-performance computing

I. INTRODUCTION

In today's digital landscape, machines use ontologies, structured knowledge representations, to process and infer information across domains, forming a larger knowledge base known as Linked Open Data (LOD) [1]. Ontologies are essential for applications such as the semantic web, artificial intelligence, and data-driven decision-making [2], enabling machines to derive insights from complex relationships [3]. However, as ontologies grow in size and complexity, especially with the rise of the Internet of Things (IoT) and social networks, reasoning over these vast knowledge graphs becomes challenging [4]. Traditional semantic reasoners, which use sequential processing, struggle to scale with the growing size of ontologies, leading to significant delays in driving inferences [4], [5].

Fortunately, parallel reasoning systems have emerged, leveraging multicore and distributed computing technologies to improve efficiency [6]. These systems divide reasoning tasks into smaller units, allowing concurrent processing between multiple computing cores or nodes, thus improving speed and

performance [7]. Despite these improvements, the rapid growth of data from IoT systems and social networks continues to add complexity to the reasoning process, demanding more scalable and efficient solutions [8]. Therefore, optimizing parallel reasoning systems to successfully manage performance while addressing the increasing complexity of ontologies is a major challenge for researchers.

Numerous automatic tuning approaches have shown significant improvements in different application areas, such as hyperparameter tuning to optimize machine learning models [9], [10], big data analytics systems [11], [12], and parallel programs [13], [14]. These approaches varied in their strategies and techniques. Optimization search methods have shown accurate results in finding optimal solutions compared to other approaches. However, studies have reported a significant challenge in reducing search time as search space increases.

Linking these challenges in the areas of semantic reasoning and optimization, we present the Eagle Framework (EF), an advanced modular and extensible tool to optimize the performance of parallel reasoning systems. EF automatically generates parallelism configurations, recording performance data, and identifying the optimal degree of parallelism. Operating as a black-box solution on top of existing parallel semantic engines, EF relieves researchers and developers of the time-consuming process of manual tuning. A key innovation of EF is its novel search algorithm, which integrates an AVL tree with a priority queue, enabling efficient exploration for the optimal thread configuration. EF is implemented in Java, ensuring compatibility with a wide range of operating systems and server environments. In addition, EF saves performance data in CSV format and high-resolution line charts for data visualization. In general, EF significantly streamlines the optimization process, ensuring optimal performance and offering valuable time efficiency for researchers to develop scalable and advanced parallel reasoning systems.

In addition to proposing a tuning framework, this study includes a comprehensive case study, in which we validate the performance and reliability of our framework, as well as its effectiveness in optimizing parallel reasoners. We utilized multiple statistical methods through the assessment process, but a key method is introducing the Interclass Correlation Coefficient (ICC) for the reliability analysis. We will explain how we applied the Interclass Correlation Coefficient (ICC) to evaluate the reliability of tuning systems. To the best of our knowledge, ICC has not been used in evaluating automatic tuning.

The remainder of this study is organized as follows. Section II provides a comprehensive review, highlighting the gaps in the existing literature. Section III presents the conceptual and technical design of the EF architecture. Section IV details the case study that evaluates EF from the perspectives of performance, reliability, and effectiveness. Section V presents the findings of the study. Section VI discusses the findings and insights derived from this study. Finally, Section VII concludes the study.

II. BACKGROUND AND RELATED WORK

In this section, we provide a brief review of existing work to highlight the gaps in the literature on semantic reasoning and optimization. This review begins with the challenges of manual tuning presented in the context of semantic reasoning. Then, we provide an overview of the automatic tuning approaches commonly applied in other domains. We review related works with existing optimization methods. For each section of this review, we grouped studies based on the tuning approach or optimization method.

A. Challenges in Tuning Parallel Semantic Reasoners

Semantic reasoning is critical in applications that require logical consistency and explainability, such as medical, bioinformatics, and law. However, the development of parallel semantic reasoning systems has been less active in the last ten years, and most of these systems have been abandoned and no longer maintained [15]. Research studies in semantic reasoning implemented manual tuning to adjust the degree of parallelism. Examples of these studies are [6], [16], [17]. Although these studies did not explicitly state the drawbacks of manual tuning approaches in their research results, recent studies highlighted the challenges of computational complexity and performance in automated reasoning [18] [19]. Reasoning tasks use algorithms that require extensive computations and resource allocation [20]. These scalability challenges are further complicated by the increase in ontology sizes and hierarchies that require intensive computation. Although parallel semantic reasoners leverage multicore and distributed architectures to tackle scalability issues, without automatic tuning strategies, reasoning scalability remains challenging. The Table I provides a comparison of various approaches based on different factors and parameters.

B. Overview of Automatic Tuning Approaches

This section reviews recent studies on optimizing parallel systems. It classifies them into six approaches, following the categorization by Herodotou et al. [21], and discusses the advantages, limitations, and methods for each approach.

1) *Search-based approach*: The search-based approach systematically searches for the optimal solution through experiments, improving performance and resource efficiency. Although this approach is reliable for identifying the optimal or near-optimal solution, it can be computationally expensive for large systems. Van Werkhoven introduced an automatic framework that integrated various optimization search, including simulated annealing and particle swarm optimization, to optimize GPU kernels in OpenCL and CUDA [22].

2) *Rule-based approach*: The rule-based approach uses predefined guidelines and domain expertise to guide tuning decisions, offering simplicity and fast execution. It works well in predictable environments where the behavior of the system follows established patterns. However, its lack of flexibility limits its effectiveness in complex or dynamic workloads. Schwarzrock et al. applied this approach to enhance performance and energy efficiency in NUMA systems. Their focus was on optimizing thread-to-core mapping, memory page mapping, and thread throttling [23].

3) *Machine learning approach*: The machine learning approach employs models, such as regression and neural networks, to predict optimal configurations by learning from historical data, capturing complex relationships for improved tuning accuracy. When trained on quality data, these models can achieve near-optimal configurations without exhaustive searches. However, they require substantial data and computational resources, and accuracy is dependent on data quality. Fan et al. used a random forest model to optimize query performance in databases by predicting optimal degrees of parallelism [24].

4) *Adaptive approach*: The adaptive approach dynamically tunes parameters in real-time, adjusting to workload changes, making it ideal for dynamic environments where static tuning fails. It can achieve near-optimal configurations quickly without exhaustive searches, though it may introduce overhead and struggle with stability in highly volatile conditions. Vogel et al. proposed reactive self-adaptive strategies to control parallelism in stream processing systems, allowing real-time adjustments without the need to restart applications [25].

5) *Cost modeling approach*: Cost modeling estimates resource costs, such as memory and CPU usage, for different tuning configurations, helping to avoid costly trial runs. It provides fast and moderately accurate estimations; however, its limitations lie in the accuracy of the model as it may not capture all the dynamics and interactions in the real world. This limits its ability to find optimal configurations in complex environments. Siddiqui et al. introduced a machine learning-enhanced framework to improve the accuracy of cost modeling in big data systems [26].

6) *Simulation-based approach*: The simulation-based approach models the behavior of the system in a simulated environment to predict optimal settings without affecting live performance. This is especially useful for difficult-to-test scenarios, offering reliable approximations if the simulation models are accurate. However, accuracy depends on model fidelity and detailed simulations can be resource-intensive. Liu et al. developed HSim, a Hadoop simulator for modeling various performance parameters in cloud computing [27].

C. Existing Optimization Search Methods

Among the six approaches previously discussed, we were particularly motivated by the reliability of search-based methods in finding optimal solutions. This section focuses on studies that implement these search-based methods. Currently, search methods are applied in system optimization in related domains, big data, machine learning, and high performance computing. These studies provide valuable insights that inform the design of our solution for optimizing parallel reasoning performance.

TABLE I. SUMMARY OF COMMON PARALLELISM TUNING APPROACHES

Approach	Search-based	Rule-based	Machine Learning	Adaptive	Cost Modeling	Simulation-based
Methods	Search algorithms.	Based on heuristics.	Data-driven predictions	Real-time dynamic adjustments.	Cost estimation models.	System behavior simulation.
Advantages	High-quality solutions.	Simple, fast decisions.	Adapts to changing conditions	Real-time tuning.	Guides resource decisions.	Tests without real-world impact.
Drawbacks	Expensive computing costs.	Expert knowledge required.	Large data needed, Slow training.	Struggles with unexpected changes.	May overestimate real-world factors.	Expensive computing costs.
Domain	Large, complex systems.	Simple, predictable systems.	Learning from history.	Workloads that change.	Resource-constrained systems.	Expensive or risky scenarios.
Data Size	Moderate to large.	Small to moderate.	Large.	Small to medium.	Small to medium.	Large.

1) *Grid Search (GS)*: GS is an optimization method that systematically explores all possible parameter combinations to determine the optimal one. However, this exhaustive approach is computationally expensive, particularly for high-dimensional search spaces [28].

Recent studies have demonstrated the potential of GS in optimizing systems in various domains. For example, George and Sumathi applied GS to optimize a random forest classifier for sentiment analysis, leading to improved accuracy [29]. Similarly, Priyadarshini and Cotton used GS to tune a deep neural network model for sentiment analysis, achieving an accuracy above 96%, which outperformed several baseline models [30]. In the big data domain, Chen et al. incorporated GS into their system to optimize MapReduce performance on Hadoop clusters, identifying optimal configurations to minimize running times [31]. In a similar study, Sewal and Singh compared GS with other optimization methods such as Evolutionary Optimization and Random Search to fine-tune Apache Spark. They found that GS was effective in reducing execution times by 23.24%. [32]. These studies are examples among others that utilize determinism in GS to enhance performance in areas like machine learning and big data systems.

2) *Hill Climbing (HC)*: HC is a simple and efficient local search algorithm that iteratively improves an initial solution by exploring neighboring options. However, it may get stuck in local optima, reducing the chances of finding the global.

Recent studies have highlighted the effectiveness of HC in various fields. For example, Sivakumar and Mangalam introduced a technique to improve adaptive cruise control systems in automated vehicles, using a combination of search methods, including HC, to optimize vehicle parameters, improving safety and fuel efficiency [33]. Zeng et al. employed a simple HC method with machine learning techniques to optimize parallelism in Parallel Nesting Transactional Memory (PN-TM) systems, achieving faster convergence and higher accuracy compared to other optimization methods [34]. Pradhan et al. applied HC to optimize a CNN model for classifying COVID-19 from chest X-ray images, improving its performance metrics and outperforming other hybrid techniques [35]. These studies are among several that demonstrate HC as a versatile and effective optimization method to improve system performance in diverse domains, from automated vehicles to machine learning and medical image classification.

3) *Simulated Annealing (SA)*: SA is a probabilistic optimization algorithm that explores various solutions, including the worst ones to escape local optima. SA is widely used method for many optimization problems due to its adaptability and capability to navigate rugged search spaces [36]. However, it can be computationally intensive, sensitive to parameter choices, and slow to converge.

Recent studies demonstrate the adaptability and effectiveness of SA in finding optimal solutions for complex optimization problems. A study by Rasch et al. utilized SA within their Auto-Tuning Framework (ATF) to optimize interdependent parameters in parallel programs, using the chain of trees and coordinate search spaces to enhance multidimensional exploration and improve tuning efficiency [7]. Gülcü and Kuş introduced the multi-objective simulated annealing algorithm for optimizing hyperparameters in convolutional neural networks, balancing classification accuracy and computational complexity, and achieving superior results compared to traditional SA [37]. Similarly, Abdel-Basset et al. combined SA with Harris Hawks Optimization to enhance feature selection for classification tasks, using SA to escape local optima and explore better feature subsets effectively [38]. These studies highlight the effectiveness of SA, both as a standalone algorithm and within hybrid frameworks, in solving diverse optimization problems across fields.

4) *Random Search (RS)*: Random Search (RS) is a simple algorithm that explores a search space by randomly sampling points and updating the best solution found. Although its simplicity and ability for global exploration, it lacks efficiency in high-dimensional spaces.

Recent studies underscore the role of RS as a simple and effective optimization method in different applications. A study by Willemsen et al. proposed a standardized benchmarking methodology for evaluating automatic tuning frameworks. This methodology integrates RS as a baseline for benchmarking optimization algorithms [39]. A similar study by Deligkaris used RS as a baseline method to benchmark evobps, an algorithm based on particle swarm optimization, against 12 related methods and models. The benchmark results demonstrated the effectiveness of RS in the search for neural architectures [40]. Hosseini et al. employed RS in optimizing 10 hyperparameters of Long Short-Term Memory (LSTM) networks used in rainfall-runoff modeling, resulting in highly precise predictions for hourly stream-flow and water levels in Spain's Basque Country [41]. Despite the advancement of more sophisticated

methods, these studies highlight the effectiveness of RS as a baseline or complementary method in advanced optimization methods.

D. Limitations and Research Gaps

Section II-A presented studies on parallel reasoning, which currently rely on manual tuning for optimization and scalability analysis. As research labs, governmental sectors, and universities continue to develop ontologies, the need for scalable and efficient reasoners has increased. Furthermore, the substantial efficiency and accuracy of automatic tuning reported in studies applied in other domains, such as machine learning and big data, highlight that the lack of automatic tuning application in semantic reasoning is a considerable gap. To the best of our knowledge, the application of automatic tuning in semantic reasoning has not been explored. Therefore, addressing this gap is the main objective of this study.

The studies reviewed in Section II-C noted the advances gained from applying search optimization methods. However, they also noted significant limitations of these optimization methods (see Table II). For GS, though it is exhaustive and guaranteed to find the optimal solution, it is computationally expensive, especially in multidimensional spaces where the time increases exponentially with increase in the search space. On the other hand, HC, SA, and RS offer greater efficiency and scalability; however, they are limited by several drawbacks, such as risks of local optima, stochastic behavior, and incomplete search space exploration. These limitations underscore the trade-off between time complexity and the guarantee of identifying the optimal solution, and this trade-off is strongly correlated with search space size. As noted by Krestinskaya et al., optimizing search time for large search spaces remains a critical gap and an open research challenge in optimization [42]. Therefore, the second objective of this study is to address the complexity of search time by designing a deterministic algorithm and a tree-based search space that aims to guarantee finding the optimal solution in efficient time.

TABLE II. SUMMARY OF SEARCH OPTIMIZATION METHODS

Aspect	Grid Search	Hill Climbing	Simulated Annealing	Random Search
Exploration	Global	Local	Global (with refinement)	Global (with random sampling)
Efficiency	Very Low	High	Medium	Low
Dimensional Scalability	Poor	Poor	Good	Good
Risk of Local Optima	No	High	Low	No
Best Use Case	Small spaces	Small spaces	Large spaces	Large spaces
Time Complexity	$O(n^k)$	$O(k \cdot n)$	$O(k \cdot n)$	$O(k \cdot n)$

III. EAGLE FRAMEWORK (EF)

This section presents the architectural and algorithmic design of EF. Before diving into architectural design, we provide an overview of the EF multi-layered system, where the EF resides in the middle layer. This overview is essential for

understanding the interoperability and portability of EF, which allows it to function with various parallel reasoning systems on different platforms. Following the system overview, we will explore the architectural details of EF from both mathematical and algorithmic perspectives.

A. System Overview

EF is a modular architecture that operates within a multi-layered system, where EF occupies one layer alongside a parallel reasoner. The abstraction view of the EF system consists of five layers, as depicted in Fig. 1.

The first layer is the Command Line Interface (CLI), which accepts parameter values to set automation parameters and is responsible for displaying output results. The second layer represents the main contribution of this study, where EF and the parallel semantic reasoner are positioned. EF comprises one main algorithm and three auxiliary algorithms: *Thread Controller*, *Speedup Calculator*, and *Parallelism Optimizer*. The third layer is the java runtime environment, which serves as a bridge between the EF implementation and the operating system. This middle layer provides the resources needed to compile and execute EF classes on any machine. The fourth layer is the operating system, which abstracts physical hardware and manages system resources. The final layer is the hardware, which represents the physical components of the machine, such as memory and the CPU. The next section focuses on the second layer, detailing the EF architectural components and their interactions with the other layers.

B. EF Architecture

EF architecture consists of a main algorithm and auxiliary algorithms. The main algorithm represents the core architecture of the EF and the interface that manages the connection between the CLI layer, the parallel reasoner, and the auxiliary algorithms. The auxiliary algorithms help the main algorithm in the tuning process by adjusting the thread count, calculating the speedup factor, and identifying the optimal thread count. Fig. 2 presents a flow chart for the EF architecture. The following sections provide an in-depth explanation of the functionality of each algorithm within the EF architecture.

1) *The Main algorithm*: As shown in Fig. 2, the main algorithm comprises five phases: automation setup, sequential reasoning, parallelism tuning, optimization, and output formatting.

a) *Phase 1: Automation setup*: The main algorithm starts by accepting three parameter values from the CLI layer: the path to the ontology file (p), the scale difference (d), which defines the incremental scale used to calculate thread configurations, and the maximum thread count (m), which serves as a threshold to limit the generation of additional configurations. Based on these parameters, the number of thread configurations is proportional to the values of d and m . Additionally, it creates a tree (A) to use in the search for the optimal thread count (o), and a list (L) to store performance data. It also sets the thread count (n) and speedup factor (s) to the seed value of 1.

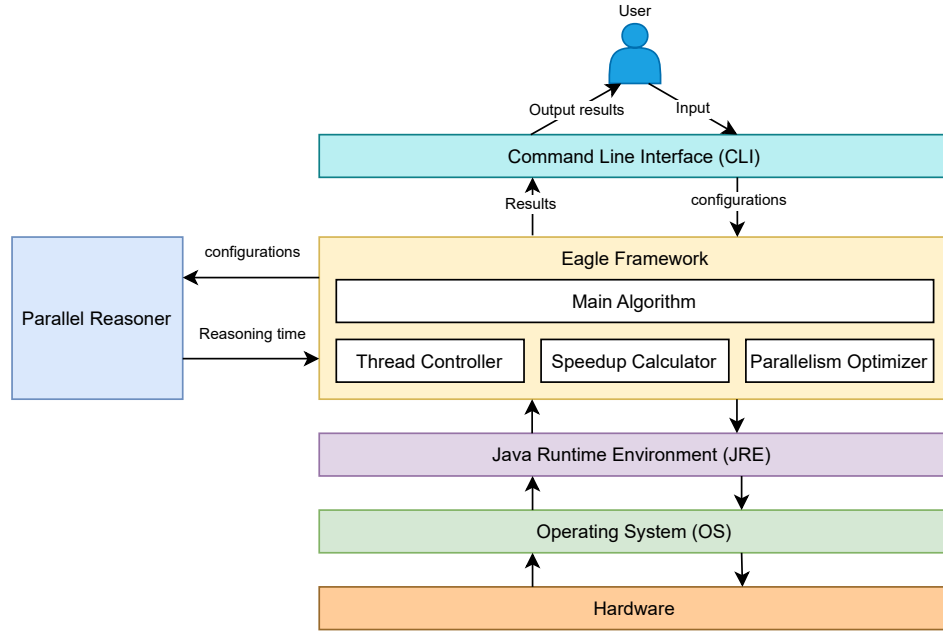


Fig. 1. EF System overview.

b) Phase 2: Sequential reasoning: In this phase, the main algorithm runs the parallel reasoner, $\text{runReasoner}(p, n)$, sequentially using one thread. Then, the main algorithm saves sequential reasoning time in a variable named T_{seq} . Since EF does not perform parallel reasoning at this phase, the speedup factor remains 1. Consequently, the main algorithm inserts the values of n and s in A and adds them with T_{seq} to L .

c) Phase 3: Parallelism tuning: This phase represents the core automation provided by EF. The main algorithm starts the automation by checking whether n does not exceed m . This step is essential to limit the EF from generating more configurations. If n is less than or equal to m , the main algorithm passes n and d in a call to the Thread Controller $\text{ctrlThreads}(n, d)$. After receiving the new n value from the Thread Controller, the main algorithm passes n with p in a call to the reasoner and stores the parallel reasoning time in a variable named T_{par} . Subsequently, the main algorithm sends T_{seq} and T_{par} in a call to the Speedup Calculator, $\text{calcSpeedup}(T_{seq}, T_{par})$, to find the ratio and save it in s . Finally, the main algorithm performs the necessary operations to store performance data in A and L .

d) Phase 4: Optimization: In this phase, the main algorithm passes A in a call to the Parallelism Optimizer, $\text{optParallelism}(A)$, which searches for o in A . Once o is identified, the main algorithm prints o on the console. The parallelism tree and optimizer will be explained comprehensively in Section III-B4.

e) Phase 5: Output formatting: This phase is the final stage in EF, where performance data L are formatted into a comma separated value file (CSV). This format was chosen for its compatibility with most statistical analysis and database systems, which allows direct processing by analysis tools. In addition, a line chart is generated to illustrate the relationship

between each thread count and the associated reasoning time.

2) Thread controller: This section explains the algorithm of Thread Controller, denoted as $\text{ctrlThreads}(n, d)$, designed to create thread configurations. Building on recommendation by Huang et al. [43], who recommended employing sampling strategies for generating configurations, this algorithm generates thread configurations systematically using a specified scale difference, d . To prevent incorrect input, we added a conditional statement that verifies the value of d entered by the user. If d is assigned a negative number or zero, the Thread Controller sets d to the seed value of 1 and recalculates n accordingly. Otherwise, the Thread Controller computes the new n based on the provided value of d .

Definition 1. Let n be a thread configuration and d be the scale difference. The Thread Controller calculates is defined as:

$$\text{ctrlThreads}(n, d) = \begin{cases} n + 1 & \text{if } d \leq 0 \\ n + d & \text{if } d > 0 \end{cases} \quad (1)$$

3) Speedup calculator: Speedup is a common metric that is used to assess performance improvements in systems running on parallel computing architectures. To measure speedup, we designed Speedup Calculator, denoted as $\text{calcSpeedup}(T_{seq}, T_{par})$, an algorithm that calculates the speedup factor as the ratio between the execution time for sequential reasoning and the time taken for parallel reasoning. Since most reasoners record reasoning times in milliseconds, we took account of scenarios where executing the reasoner on massively parallel computing resources results in reasoning times in fraction of a millisecond rounded to zero (i.e. in nanoseconds). To handle this, the Speedup Calculator first checks if the value of T_{par} is non-zero. If this condition is

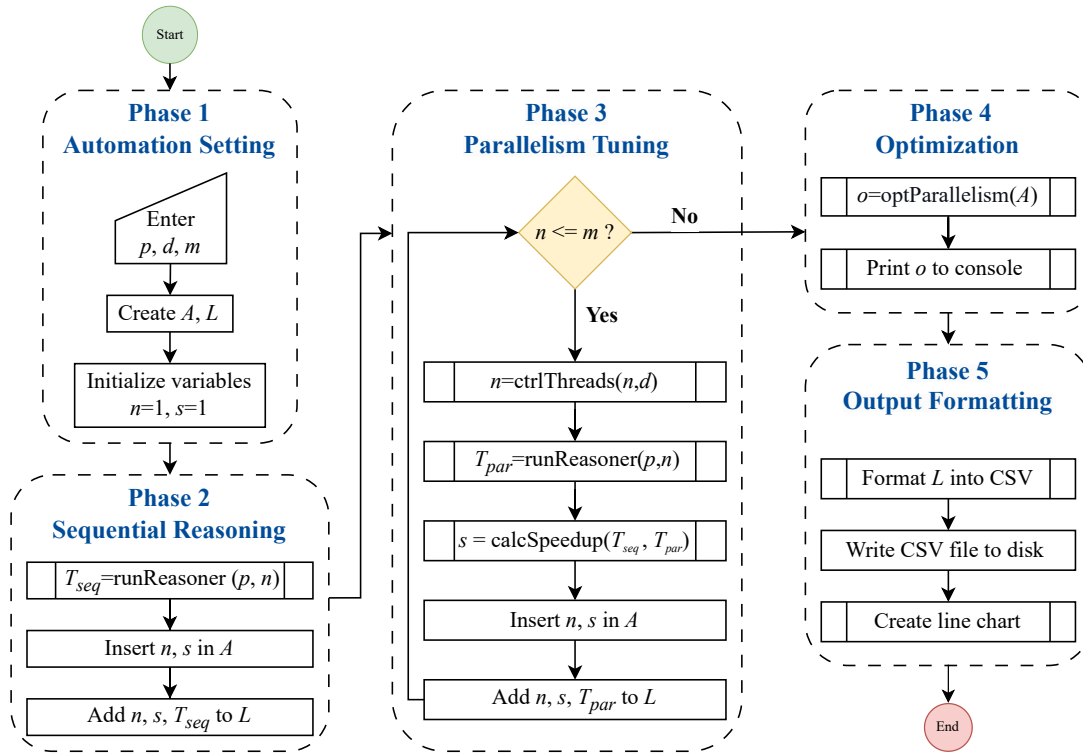


Fig. 2. EF Architecture.

met, the calculator proceeds with the division. However, if T_{par} equals zero, the calculator throws an arithmetic exception, which is treated as “undefined” in arithmetic, and the process safely halts.

Definition 2. Let T_{seq} be the execution time of the sequential reasoning, and T_{par} be the execution time of the parallel reasoning. The Speedup Calculator is defined as:

$$calcSpeedup(T_{seq}, T_{par}) = \begin{cases} Undefined & \text{if } T_{par} = 0 \\ \frac{T_{seq}}{T_{par}} & \text{if } T_{par} \neq 0 \end{cases} \quad (2)$$

4) *Parallelism Tree (PT) and Parallelism Optimizer (PO)*: An AVL tree is a self-balanced binary search tree characterized by its speed in most operations, including insertion and searching [44] [45]. We chose the AVL tree to construct the search space in EF due to its time efficiency, with a worst-case time complexity of $O(\log n)$ for core operations, which outperforms other data structures to align with the objective of this study. For clarity, we used Parallelism Tree (PT) to refer to the tree-based search structure and Parallelism Optimizer (PO) to refer to the associated optimization algorithm.

PT is a modified version of the AVL tree, where each node consists of performance data (s) and its associated thread configuration (n). In addition, each node has pointers to a left child (l) and a right child (r). Unlike the AVL tree, PT uses s as the key to determine the correct position to insert a new node. In PT, the sequential reasoning node is always the root, while the parallel reasoning nodes are placed based on their

speedup factor. Algorithm 1 presents the insert procedure for PT.

Algorithm 1: insert(s, n)

- 1: **Input:** s, n
 - 2: **Output:** rebalance($root$)
 - 3: **if** $root = null$ **then**
 - 4: **return** new Node(s, n)
 - 5: **else**
 - 6: **if** $root.s > s$ **then**
 - 7: $root.l \leftarrow$ insert(s, n)
 - 8: **else if** $root.s < s$ **then**
 - 9: $root.r \leftarrow$ insert(s, n)
 - 10: **else**
 - 11: $root.nQueue.add(n)$
 - 12: **end if**
 - 13: **end if**
 - 14: **return** rebalance($root$)
-

In the initial experiments, we observed that different thread configurations produced identical speedup factors. Consequently, PT prevents the insertion of nodes with duplicate speedup. To resolve this issue, we integrated a priority queue within the PT node structure to store all thread configurations associated with the same speedup factor. This approach enables PT to encapsulate each speedup factor with its corresponding thread configurations in the same node. The priority queue orders the configurations from the smallest to the largest, enabling efficient exploration by neglecting the less effective configurations. In this study, we designed PO to select the

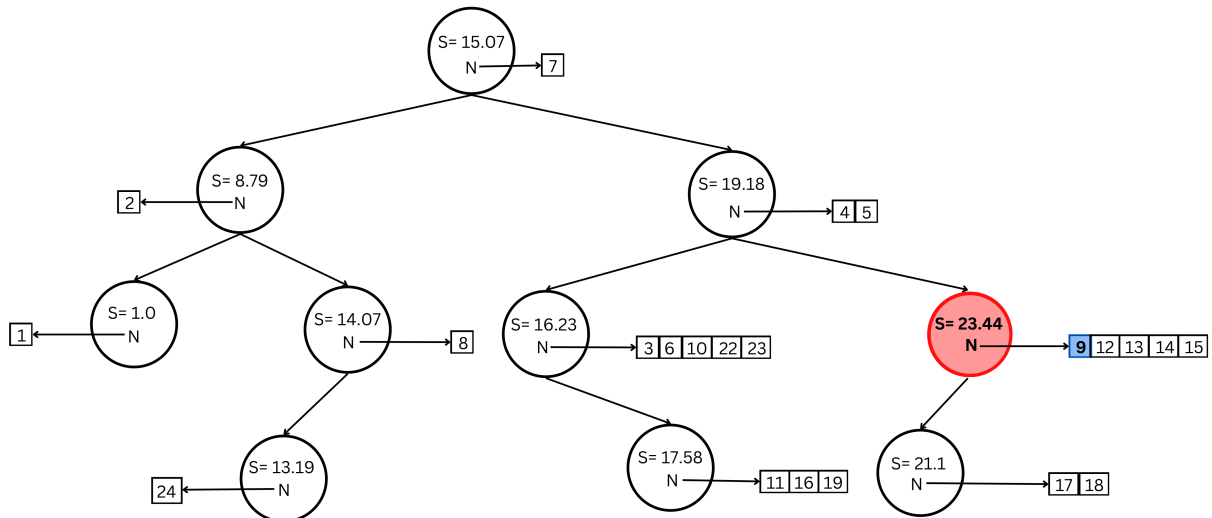


Fig. 3. Parallelism tree (PT) in EF where each node contains speedup factor (s) and the associated thread configuration queue (n).

smallest thread configuration as the optimal one, under the assumption that this configuration achieves the highest speedup factor and that performance will not improve beyond this point. Fig. 3 illustrates an example of PT resulting from one tuning experiment. The red circle denotes the node with the maximum speedup value, while the blue square indicates the optimal degree of parallelism PO selects from the thread queue.

Similarly to the AVL tree, PT is ordered and places the node with the highest speedup factor at the end of its rightmost path. Therefore, we designed the PO algorithm, denoted as $\text{optParallelism}(A)$, to search for the optimal configuration only in the rightmost path of the PT. Such an approach efficiently saves time compared to search in a multi-dimensional space structure. PO starts the optimization search by checking whether the root's right pointer points to NULL. If so, it returns the root node because it contains the optimal degree of parallelism. If not, it performs this check recursively until it finds the node whose right pointer refers to NULL as the node containing the optimal thread configuration. The PO algorithm is shown in Algorithm 2.

Algorithm 2: $\text{optParallelism}(\text{node})$

- 1: **Input:** node
 - 2: **Output:** node
 - 3: **if** $\text{node.right} == \text{null}$ **then**
 - 4: **return** node
 - 5: **else**
 - 6: **return** $\text{optParallelism}(\text{node.right})$
 - 7: **end if**
-

IV. EXPERIMENTAL DESIGN AND SETUP

This study aims primarily to evaluate EF in terms of performance, reliability, and effectiveness in assessing the scalability of parallel reasoners. To achieve this, we conducted our experiments on a case study on the ELK reasoner, a reasoning engine specifically designed for OWL2 EL ontologies [16].

ELK is one of the few actively maintained parallel reasoners for OWL2, as many other reasoners have been discontinued [15]. Although ELK provides a variety of reasoning services to support ontology development and querying, our experiment scope is only on the classification reasoning service.

We conducted the experiments on the Aziz Supercomputer, where each node consists of two 12-core processors (Intel Xeon CPU E5-2695v2, 2.40 GHz) that support Hyper-Threading Technology, providing a total of 48 logical cores and a total memory of 256 GB (128 GB per processor). Since ELK was designed to operate exclusively on shared memory servers [16] [46], we performed all experiments on a single node. To maintain the integrity of our results, we secured exclusive access to server resources, thereby preventing the interleaving of jobs which could potentially compromise reasoning time and speedup factor. Furthermore, we configured the EF parameters with a scale difference of 1 and set the maximum thread count to 240 in all experiments. Section V-D1 will explain our choice of these parameter values.

Our research used a systematic sampling technique to select the ontologies. First, we downloaded ontologies from BioPortal¹ and OBO Foundry² that support the OWL2 EL profile. Then, we further classified the ontologies based on the number of TBox axioms into different sizes, ranging from tiny to medium. We based our categorization on the number of TBox axioms since the classification reasoning service is associated with only TBox axioms. We selected three ontologies from each size category, resulting in a total of nine ontologies. Each ontology was examined in a separate experiment, and each experiment was repeated three times to ensure reliability, resulting in a cumulative total of 27 experiments. This approach allowed us to ensure a diverse range of ontologies for our experiments while ensuring that the samples represented the entire population of OWL2 EL ontologies.

¹<https://bioportal.bioontology.org/>

²<https://obofoundry.org/>

V. CASE STUDY: VALIDATING THE PERFORMANCE, RELIABILITY, AND EFFECTIVENESS OF EF IN TUNING THE ELK REASONER

ELK is a specialized OWL reasoner that classifies ontologies in the OWL2 EL profile. It is known for its high performance due to its parallel reasoning and robust optimization techniques [16]. ELK has expanded its capabilities to include incremental classification and proof tracing, with optimizations for handling role composition axioms and rewriting low-level inferences. These improvements simplify incremental reasoning, proof generation, and enable automated verification and ontology debugging.

This case study begins with an assessment of EF from performance and readability perspectives and ends with an evaluation of the effectiveness of EF in analyzing the scalability of the reasoning system.

A. Overall Framework Performance

The analysis starts by evaluating the performance of the EF for the overall tuning process, covering ontology loading, configuration generation, performance monitoring, and reasoning optimization. Table III presents the minimum, maximum, average and standard deviation of execution times (in seconds) required for EF to tune and optimize the ELK reasoner. Execution times range from a few seconds to just under five minutes.

For tiny ontologies, such as OLATDV, INO, and FBDV, the tuning process was completed in a few seconds, demonstrating the framework's efficiency in quickly exploring thread configurations. Small-sized ontologies, including PLANA and PDON, exhibited slightly longer tuning times, ranging from 17 to 28 seconds. In contrast, medium-sized ontologies, such as OBA, EMAPA, and ORDO, required significantly more time, with ORDO taking the longest at 4 minutes and 29 seconds. This variation in execution times reflects the influence of ontology size on EF performance, with larger ontologies requiring longer durations.

Narrowing the focus from overall framework performance, the next section presents a detailed comparative analysis of the EF's optimization algorithm against baseline search algorithms commonly used in optimization studies.

B. Parallelism Optimizer vs. Existing Optimization Methods

To evaluate EF's optimization performance, we compared its Parallelism Optimizer (PO) with existing search methods commonly used in optimization studies. Specifically, we compared PO with grid search (GS), hill climbing (HC), simulated annealing (SA) and random search (RS). To conduct a fair comparison, we separated PO from the EF architecture. We used one data set resulted from one tuning experiment for all the algorithms involved in the comparison. For SA, we set the initial temperature at 1000 and the cooling rate to 0.95, while for RS, we set the number of iterations to 100.

A summary of the comparative analysis is shown in Table IV. The results showed that PO significantly outperforms its competitors, achieving an average reasoning time of just 0.003 ms. In contrast, the average reasoning times for the other algorithms were 0.167 ms for GS, 0.090 ms for HC, 0.196 ms

for SA, and 0.128 ms for RS. This high efficiency is further demonstrated by the success rate in identifying the optimal thread configuration. PO perfectly found the optimal thread configuration in all ten attempts, while both HC and SA failed in all attempts, and RS succeeded in only three trials.

C. Data Quality and Reliability

In data assessment, we focus on evaluating the quality of the EF measurements gathered during the tuning process and the consistency between these measurements. Listing 1 represents a sample console output of the type of variable data input, missing values, and duplicate rows in the data collected by the EF. As shown in Listing 1, the data underwent evaluation included the test ID for referencing purposes, thread count, reasoning time in milliseconds and in a format of days, hours, minutes and seconds to ease readability, and corresponding speedup factor. The assessment showed that each variable was correctly formatted in a suitable data type. In addition, it revealed that neither missing values nor duplicate rows were found in the data, reporting data integrity and quality.

```
=====
File: ELK_ORDO_1_240_2.csv
=====
| Variable | Data Types | Missing |
|-----|-----|-----|
| Test Number | object | 0 |
| Number of Threads | int64 | 0 |
| Total Reasoning Time (ms) | int64 | 0 |
| Total Reasoning Time (d:h:m:s) | object | 0 |
| Speedup Factor | float64 | 0 |
| Duplicate Rows: | | 0 |
=====
```

Listing 1. Sample Console Output For EF's Data Quality Assessment.

To assess the reliability of the EF, we used the intraclass correlation coefficient (ICC). We selected ICC over other statistical methods because our study involves repeated experiments conducted in the same environmental settings. ICC is a highly precise statistical method that is sensitive to variance [47]. In addition, ICC can assess both the consistency within and between configurations. It is widely used in other fields such as medicine, psychology, biology, and genetics, especially to evaluate the reliability of measurement tools such as medical instruments and computer-aided detection (CAD) systems [48]. To our knowledge, this is the first study to use ICC in assessing the reliability of automatic tuning methods.

Because the speedup factor is a ratio, we applied the ICC assessment exclusively to the reasoning time. The results of the reliability analysis for different ontology sizes are presented in Table V. Each assessment applied the ICC(3,k) model to a dataset of 720 measurements, calculated as each experiment repeated three times with 240 measurements per trial. As shown in Table V, the analysis of nine ontologies revealed high ICC scores, ranging from 0.789 to 0.992, indicating strong consistency between measurements. All ontologies showed statistically significant results ($p < 0.001$) with narrow confidence intervals, indicating precise measurements. The highest ICC was observed for INO (0.992), followed by OLATDV (0.990) and PLANA (0.962). ORDO and PDON also showed strong ICC scores of 0.951 and 0.932, respectively. However,

TABLE III. REASONING TIME CALCULATED BY EF IN TUNING ELK REASONER (IN SECONDS)

Ontology	TBox Axiom count	Ontology Size ^a	Min. Time	Max. Time	Avg. Time	Median Time	SD Time
OLATDV	88	Tiny	9.24	12.25	10.377	9.637	1.636
INO	384	Tiny	11.36	11.63	11.491	11.489	0.134
FBDV	646	Tiny	10.17	11.71	11.097	11.413	0.815
PDON	1252	Small	15.86	34.64	28.311	34.433	10.786
PLANA	2755	Small	15.89	18.21	17.071	17.113	1.157
WBPHENOTYPE	4026	Small	36.86	43.34	39.703	38.910	3.316
OBA	17811	Medium	137.06	295.03	194.653	151.877	87.241
EMAPA	23029	Medium	74.49	78.92	77.309	78.518	2.448
ORDO	53861	Medium	233.28	287.40	269.103	286.634	31.029

^a Ontology sizes categorized by TBox axiom count: Tiny – fewer than 1,000 axioms; Small – 1,000 to 10,000 axioms; Medium – 10,000 to 100,000 axioms; Large – 100,000 axioms or more.

TABLE IV. BENCHMARKING THE PARALLELISM OPTIMIZER AGAINST EXISTING OPTIMIZATION SEARCH ALGORITHMS

Algorithm	Grid Search	Hill Climbing	Simulated Annealing	Random Search	Parallelism Optimizer
Average Reasoning Time (ms)	0.167	0.090	0.196	0.128	0.003
Success Rate for Optimal Thread Identifications (out of 10)	10/10	0/10	0/10	3/10	10/10
Time Complexity	$O(n)$	$O(n)$	$O(n)$	$O(n)$	$O(\log n)$

TABLE V. RELIABILITY ANALYSIS FOR EF MEASUREMENTS AMONG DIFFERENT SIZES OF ONTOLOGIES

Ontology	ICC	F	P	CI95%
OLATDV	0.989824	98.267647	8.143461e-311	[0.99, 0.99]
INO	0.992448	132.417193	0.0	[0.99, 0.99]
FBDV	0.880456	8.365113	3.314461e-85	[0.85, 0.90]
PDON	0.932163	14.741265	1.651629e-130	[0.92, 0.95]
PLANA	0.962178	26.439423	9.941877e-183	[0.95, 0.97]
WBPHENOTYPE	0.789731	4.755818	9.179020e-48	[0.74, 0.83]
OBA	0.883609	8.591758	3.438310e-87	[0.86, 0.91]
EMAPA	0.889079	9.015433	8.344745e-91	[0.86, 0.91]
ORDO	0.951172	20.48018	2.327254e-159	[0.94, 0.96]

WBPHENOTYPE, with an ICC score of 0.790 and a 95% confidence interval of [0.74, 0.83], showed moderate consistency. Although the ICC score for WBPHENOTYPE was statistically significant, its lower ICC and broader confidence interval indicated weaker consistency compared to the other ontologies.

D. EF Effectiveness in Analyzing Reasoning Performance

This section explores the role of EF in helping researchers with scalability assessments to improve the performance of parallel semantic reasoners. To validate EF effectiveness, we performed exploratory and regression analysis.

Before conducting the evaluation, we combined all EF data resulted from all experiments. Additionally, we added characteristics information for each ontology, including the TBox axiom count and size. we performed the necessary pre-processing and normalization .

1) *Exploratory analysis:* This section explores the impact of varying thread configurations and ontology sizes on the performance of the ELK reasoner. It also examines the relationship between the optimal degree of parallelism and the total number of logical cores. To examine these relationships, we define three examination areas:

- Area 1: less than the total of logical cores.
- Area 2: equal to the total of logical cores.
- Area 3: greater than the total of logical cores.

As mentioned in Section IV , we set the scale difference to 1 and the maximum thread count to 240. These settings ensured the gradual increase in thread configurations with a threshold exceeding the total number of logical cores. In addition, these settings allowed us to cover all the examination areas in a single execution. Fig. 4a, 4b, and 4c demonstrate the scalability of the ELK reasoner with varying ontology sizes and thread configurations. The red dashed line presents a reference mark pointing to the 48 logical cores.

In Fig. 4a, processing tiny ontologies OLATDV, INO, and FBDV displayed optimal performance at around 48 threads, with OLATDV achieved a speedup factor of 30, followed by FBDV of 25 and INO of 16. For OLATDV and INO, the optimal thread configuration was found in Area 1, while for FBDV the optimal solution was found in the first portion of Area 3, after which the performance decreased significantly. Similarly to tiny ontologies, the small ontologies PDON,

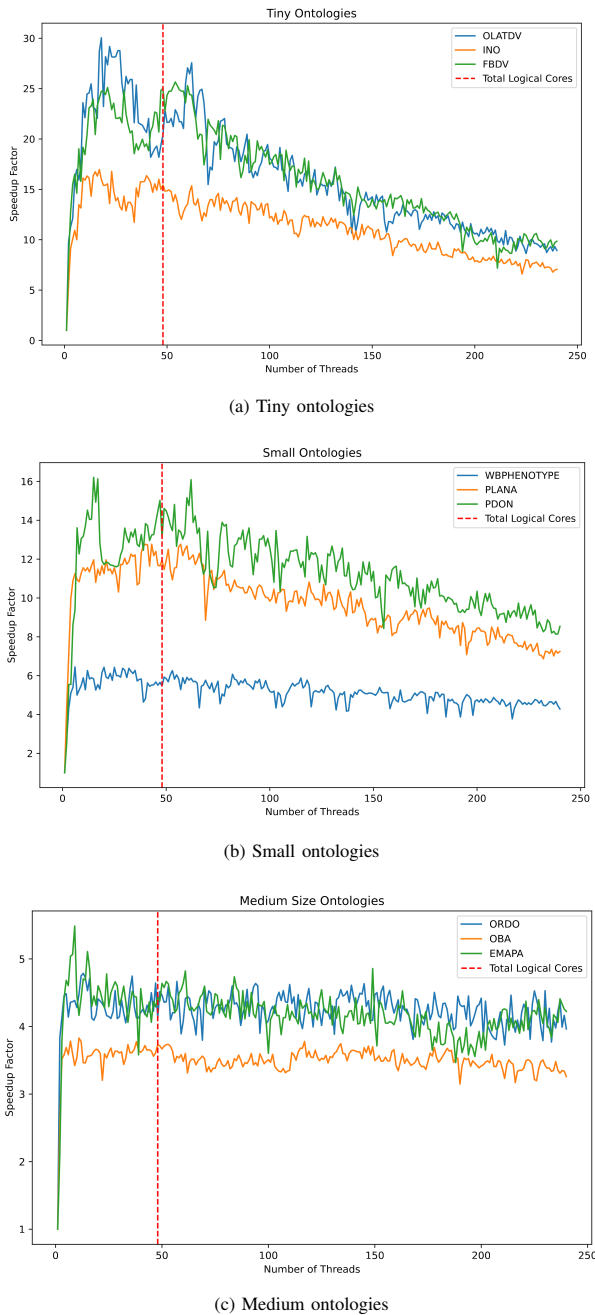


Fig. 4. ELK's Scalability with varying ontology sizes and thread configurations.

PLANA, and WBPHENOTYPE reached the speed factor of 16, 12, and 6, respectively, with optimal configuration found in Areas 1 and 3, as shown in Fig. 4b. However, the range of speedup factors for the small ontologies was narrower than that of the tiny one, as indicated by the reduced scale of the speedup axis in Fig. 4b compared to the axis in Fig. 4a.

In Fig. 4a, processing small ontologies like OLATDV, INO, and FBDV showed optimal performance at approximately 48 threads. Processing OLATDV achieved a speedup factor of 30, FBDV reached 25, and INO managed 16. The best thread

configuration for OLATDV and INO was in Area 1, while FBDV's optimal performance was in the initial part of Area 3, followed by a significant drop. Similarly, small ontologies PDON, PLANA, and WBPHENOTYPE achieved speed factors of 16, 12, and 6, respectively, with optimal configurations in Areas 1 and 3, as shown in Fig. 4b. The speedup range for small ontologies was narrower compared to tiny ones, reflected by the smaller scale of the speedup axis in Fig. 4b compared to Fig. 4a.

Fig. 4c presents notable performance gains for medium-sized ontologies compared to the small ones, where the average speedup for processing ORDO and EMAPA achieved factors between 4 and 5, while performance in processing OBA stabilized at a less speedup factor. Key observations were deduced from this figure. First, there is a notable decrease in the range of speedup factors compared to Fig. 4a and 4b. Second, the optimal thread configurations were identified in Area 1, indicating that adding more threads did not lead to further enhancements. Third, the trend line for this category shows a performance stabilization, suggesting that ELK reasoners benefits from parallelization in reasoning large ontologies more than small ones.

2) *Regression analysis:* In this analysis, we used an Ordinary Least Squares (OLS) regression model to quantify the impact of thread configurations and ontology size, measured in terms of the TBox axiom count, on the reasoning time. This model, with standardized predictors, explained 78.4% of the variance in reasoning time ($R^2 = 0.784$), highlighting its effectiveness in capturing the relationship between predictors and reasoning time. The results, presented in a 3D scatter plot shown in Fig. 5, revealed that while the TBox axiom count significantly affected the time of reasoning ($\beta_1 = 0.8855$, $p < 0.001$), the thread count has a negligible impact ($\beta_2 = 0.0066$, $p = 0.512$). Furthermore, the model showed a notable predictive accuracy, with an average Mean Squared Error (MSE) of 0.2165 and a Root Mean Squared Error (RMSE) of 0.4653.

VI. DISCUSSION

The rapid expansion of ontologies and the lack of automatic tuning approaches have poses challenges on advancing parallel semantic reasoners. In related domains, several sophisticated tuning frameworks have been developed, applying existing search methods for optimization. Although existing search methods presented significant improvements, reducing their search time is still an active research field. This study addressed these gaps by proposing an automatic tuning methodology with an innovative tree-based search algorithm.

Our case study presented in Section V validated the performance gains, reliability, and effectiveness of automatic tuning in optimizing parallel semantic reasoners. Using ELK reasoner as a case study, our methodology, Eagle Framework (EF), efficiently completed the entire tuning process, from ontology loading to final optimization results, in less than five minutes across 240 thread configurations. Practically, such efficiency cannot be achieved in manual tuning approaches, underscoring the importance of applying automatic tuning methods to optimize the performance of semantic reasoner. These findings align with the conclusion of Mustafa's study,

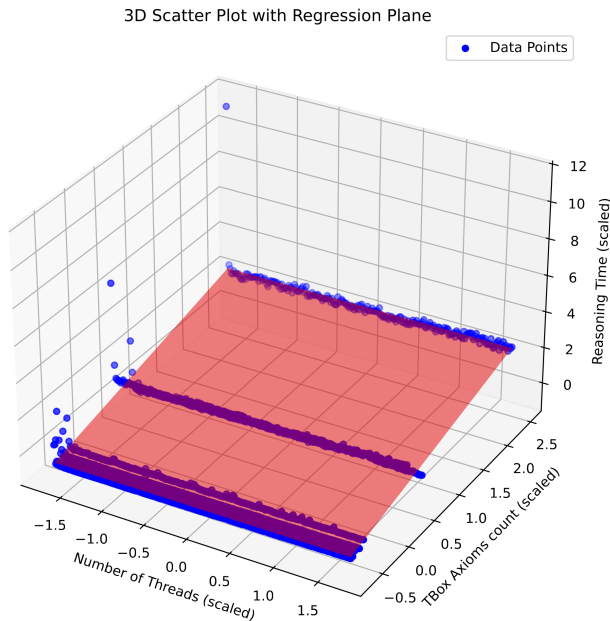


Fig. 5. A 3D Scatter plot with regression plane illustrating the relationship between TBox axiom count, number of threads, and reasoning time.

who also found through his survey study that automatic tuning ultimately outperforms manual tuning and becomes a crucial demand for optimizing parallel architectures [49].

We benchmarked our search algorithm, Parallelism Optimizer (PO), against the methods reviewed in Section II-C, namely: Grid search (GS), hill climb (HC), simulation annealing (SA), and random search (RS). We ensured a fair comparison by isolating the implementation of PO from other components in EF, similar to the isolation strategy employed in [40]. The results demonstrated the superiority of PO over its counterparts. PO exhibited logarithmic growth in search time and achieved a perfect success rate in identifying the optimal degree of parallelism. In contrast, other algorithms showed linear growth in search time and varying success rates. This superiority is gained from the deterministic characteristics of PO combined with the structural design of the Parallelism Tree (PT) search space. Specifically, the ordering nodes in PT based on a performance metric (i.e. the speedup factor for this study) and the queuing mechanism for storing parallelism configurations significantly decreased the search time and reduced tree size. Compared to the methodology proposed in [7], where a long chain of trees with a distinct node was used only for storage purposes, our methodology leveraged the efficiency of an integration between the tree-based and priority queue in storage and exploration purposes. However, our methodology provides efficiency for one-parameter optimization, and multi-dimensionality is not supported yet.

The analysis in Section V-D investigated the influence of increasing thread count and ontology size on ELK's scalability. For small ontologies, the results showed severe performance degradation as the number of threads increases. On the other hand, larger ontologies exhibited a lower speedup factor but maintained stable performance with the increase in thread

count. Our findings demonstrate that the ELK reasoner scales efficiently with larger ontologies using a high degree of parallelism, while for small ontologies it performs poorly due to over-utilization of processing units. These findings emphasize those in [16], where the authors stated that their ELK reasoner benefited more from increased parallelism when processing larger ontologies than smaller ones. In summary, this study highlights the need for adaptive tuning approaches to develop efficient and scalable reasoning systems.

This study effectively applied the intraclass correlation coefficient (ICC) method to analyze the reliability of EF. This effectiveness was the result of the following conditions. First, the sample size used in each ICC assessment were relatively large. Second, the massively parallel resources in HPC environment led to precise variance in the reasoning time measurements. Third, the experimental setup ensured exclusive access to HPC resources, leading to clean results, demonstrating the robustness of EF. These conditions enabled the ICC to effectively detect variance in time measurements both within and between parallelism configurations, contributing to a narrower confidence interval range. Based on these findings, we recommend future investigations to explore the viability of ICC in evaluating tuning results implemented under the same conditions.

This study offers significant contributions to revitalizing the domain of semantic reasoning and expanding existing research on optimization approaches. However, it was constrained by the limitations of the ELK reasoner, which operates only on a shared memory system. Furthermore, the hardware resources of the experimental environment restricted us from using massive-size ontologies. Future experimentation is required to validate the effectiveness of our methodology in optimizing different semantic reasoners on different computing architectures.

VII. CONCLUSION

As the size and complexity of ontologies expand, particularly with the advent of the Internet of Things and other data-driven systems, optimizing parallel semantic reasoners has become a significant challenge. This study proposed the Eagle Framework (EF), an innovative automatic tuning framework aimed at helping researchers optimize the performance of semantic reasoners. EF automatically generates thread configurations and effectively records performance data. EF differentiates itself through its modular design and adaptability, operating as a black-box solution that integrates seamlessly with various parallel reasoning engines. By designing a novel tree-based search algorithm, EF efficiently identifies the optimal number of threads. EF's methodology significantly reduces the manual effort required for tuning parallelism, saving researchers time and enabling them to focus on higher-level tasks. EF's ability lies in writing the performance measurements in CSV files, making them ready for data analysis. In addition, EF represents performance data in high-resolution visualization, offering researchers a comprehensive understanding of how different configurations impact reasoning efficiency.

Through a case study, this research validated the efficiency of EF in tuning thread configurations for the ELK reasoner across varied ontology sizes. Comparative analysis shows that

EF efficiently identifies optimal parallelism, outperforming existing search algorithms applied in optimization studies. Furthermore, this study validated the effectiveness of EF in addressing key research questions commonly discussed in the literature, such as the relationship between optimal performance and the full utilization of logical cores and the scalability of parallel reasoners to increase both processing resources and ontology size. In addition, this study introduced the application of the intraclass correlation coefficient (ICC) in assessing the reliability of performance tuning tools. The findings validated the consistency of the EF tuning measurements within and between configurations, suggesting the accuracy of ICC in assessing the reliability of tuning systems executed on a high-performance computing architecture.

REFERENCES

- [1] M. Lnenicka and J. Komarkova, "Big and open linked data analytics ecosystem: Theoretical background and essential elements," *Government Information Quarterly*, vol. 36, pp. 129–144, 1 2019.
- [2] M.-C. Valiente and J. Pavón, "Web3-dao: An ontology for decentralized autonomous organizations," *Journal of Web Semantics*, vol. 82, p. 100830, 10 2024. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1570826824000167>
- [3] C. Yang, Y. Zheng, X. Tu, R. Ala-Laurinaho, J. Autiosalo, O. Seppänen, and K. Tammi, "Ontology-based knowledge representation of industrial production workflow," *Advanced Engineering Informatics*, vol. 58, p. 102185, 10 2023.
- [4] P. Bonte, F. D. Turck, and F. Ongenaë, "Bridging the gap between expressivity and efficiency in stream reasoning: a structural caching approach for iot streams," *Knowledge and Information Systems*, vol. 64, pp. 1781–1815, 7 2022.
- [5] S. Arslan and O. Ünsal, "Efficient thread-to-core mapping alternatives for application-level redundant multithreading," *Concurrency and Computation: Practice and Experience*, vol. 35, 11 2023.
- [6] Z. Quan and V. Haarslev, "A parallel computing architecture for high-performance owl reasoning," *Parallel Computing*, vol. 83, pp. 34–46, 4 2019. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S016781911830142X>
- [7] A. Rasch, R. Schulze, M. Steuwer, and S. Gorlatch, "Efficient auto-tuning of parallel programs with interdependent tuning parameters via auto-tuning framework (atf)," *ACM Transactions on Architecture and Code Optimization*, vol. 18, pp. 1–26, 3 2021. [Online]. Available: <https://dl.acm.org/doi/10.1145/3427093>
- [8] M. Noura, M. Atiquzzaman, and M. Gaedke, "Interoperability in internet of things: Taxonomies and open challenges," *Mobile Networks and Applications*, vol. 24, pp. 796–809, 6 2019.
- [9] E. P. Cynthia, S. B. M. Samuri, W. S. Li, E. Ismanto, L. Afriyanti, and M. I. Arifandy, "Improved machine learning algorithm for heart disease prediction based on hyperparameter tuning," in *2023 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAJET)*. IEEE, 9 2023, pp. 176–181.
- [10] M. A. Ramadhani, Y. Azhar, and G. W. Wicaksono, "A study on the implementation of yolov4 algorithm with hyperparameter tuning for car detection in unmanned aerial vehicle images," in *2023 11th International Conference on Information and Communication Technology (ICoICT)*. IEEE, 8 2023, pp. 639–644.
- [11] G. Cheng, S. Ying, and B. Wang, "Tuning configuration of apache spark on public clouds by combining multi-objective optimization and performance prediction model," *Journal of Systems and Software*, vol. 180, p. 111028, 10 2021.
- [12] D. Nikitopoulou, D. Masouros, S. Xydis, and D. Soudris, "Performance analysis and auto-tuning for spark in-memory analytics," in *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2 2021, pp. 76–81.
- [13] J. J. Durillo, P. Gschwandtner, K. Kofler, and T. Fahringer, "Multi-objective region-aware optimization of parallel programs," *Parallel Computing*, vol. 83, pp. 3–21, 4 2019.
- [14] J. B. Fernandes, F. H. S. da Silva, T. Barros, I. A. Assis, and S. X. de Souza, "Patsma: Parameter auto-tuning for shared memory algorithms," *SoftwareX*, vol. 27, p. 101789, 9 2024.
- [15] A. N. Lam, B. Elvesaeter, and F. Martín-Recuerda, "A performance evaluation of owl 2 dl reasoners using ore 2015 and very large bio ontologies," in *DMKG2023: 1st International Workshop on Data Management for Knowledge Graphs*, vol. 3443. Technical University of Aachen, 5 2023, p. 13. [Online]. Available: <https://dmkg-workshop.github.io/papers/paper2861.pdf>
- [16] Y. Kazakov, M. Krötzsch, and F. Simančík, "The incredible elk," *Journal of Automated Reasoning*, vol. 53, pp. 1–61, 6 2014. [Online]. Available: <http://link.springer.com/10.1007/s10817-013-9296-3>
- [17] G. Santipantakis and G. A. Vouros, "Distributed reasoning with coupled ontologies: the e-shiq representation framework," *Knowledge and Information Systems*, vol. 45, no. 2, pp. 491–534, November 2015. [Online]. Available: <http://link.springer.com/10.1007/s10115-014-0807-2>
- [18] T. Wang, Y. Zhu, P. Ye, W. Gong, H. Lu, H. Mo, and F.-Y. Wang, "A new perspective for computational social systems: Fuzzy modeling and reasoning for social computing in cps," *IEEE Transactions on Computational Social Systems*, vol. 11, pp. 101–116, 2 2024.
- [19] Y.-B. Kang, S. Krishnaswamy, W. Sawangphol, L. Gao, and Y.-F. Li, "Understanding and improving ontology reasoning efficiency through learning and ranking," *Information Systems*, vol. 87, p. 101412, 1 2020. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0306437917306476>
- [20] S. Borgwardt and R. Peñaloza, "Algorithms for reasoning in very expressive description logics under infinitely valued gödel semantics," *International Journal of Approximate Reasoning*, vol. 83, pp. 60–101, 4 2017.
- [21] H. Herodotou, Y. Chen, and J. Lu, "A survey on automatic parameter tuning for big data processing systems," *ACM Computing Surveys*, vol. 53, pp. 1–37, 3 2021. [Online]. Available: <https://dl.acm.org/doi/10.1145/3381027>
- [22] B. van Werkhoven, "Kernel tuner: A search-optimizing gpu code autotuner," *Future Generation Computer Systems*, vol. 90, pp. 347–358, 1 2019. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0167739X18313359>
- [23] J. Schwarzrock, H. M. G. de A. Rocha, A. C. S. Beck, and A. F. Lorenzon, "Effective exploration of thread throttling and thread/page mapping on numa systems," in *2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems (HPCC/SmartCity/DSS)*. IEEE, 12 2020, pp. 239–246. [Online]. Available: <https://ieeexplore.ieee.org/document/9408014/>
- [24] Z. Fan, R. Sen, P. Koutris, and A. Albaghouthi, "Automated tuning of query degree of parallelism via machine learning," in *Proceedings of the Third International Workshop on Exploiting Artificial Intelligence Techniques for Data Management*. ACM, 6 2020, pp. 1–4. [Online]. Available: <https://dl.acm.org/doi/10.1145/3401071.3401656>
- [25] A. Vogel, D. Griebler, and L. G. Fernandes, "Providing high-level self-adaptive abstractions for stream parallelism on multicores," *Software: Practice and Experience*, vol. 51, pp. 1194–1217, 6 2021. [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1002/spe.2948>
- [26] T. Siddiqui, A. Jindal, S. Qiao, H. Patel, and W. Le, "Cost models for big data query processing: Learning, retrofitting, and our findings," in *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. ACM, 6 2020, pp. 99–113. [Online]. Available: <https://dl.acm.org/doi/10.1145/3318464.3380584>
- [27] Y. Liu, M. Li, N. K. Alham, and S. Hammoud, "Hsim: A mapreduce simulator in enabling cloud computing," *Future Generation Computer Systems*, vol. 29, pp. 300–308, 1 2013. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0167739X11000884>
- [28] R. Andonie, "Hyperparameter optimization in learning systems," *Journal of Membrane Computing*, vol. 1, pp. 279–291, 12 2019. [Online]. Available: <http://link.springer.com/10.1007/s41965-019-00023-0>
- [29] S. G. C. G and B. Sumathi, "Grid search tuning of hyperparameters in random forest classifier for customer feedback sentiment prediction," *International Journal of Advanced Computer Science and Applications*, vol. 11, 2020. [Online]. Available: <http://thesai.org/Publications/ViewPaper?Volume=11&Issue=9&Code=IJACSA&SerialNo=20>

- [30] I. Priyadarshini and C. Cotton, "A novel lstm-cnn-grid search-based deep neural network for sentiment analysis," *The Journal of Supercomputing*, vol. 77, pp. 13911–13932, 12 2021. [Online]. Available: <https://link.springer.com/10.1007/s11227-021-03838-w>
- [31] D. Chen, R. Zhang, and R. G. Qiu, "Noninvasive mapreduce performance tuning using multiple tuning methods on hadoop," *IEEE Systems Journal*, vol. 15, pp. 2906–2917, 6 2021. [Online]. Available: <https://ieeexplore.ieee.org/document/9205847/>
- [32] P. Sewal and H. Singh, "Algorithmic proficiency in spark configuration tuning: An empirical study using execution time metrics across varied workloads," *Procedia Computer Science*, vol. 235, pp. 2307–2317, 2024. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1877050924008950>
- [33] R. Sivakumar and H. Mangalam, "Ensemble hill climbing optimization in adaptive cruise control for safe automated vehicle transportation," *Journal of Supercomputing*, vol. 76, pp. 5780–5800, 8 2020.
- [34] J. Zeng, P. Romano, J. Barreto, L. Rodrigues, and S. Haridi, "Online tuning of parallelism degree in parallel nesting transactional memory," in *Proceedings - 2018 IEEE 32nd International Parallel and Distributed Processing Symposium, IPDPS 2018*. Institute of Electrical and Electronics Engineers Inc., 8 2018, pp. 474–483.
- [35] A. K. Pradhan, D. Mishra, K. Das, M. S. Obaidat, and M. Kumar, "A covid-19 x-ray image classification model based on an enhanced convolutional neural network and hill climbing algorithms," *Multimedia Tools and Applications*, vol. 82, pp. 14219–14237, 4 2023. [Online]. Available: <https://link.springer.com/10.1007/s11042-022-13826-8>
- [36] A. Kuznetsov, M. Karpinski, R. Ziubina, S. Kandy, E. Frontoni, O. Peliukh, O. Veselska, and R. Kozak, "Generation of nonlinear substitutions by simulated annealing algorithm," *Information (Switzerland)*, vol. 14, 5 2023.
- [37] A. Gülcü and Z. Kuş, "Multi-objective simulated annealing for hyper-parameter optimization in convolutional neural networks," *PeerJ Computer Science*, vol. 7, p. e338, 1 2021. [Online]. Available: <https://peerj.com/articles/cs-338>
- [38] M. Abdel-Basset, W. Ding, and D. El-Shahat, "A hybrid harris hawks optimization algorithm with simulated annealing for feature selection," *Artificial Intelligence Review*, vol. 54, pp. 593–637, 1 2021. [Online]. Available: <https://link.springer.com/10.1007/s10462-020-09860-3>
- [39] F.-J. Willemsen, R. Schoonhoven, J. Filipovič, J. O. Tørring, R. van Nieuwpoort, and B. van Werkhoven, "A methodology for comparing optimization algorithms for auto-tuning," *Future Generation Computer Systems*, vol. 159, pp. 489–504, 10 2024. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0167739X24002498>
- [40] K. Deligkaris, "Particle swarm optimization and random search for convolutional neural architecture search," *IEEE Access*, vol. 12, pp. 91229–91241, 2024. [Online]. Available: <https://ieeexplore.ieee.org/document/10577981/>
- [41] F. Hosseini, C. Prieto, and C. Álvarez, "Hyperparameter optimization of regional hydrological lstms by random search: A case study from basque country, spain," *Journal of Hydrology*, vol. 643, p. 132003, 11 2024. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0022169424013994>
- [42] O. Krestinskaya, M. E. Fouda, H. Benmeziane, K. E. Maghraoui, A. Sebastian, W. D. Lu, M. Lanza, H. Li, F. Kurdahi, S. A. Fahmy, A. Eltawil, and K. N. Salama, "Neural architecture search for in-memory computing-based deep learning accelerators," *Nature Reviews Electrical Engineering*, vol. 1, pp. 374–390, 5 2024. [Online]. Available: <https://www.nature.com/articles/s44287-024-00052-7>
- [43] C. Huang, Y. Li, and X. Yao, "A survey of automatic parameter tuning methods for metaheuristics," *IEEE Transactions on Evolutionary Computation*, vol. 24, pp. 201–216, 4 2020. [Online]. Available: <https://ieeexplore.ieee.org/document/8733017/>
- [44] C. C. Foster, "A generalization of avl trees," *Communications of the ACM*, vol. 16, pp. 513–517, 8 1973.
- [45] N. G. Bronson, J. Casper, H. Chafi, and K. Olukotun, "A practical concurrent binary search tree," *ACM SIGPLAN Notices*, vol. 45, pp. 257–268, 5 2010.
- [46] G. Antoniou, S. Batsakis, R. Mutharaju, J. Z. Pan, G. Qi, I. Tachmazidis, J. Urbani, and Z. Zhou, "A survey of large-scale reasoning on the web of data," *The Knowledge Engineering Review*, vol. 33, p. e21, 12 2018.
- [47] D. Liljequist, B. Elfving, and K. S. Roaldsen, "Intraclass correlation – a discussion and demonstration of basic features," *PLOS ONE*, vol. 14, p. e0219854, 7 2019.
- [48] H. Kim, C. M. Park, and J. M. Goo, "Test-retest reproducibility of a deep learning-based automatic detection algorithm for the chest radiograph," *European Radiology*, vol. 30, pp. 2346–2355, 4 2020.
- [49] D. Mustafa, "A survey of performance tuning techniques and tools for parallel applications," *IEEE Access*, vol. 10, pp. 15036–15055, 2022. [Online]. Available: <https://ieeexplore.ieee.org/document/9698048/>

Imbalance Datasets in Malware Detection: A Review of Current Solutions and Future Directions

Hussain Almajed, Abdulrahman Alsaqer, Mounir Frikha

Department of Computer Networks and Communications, College of Computer Sciences and Information Technology
King Faisal University, Al-Ahsa, 31982, Saudi Arabia

Abstract—Imbalanced datasets are a significant challenge in the field of malware detection. The uneven distribution of malware and benign samples is a challenge for modern machine learning based detection systems, as it creates biased models and poor detection rates for malicious software. This paper provides a systematic review of existing approaches for dealing with imbalanced datasets in malware detection such as data-level, algorithm-level, and ensemble methods. We explore different techniques including Synthetic Minority Oversampling Technique, deep learning techniques including CNN and LSTM hybrids, Genetic Programming for feature selection, and Federated Learning. Furthermore, we assess the strengths, weakness, and areas of application of each approach. Computational complexity, scalability, and the practical applicability of these techniques remains as challenges. Finally, the paper summarizes promising directions for future research like lightweight models and advanced sampling strategies to further improve the robustness and practicality of malware detection systems in dynamic environments.

Keywords—Malware detection; machine learning; imbalance datasets; oversampling; SMOTE

I. INTRODUCTION

Cybersecurity is a critical area in today's world and malware detection is a critical area of cybersecurity, because malicious software is proliferating at a rapid rate, and it is getting more sophisticated [1], [2]. Moreover, Malware detection solutions are essential given the urgent need to solve the issue. However, detecting malware more effectively has become increasingly difficult because of the complexity of modern malware and the volume of data being generated. In response to this challenge, Machine Learning (ML) techniques have risen in prominence by learning malware patterns and determining their difference from benign software [3]. But the problem of imbalanced datasets is a major obstacle in developing effective malware detection systems. This comes from having a dataset used to train ML models that contain a much lower number of malware samples than data associated with benign samples, leading to biased models that cannot adequately detect malicious activity.

In this paper, we present techniques for addressing imbalanced datasets in malware detection and evaluate their effectiveness through a systematic review.

II. BACKGROUND

A. Imbalanced Datasets in Malware Detection

Malware detection datasets are imbalanced when the malware samples (minority class) have a very small distribution compared to benign samples (majority class) [4], [3], [5].

As a result, model predictions become skewed towards the majority class and ignore important minority samples, which are most often the focus in cybersecurity. Fig. 1 demonstrates the imbalanced data distribution.

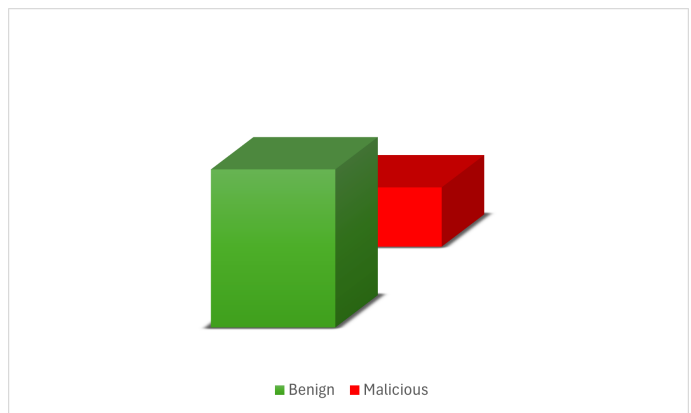


Fig. 1. Visual representation of an imbalanced malware dataset.

B. Balanced Datasets in Malware Detection

Malware detection datasets that are balanced are those which have approximately the same number of samples in the majority class (non-malicious data) and minority class (malicious data) [4], [2]. This balance prevents classifiers from biasing towards any one class, and results in more accurate detection of both non-malicious and malicious data. Fig. 2 demonstrates the balanced data distribution where both data are equal in the count.

C. Challenges of Imbalanced Datasets

Imbalanced datasets raise the following challenges:

- **Biased Prediction:** Datasets with imbalanced classes, therefore, often lead to classifiers that are skewed towards the majority class, and would often then perform poorly on the minority class [4], [2].
- **Poor Generalization:** Insufficient training examples lead to failure of the classifiers to generalize well on minority class predictions [3].
- **Metric Misleading:** As high accuracy can be obtained by ignoring the minority class, standard accuracy measures become unreliable [2], [6].
- **Class Overlapping:** Classes of imbalanced datasets might overlap, and there will be no clear boundaries

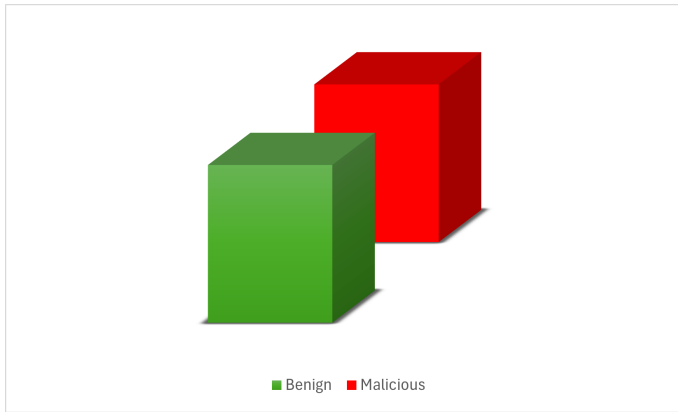


Fig. 2. Visual representation of balanced malware datasets.

TABLE I. OVERVIEW OF DATA-LEVEL METHODS

Method	Definition	Advantages	Limitations
Over-Sampling [4]	It add synthetic examples to the minority class to balance the dataset.	Balances class distribution without losing existing data.	Risk of overfitting and computational overhead in managing large synthetic datasets [6].
Under-Sampling [6]	Reduces the majority class samples to balance the dataset by either randomly removing examples or applying heuristic methods.	Simplifies the dataset, and encourages the model to focus equally on both classes.	Loss of potentially valuable data from the majority class.

separating classes, which complicates distinguishing between majority and minority samples [6], [2].

- Overfitting and Underfitting: On the other hand, over-sampling the minority class results to overfit while under-sampling the majority class results to underfit [6].

D. Approaches to Address Imbalanced Datasets

In this section, different techniques are introduced to address the imbalanced dataset problem in malware detection. We broadly categorize these approaches into data level methods, algorithm level methods, and ensemble methods that solve the imbalance problem from different angles.

1) *Data-Level Methods*: Data level approaches try to balance the class distribution by changing the data before applying any ML algorithm [6]. Table I shows the data-level method.

The Fig. 3 shows the illustration of over-sampling and under-sampling.

2) *Algorithm-Level Methods*: Algorithm level methods adapt existing learning algorithms to make them more sensitive to imbalanced data [7]. Unlike these methods, they do not change the dataset but rather change the training process. Common algorithm-level methods show in Table II.

3) *Ensemble Methods*: It is a combination of multiple classification techniques from the above mentioned categories and can be seen as a wrapper of other methods such as nsembling which is widely used as a classification technique [7].The method consists of pretraining and fine tuning on the original



Fig. 3. Illustration of OverSampling and UnderSampling methods for handling imbalance datasets.

TABLE II. SUMMARY OF ALGORITHM-LEVEL METHODS

Method	Definition	Advantages	Limitations
Cost-Sensitive Learning [7]	Assigns higher mis-classification costs to minority classes.	Improves focus on minority samples, and aligns learning with real-world impact.	Requires precise cost estimation; may still bias towards majority class.
Thresholding [7], [2]	To balance the class distribution, the decision threshold is adjusted.	Simple Implementation, No data loss.	heavily depends on the choice of the optimal threshold value and not be effective for all types.
One class classification [7], [2]	It learns from one class (typically the minority class) and seeks to identify instances that belong to this class, rejecting all others.	Useful for high-dimensional datasets and more robust to noisy data.	More complex to implement and not generalize well to new, unseen data.

imbalanced dataset. Also, it combines predictions from multiple models to increase robustness and decrease bias [3], [4]. Bagging and Boosting are techniques. where it Combines the strengths of individual classifiers for better overall performance and reduces the impact of minority class under representation by focusing on difficult to classify samples [6], [4].

E. Motivation

This systematic literature review is motivated by the necessity of improving malware detection capabilities in the presence of:

- The Growing Threat of Malware: Malware attacks have been increasing in frequency and sophistication, making risks to individuals and organizations. As stated by the report of AV-atlas, where over three millions new malware were found in the first two weeks of November 2024 [8]
- Importance of Effective Malware Detection: Undetected malware can lead to the loss of sensitive information, financial implications, operational disruption.
- Challenges with Imbalanced Datasets: Non-malicious samples outnumber malware samples, leading to model bias and high false negatives.
- Need to Address Data Imbalance: To enhance security, improve malware detection accuracy and strengthens overall cybersecurity defenses.

F. Problem Statement

Imbalanced datasets in malware detection pose a big problem for ML models, which leads to biased detection systems that fail to well detect malware [6], [5]. Failure in the identification of the minority class leads to models that perform poorly when it comes to classifying benign against malicious samples, this being due to the current imbalance between benign and malicious samples. This work is a systematic literature review to investigate and assess existing solutions to solve this problem, and to gain insights to develop better methods to deal with imbalanced datasets in malware detection.

G. Scope

The scope of this SLR is to review the literature on imbalance in datasets for malware detection. It includes data level, algorithm level and ensemble methods used to handle the imbalanced datasets. The scope is to evaluate these methods, to identify the challenges and limitations of applying them, and to suggest potential directions for future research. In addition, the review will point out how different solutions have been used in the case of malware detection and their pros and cons.

H. Objective

The objectives of this research are as follows:

- Conduct a Comprehensive Literature Review: In order to systematically review the existing literature regarding how to handle imbalanced datasets in malware detection.
- Investigate Current Solutions: In order to identify and evaluate different approaches used to tackle imbalanced datasets.
- Assess Effectiveness: Focusing on metrics such as accuracy and F1-score, these approaches will further be assessed for their effectiveness in improving malware detection.
- Identify Challenges and Gaps: The challenges, limitations, and gaps of existing methods dealing with imbalanced datasets in malware detection will be identified.
- Suggest Future Directions: propose several directions that could become future research paths in regard to imbalanced datasets in malware detection.

By addressing these objectives, this review aims to offer a clear understanding of the current landscape of imbalanced dataset in malware detection.

III. RESEARCH METHODOLOGY

We follow a systematic approach to review the existing literature on imbalance datasets in malware detection, following the Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA) guidelines. It includes defining the research questions, selecting databases, developing search strings, establishing of inclusion exclusion criteria, and applying a quality assessment framework. The methodology is organized as follows:

A. Data Sources and Search Strategy

To ensure comprehensive coverage of relevant studies, the search was conducted across the following academic databases:

- IEEE Xplore
- MDPI
- SpringerLink
- ScienceDirect

The keywords used for the selection based on the related research objectives:

(“Imbalanced Datasets”) AND (“Malware Detection”)

Only peer-reviewed journal articles and conference papers published between 2020 and 2024 were considered to capture recent developments.

B. Inclusion and Exclusion Criteria

To filter search results for relevant studies, we established the following inclusion and exclusion criteria:

1) Inclusion Criteria:

- Studies that focus on imbalanced datasets and malware detection
- Peer-reviewed journal articles, conference papers.
- Studies that provide empirical results or evaluations using datasets relevant to imbalanced datasets.
- Publications written in English.
- Propose novel methods or provide empirical evaluations.

2) Exclusion Criteria:

- Studies not related to imbalanced datasets and malware detection.
- Publications that only provide theoretical models without empirical validation.
- Non-peer-reviewed sources such as theses, white papers, and editorials.

C. Study Selection Process

The study selection process adhered to the PRISMA framework, proceeding in three stages:

- Initial Screening: All retrieved articles were screened by titles and abstracts to exclude irrelevant studies and choose those meeting the inclusion criteria for full-text review.
- Full-Text Review: Full texts of selected articles were reviewed to determine their relevance and quality. Excluded articles that did not provide detailed information on balancing techniques, datasets, or empirical evaluations.
- Data Extraction and Coding: A standardized form was used to extract data from the final set of articles, including balancing techniques, datasets, and

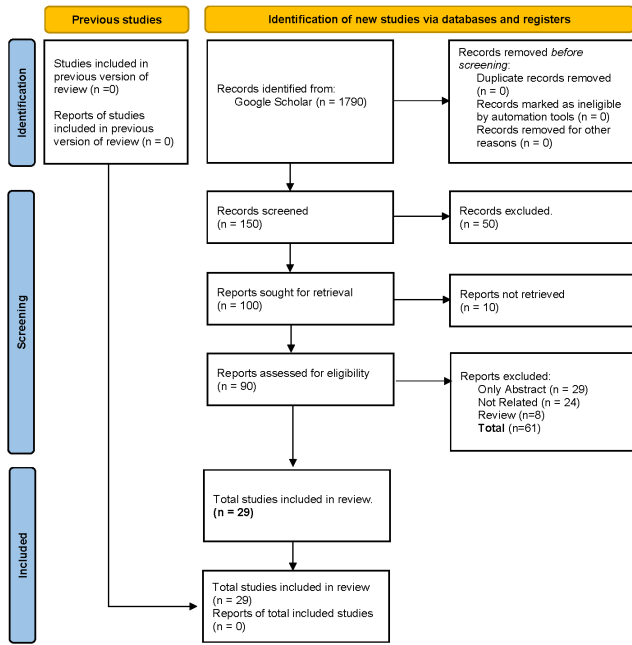


Fig. 4. PRISMA flow diagram summarizing the study selection process.

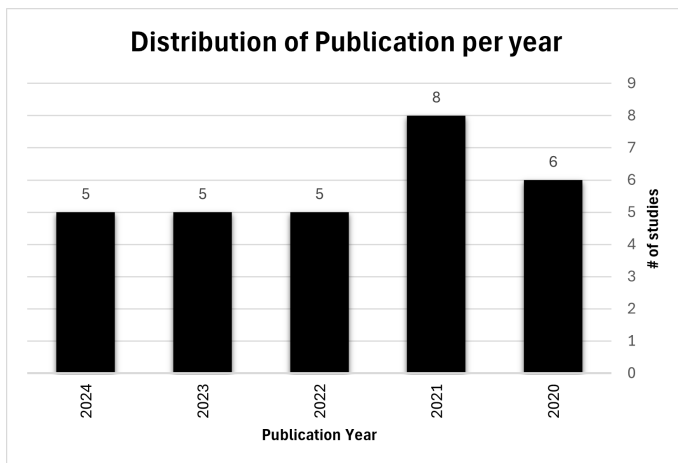


Fig. 5. Distribution of publications included in the review based in year.

evaluation metrics, as well as identify challenges and contributions.

Fig. 4 shows the PRISMA flow diagram.

Fig. 5 shows the distribution of the number of papers selected for this SLR per year.

IV. LITERATURE REVIEW

The problem of imbalanced datasets in malware detection in Android devices is addressed by Dehkordy and Rasoolzadegan [9]. They obtained a dataset from Drebin and AMD datasets containing 9,223 applications, and was heavily pre-processed to reduce the features from 1,262 to 78 for faster learning. The authors used SMOTE (Synthetic Minority Over-sampling Technique), undersampling, and a hybrid approach to

solve the imbalanced issue. To improve the accuracy of detection they employed dataset preprocessing, ranking of features and using multiple classifiers like K-nearest neighbors (KNN), Support Vector Machines (SVM) and Iterative Dichotomiser 3 (ID3). The best results were obtained by a combination of KNN with SMOTE, with an accuracy of 98.69%. However, the study limited to false positive rates of 2.09% to 4.77% and an approach that is only applicable to a limited number of malware families, which limits the model's generalizability. Guan et al. [10] propose n Class Imbalance Learning (CIL) approach to address the class imbalance problem for Android malware detection. It applies the K-means clustering-based under-sampling, which retains the representative majority samples, and then the SMOTE algorithm to generate the synthetic minority samples. A Random Forest (RF) classifier is then trained on the combined dataset. The dataset used for evaluation consists of 10,182 malware samples from VirusShare and 127 benign samples, with a class imbalance ratio of 1:80. They showed that the CIL method outperforms other traditional methods such as SMOTE and random under-sampling. In general, CIL shows good generalizability to other imbalanced datasets, and it is a promising solution to the class imbalance problem in malware detection.

Imbalanced datasets in malware detection for edge computing in Android based Internet of Things (IoT) environments is addressed by Khoda et al. [11]. The authors propose two methods a dynamic class weighting technique and modified Fuzzy-SMOTE for synthetic oversampling. The first approach generates valid synthetic malware samples preserving the malicious functionality, the second approach dynamically adjusts class weights during training to improve malware detection. The evaluation show over 9% improvement in F1 score over traditional imbalanced learning techniques. 50,000 Android applications and 500 malware samples in the dataset. The fuzzy approach is limited by the requirement of careful tuning, while the dynamic class weighting method is less sensitive to such parameters.

The challenge of detecting Android ransomware in an imbalanced dataset is addressed by ALMOMANI et al. [12]. A hybrid evolutionary approach using Binary Particle Swarm Optimization (BPSO) and SVM is employed for feature selection and classification to improve classification performance by effective optimization. The SMOTE was used to balance the dataset. Sensitivity, specificity and g-mean were used to evaluate the model, scoring 96.4%, 98.7% and 97.5%, respectively. The dataset has 10,153 Android applications out of 500 ransoms. However, the dataset is small making it difficult to generalize, especially for new ransomware variants.

Hemalatha et al. [13] suggest a DenseNet model based on Deep Learning (DL), with a class balanced categorical cross entropy loss to overcome class imbalances. Malware binaries are transformed into grayscale images and malware detection is framed as a multi-class image classification problem. The experiments were performed on Maling, BIG 2015, and MaleVis datasets with high accuracies of 98.23%, 98.46%, and 98.21%, respectively; and 89.48% on the unseen Malicia dataset. However, the model lacks in accuracy on unseen data, and struggles with novel malware (zero day attacks). Future work could involve improving generalization to deal with zero day attacks.

Goyal and Kumar [14] discuss malware detection and the effect of data imbalance. To balance the dataset, the researchers use random under-sampling to reduce 42,797 malware samples to 1,079 benign samples. They compared different ML classifiers (KNN, Decision Tree, RF). RF achieved the best accuracy of 98.94% on the imbalanced dataset, and 90.38% on the balanced dataset. They show the impact of data imbalance on model accuracy, and that more reliable results can be obtained from balanced datasets. The study concludes that balanced datasets are necessary to reduce bias and increase reliability, and future research could include further investigation of more sophisticated balancing techniques to improve the applicability of the model to real world scenarios.

Salas and Geus [15] addresses the challenge of class imbalance. The authors propose the MobileNet Fine-Tuning (MobileNet FT) model, a fine tuned version of MobileNet that utilizes bicubic interpolation and class weight estimation techniques. Experimental results show that the proposed model reaches accuracy rates for different datasets, such as Microsoft Big 2015 (98.71%), Maling (99.08%), MaleVis (96.04%), and a new Fusion dataset (98.04%). These results show that the model is robust to a range of malware families. The approach is also shown to have limitations, such as a degradation in performance as the number of malware families increases and problems with unseen malware. This motivates further investigation into more adaptive models to improve scalability and robustness to new threats.

Almaleh et al. [16] suggest to improve the detection of malware in Windows using a hybrid method. They use logistic regression with Recurrent Neural Network (RNN) to detect malware from Application programming interfaces (API) call sequences. The study presents a solution to the problem of imbalanced datasets through the use of an undersampling technique that creates a balanced dataset of 2,158 samples of malicious and non-malicious samples. They initialize the RNN weights using logistic regression for improved model accuracy. For the balanced dataset, the model reached an accuracy of 83%, and for the imbalanced, an accuracy of 98%. Limitations include a relatively small trained dataset after balancing due to its restriction on generalizability. Future work could build on the model for other operating systems and overcome these limitations so that the model is more applicable and robust in different scenarios.

Yu Ding et al. [17] proposed self-attention based approach, considering malware ASM files as text sequences to distinguish the malware families. The imbalance dataset technique used to represent ASM files as integer vectors and use a self attention neural network to improve minority class recognition. The sequence classification accuracy is improved by capturing internal dependencies within sequences using this approach. The model is evaluated using the Microsoft Malware Classification Challenge dataset, and shows a robustness to different datasets with 98.48% accuracy and 89.66% F1 score for Simda class. However, small sample recognition problems are not completely solved, and the interpretability of the model is restricted. The future work could include improving early detection of new malware and make neural networks easier to interpret for practical application in cybersecurity.

Moti et al. [18] handles the problem of imbalanced dataset for malware detection. The synthetic samples for minority

classes are generated using a hybrid model composed of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks with Sequence Generative Adversarial Networks (SeqGAN), so that the dataset is balanced. The classification accuracy is improved to 98.99% using this approach. They evaluate on a Microsoft dataset from a Kaggle competition that contains nine malware families. While the high accuracy, the model depends only on opcode sequences without preprocessing, which can restrict its feature diversity. Moreover, the training overhead of SeqGAN is not high and the model still needs to be adapted to different datasets and zero day threats.

The problem of Android malware detection is addressed by Almomani et al. [19]. They present a vision based DL model that converts Android Application Package (APK) bytecodes to visual images and uses CNNs for classification. They evaluate the model on an imbalanced dataset (14,733 malware and 2,486 benign samples), without using data augmentation. The main contribution is the development of 16 fine tuned CNN algorithms that are efficiently able to classify malware, which demonstrates 99.40% accuracy on balanced datasets and 98.05% on imbalanced datasets. The study highlighted reduced computational cost due to no longer requiring the manual feature extraction. The limitations include dependence on pre trained CNN weights and uncertain adaptability to new malware types or other datasets.

In the problem of detecting macro malware in Microsoft document files, Mimura [20] tackles the problem of highly imbalanced datasets. They propose a method to combine Doc2Vec and Latent Semantic Indexing (LSI) with four classifiers (SVM, RF, Multilayer Perceptron (MLP), CNN) to increase the accuracy of malware detection. The highest F-measure of 0.99 indicated a high accuracy. The dataset consisted of more than 30,000 samples from VirusTotal and Stack Overflow with an imbalanced distribution favoring benign samples. Limitations include possible lack of generalizability from dataset composition, and future work is to collect more data for robust evaluation. The results of the study demonstrate that LSI is robust to class imbalance and promising results in practical malware detection applications.

Nikale and Purohit [21] addresses the issue of class imbalance in the dataset. The authors used dynamic features such as system calls and binder calls to classify Android APKs into five families: Ransomware, smware, adware, scareware and benign. The dataset consists of 525 APK samples from different sources, including Contagio Mobile and Google Play Store. The research introduced SMOTE, Adaptive Synthetic sampling (ADASYN), and balanced cost. They tested various classifiers, and the highest accuracy of 91% was obtained with Extreme Gradient Boosting (XGBoost) combined with SMOTE. However, the study is limited by small dataset, hence, restricted in generalizing its findings. Furthermore, constraint is noted on the focus on a fixed set of dynamic features without investigating more specific behavioral characteristics.

SAWADOGO et al. [22] evaluate the impact of data imbalance on eleven ML algorithms, the authors use CICMal-Droid 2020, a malware dataset. For comparison, they created two subsets: one imbalanced and the other balanced. They claim that traditional evaluation metrics (Accuracy, Precision and Recall) are inappropriate for imbalanced datasets, while

Balanced Accuracy and Geometric Mean are more appropriate. The results show that algorithms such as AdaBoost and SVM perform very poorly on imbalanced data, whereas Extra Trees and RF are less sensitive. Therefore the authors suggest to use balanced evaluation metrics to better represent the model performance on imbalanced dataset. A study limitation is that a single dataset was used. Additionally, future work should study more sophisticated learning techniques on more diverse datasets to improve the robustness of Android malware detection models.

In this paper, Haluška et al. [23] compare 16 data preprocessing methods for imbalanced classification problems, with a focus on cybersecurity. The authors extensively used six cybersecurity datasets and 17 other public imbalanced datasets from different domains as benchmarks. Overall, the performance of oversampling methods is better than that of undersampling methods, and the standard SMOTE algorithm gives a substantial performance boost. Experiment results indicate that SMOTE and its variants, e.g. generalization of SMOTE and SVM SMOTE, work well in various datasets and metrics, e.g. PR AUC, ROC AUC, and P-ROC AUC. This study shows that to improve predictive performance on multiple tasks in cybersecurity and other domains, it is essential to choose appropriate preprocessing methods and carefully consider method choice and hyperparameter tuning.

Alzammam et al. [24] provide a comparative view of different approaches to the problem of imbalanced multi-class classification in malware detection using CNN. The study focuses on evaluating the effectiveness of different methods, such as cost-sensitive learning, oversampling, and cross-validation, to mitigate the imbalance issue in three publicly available malware datasets. For instance, the study shows that oversampling outperforms other methods in boosting the accuracy and F1-score of the CNN model on every dataset, while the proposed model achieves substantial improvements in accuracy and F1-score when oversampling, with accuracy reaching as high as 99.94% for the Maling dataset. Finally, this research highlights the importance of dataset characteristics when selecting a method to correct for imbalance, as well as other data factors (including noise and overlapping) and the complexity of applying pre-trained models to malware classification.

Phung and Mimura [25] suggest a way to detect malicious JavaScript by using ML, but specifically dealing with the class imbalance problem. Once the balance between the benign and malicious datasets is adjusted through an oversampling technique, the authors use them to train a classifier for prediction. Experimental results indicate that the proposed method can effectively detect new malicious JavaScript with higher accuracy and efficiency (0.72 recall with Doc2Vec). With the same training and test time per sample, this outperforms the baseline method by 210% in terms of recall score on the dataset with over 30,000 samples: 21,745 benign samples from popular websites and 214 malicious samples from PhishTank, along with an additional 8000 malicious samples from GitHub. The research limits itself to various resampling techniques without exploring or comparing them in a more comprehensive way.

In mobile malware detection, Khoda et al. [26] propose a novel way to handle the problem of imbalanced datasets

via synthetic oversampling. This method proposes the addition of features to existing malware samples to generate synthetic malware samples that are valid and retain the malicious functionality. They test the approach using a Deep Neural Network (DNN) on the Drebin Android malware dataset. Results indicate that the proposed method achieves higher precision, recall and F1 score than the oversampling and undersampling techniques in general, and especially at lower imbalance ratios. The performance of the proposed model is much more accurate than previous methods, achieving an F1 score of 94.2% at 10% imbalance ratio and accuracy of 98.8%. The dataset used is the Drebin dataset which has 50,000 apps and 500 malicious apps as the minority class.

Reshi and Singh suggest a new method to handle the imbalance issue in malware datasets through the use of Variational Autoencoder (VAE) [27]. The proposed solution uses VAEs to extract and compress features from the given data and, thus, the model is able to learn features that are resistant to noise and distinguish between real malware and other types of noise. In addition, VAEs improve the data augmentation technique by generating synthetic malware samples from the learned latent space to help overcome the imbalanced problem. This approach expands the training set and, therefore, improves the model's ability to generalize and increase the detection rate for new or less frequent variants of malware. The research contributes by enhancing the VAEs by combining them with CNNs for malware detection. The proposed model gives an accuracy of 98% on the Maling dataset, which is better than the baseline model. The dataset used in this work is the Maling dataset, which has a highly unbalanced class distribution. The main drawback of the proposed work is that the integrated VAE-CNN model is relatively complicated and may need appropriate resource allocation and hyperparameter optimization.

Faridun and Im propose a novel malware detection approach using the TabNetClassifier, which is a DL architecture designed explicitly for tabular data analysis [28]. In this research, they enhance malware detection by utilizing the TabNetClassifier in conjunction with the SMOTE to address class imbalance in datasets. Initially, the dataset of 138,047 Portable Executable (PE) header samples is trained using the TabNetClassifier, which is imbalanced with 41,323 benign and 96,724 malware samples. SMOTE is applied to balance the dataset and improves model performance significantly. The main contribution of this research is to show the success of combining TabNetClassifier with SMOTE in improving malware detection accuracy and sensitivity. After applying SMOTE, the model achieves an accuracy of 99.10%, precision of 99.03%, and recall of 99.19%.

Li et al. [29] present a novel method for malicious family classification based on multimodal fusion and weight self-learning. The method deals with the problem of imbalanced datasets and concept drift in malware family classification. This approach integrates multiple features of byte, format, statistic, and semantic types to improve the robustness of the classification model. Experimental results show high efficiency and small resource overhead in classifying highly imbalanced malware family datasets while delivering very good classification performance. The dataset used consists of some types of malware, namely ransomware, Trojans, viruses, and malicious

mining programs. However, the research is limited by the reliance on static analysis and may not find the dynamic behaviors of malware.

In order to improve ransomware detection and classification, Onwuegbuche et al. [30] propose a three-stage feature selection method. This method applies chi-square (CHI2), Duplicated Features (DUF), and Constant Features (COF) filter feature selection techniques to reduce the dimensionality of the dataset, taking into account the different importance of different feature groups. Further, the study addresses the class imbalance problem by employing the SMOTE and cost-sensitive ML methods. The performance of this method is evaluated on the Elderan ransomware dataset and several ML models (XGBoost, Logistic Regression, RF, Decision Trees, and SVM). The results demonstrate that the proposed feature selection method leads to a 10% average improvement in binary classification and 21.79% in multi-class classification over previous studies. Among binary classifiers, XGBoost with cost-sensitive learning and SMOTE is the best with 98.78% balanced accuracy, while the best multi-class classifier is the RF model with cost-sensitive learning achieving 61.94% balanced accuracy.

Andelic et al. [31] deal with the problem of malware detection in imbalanced datasets. The authors suggest combining a Genetic Programming Symbolic Classifier (GPSC) with dataset oversampling techniques to increase the detection accuracy. They apply the GPSC algorithm to an open dataset containing hybrid features consisting of binary hexadecimal and Dynamic Link Library (DLL) calls of Windows executables. The dataset is initially imbalanced, containing 301 malicious and 72 non-malicious samples. In order to address this imbalance, the authors use oversampling techniques such as ADASYN, BorderlineSMOTE, KMeansSMOTE, SMOTE, and SVMSMOTE, and they train the GPSC with Five-Fold Cross-Validation(5FCV) and Random Hyperparameter Value Search (RHVS) method to select the best combination of hyperparameters. The classification accuracy of the proposed method is 0.9962. GPSC is used to generate Symbolic Expressions (SEs) that can be easily applied to and implemented into malware detection models, overcoming the limitations of traditional ML models, which are hard to interpret and transform into mathematical equations.

According to Çayır et al. [32], a new ensemble model called the Random CapsNet Forest (RCNF) is proposed to tackle the imbalance in malware type classification. The authors use the Capsule Network (CapsNet) architecture to preserve spatial information without the use of pooling layers and incorporate the bootstrap aggregating (bagging) technique to form an ensemble model. The idea is to reduce the variance of CapsNet models and improve the robustness of classification by using this approach, which is tested on two highly imbalanced datasets, Malimg and BIG2015, where the RCNF model is also shown to outperform other competitors with fewer trainable parameters. It achieves an F-Score of 98.20% for the BIG2015 dataset and 96.61% for the Malimg dataset. Advantages noted regarding the simplicity of the architecture and the ability to train from scratch without the need for transfer learning.

LIN et al. [33] present a ML framework based on a VAE and a MLP that helps overcome the problem of imbalanced

datasets in intrusion detection systems (IDS). An efficient range-based sequential search algorithm is included in the framework to determine the optimal sequence length for data segmentation from multiple sources, including network packets and system logs. Experimental results on HDFS dataset demonstrate that the proposed method achieves an F1 score of around 97% and recall rate of 98%, better than other solutions. Imbalanced datasets are treated using the proposed approach, which increases the F1-score by up to 35% and the recall rate by 27%. In addition, the work also points out the necessity of the appropriate data segmenting and the possibility of the proposed model detecting the new attack variants. The dataset used is the HDFS dataset, a public system log dataset.

In the context of Federation Learning (FL), ransomware detection, and attribution, Vehabovic et al. [34] address an essential problem of dataset imbalance. The authors suggest a modification of the FL scheme where the weighted cross entropy loss function is used to combat bias in datasets distributed across various clients. This approach is particularly applicable since ransomware data distribution and quantity can differ significantly also across different locations and companies. The performance of the proposed FL scheme is evaluated using an up-to-date repository of Windows-based ransomware families and benign applications. The results indicate that the weighted cross entropy loss function approach can mitigate the effect of dataset imbalance, especially in the case of binary ransomware detection with an average accuracy of 94.67%, but the study also points out the difficulties of multi-class attribution with imbalanced datasets, which results in more decline in performance compared to the balanced datasets.

As a form of unsupervised learning, Shi et al. [35] explore using One-Class Classification (OCC) to detect malware in the Internet of Things (IoT) domain. To combat dimensionality and information loss, the authors suggest that categorical features should be changed into numerical formats by using the Term Frequency-Inverse Document Frequency (TF-IDF) method. They compare the performance of OCC models, such as Isolation Forest and deep autoencoder, trained on benign NetFlow samples alone. It is shown that these models achieve 100% recall with precision rates greater than 80% and 90% on a number of test datasets, highlighting the adaptability of unsupervised learning to time-evolving malware threats in IoT, and making an important contribution to the study of malware detection in IoT, particularly when labeled malicious data are scarce. TF-IDF is used for feature transformation, and the comparison of various OCC algorithms leads to valuable improvements in the IoT security framework.

Using ML techniques, in particular, the use of genetic programming for feature selection, Al-Harashsheh et al. [36] propose how to enhance malware detection. The researchers built a malware detection model in which the features are selected by a genetic programming algorithm, and then a set of parallel classifiers is used to enhance detection accuracy at a lower cost. The proposed model employs five feature selection methods: Filter-based, wrapper-based, Chi-Square, Genetic Programming Mean (GPM), and Genetic Programming Mean Plus (GPMP). Experimental results demonstrate that the GPMP method (which uses fewer features than the Filter-based method) results in better accuracy and F1-score values of 0.881066 and 0.867546, respectively. The research

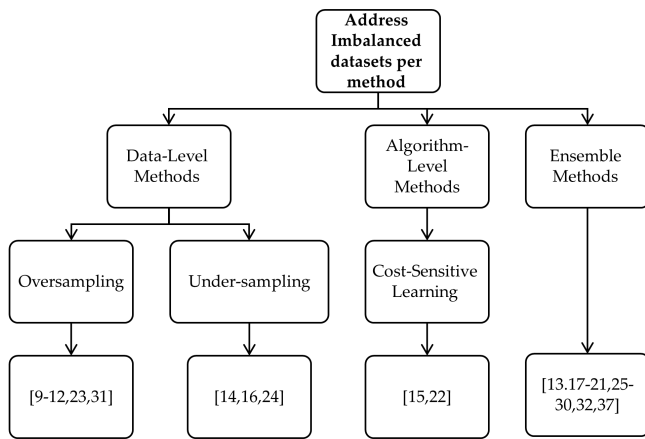


Fig. 6. Taxonomy of the literature review per method used.

indicates that genetic programming is able to select features to improve the performance of malware detection effectively. A number of classifiers, such as RF, Random Tree, and SVM, were used to compare the performance of the proposed feature selection methods.

Al-Khshali et al. [37] propose a new technique employing subspace learning-based OCC methods to detect malware. In this work, they address the issues of class imbalance and the curse of dimensionality in the application of traditional ML algorithms for malware detection. In order to overcome these problems, the researchers introduce a pipeline that uses subspace learning techniques such as Subspace Support Vector Data Description (SSVDD) and Graph Embedded Subspace Support Vector Data Description (GESSVDD). The proposed framework solves multiple problems at once, including class imbalance and the curse of dimensionality. The results show promising performance, with a True Positive Rate (TPR) of 100% for subspace-learning-based OCC. The datasets used in this study include (Benign and Malicious PE Files, ClaMP, and Malware Analysis Datasets by Oliveira) which are diverse but representative of a wide variety of malware types.

Table III shows a summary of the literature review conducted previously.

Table IV shows the limitations and contributions of studies conducted.

A. Taxonomy of the Research

Fig. 6 shows the taxonomy clearly categorizes the different techniques utilized by the studies to mitigate the imbalances, with the data level set of methods focusing on manipulating the dataset distribution, the algorithm level which trains towards the adaption of the learning algorithms and the ensemble where a combination of a variety of methods is used to get better results.

V. DISCUSSION OF THE LITERATURE REVIEW

The literature review reveals that many ML approaches applied to address imbalanced datasets in malware detection. Many studies indicate handling class imbalance is important to improve detection performance.

1) Oversampling Techniques:

- In several studies SMOTE was commonly employed to increase the minority class representation, leading to improvements in detection rates for families of malware like adware, ransomware or smsware [9] [21] [30]. In the case of imbalanced learning, SMOTE was found to be effective at improving metrics such as F-measure and MCC [9]. Complex oversampling methods gave incremental improvements, indicating the need to balance computational cost with performance gains [23].
- However, oversampling methods like the conventional SMOTE have some limitations that need to be overcome or minimized through the use of more advanced forms like the Modified Fuzzy-SMOTE whose oversampling strategies produces synthetic data that is more reflective of real data distribution. However, these techniques are still inefficient with large, high-dimensional data sets when they are applied.

2) Feature Selection Techniques:

- The improvement in model efficiency and accuracy was greatly aided by feature selection techniques. The genetic programming based feature selection methods, including GPMP, demonstrated that selecting fewer but more relevant features can reduce computational complexity and improve classifier performance [36]. Multi stage feature selection was used in other studies to select features such as API calls, registry operations, and directory logs, which improves model interpretability and classification performance [30]. Moreover, swarm intelligence based optimization, and in particular BPSO was also efficient to choose the best features in order to achieve a large performance gain when dealing with highly imbalanced data for Android malware detection [12].
- The problem of selecting features often demands an expert's input in the process for feature selection. Perhaps, even more, automated approaches, such as feature selection by applying AI methods, could be more beneficial for this step.

3) ML Approaches:

- RF and SVM are used frequently as they are robust, and can handle non linear relationships in data. For instance, Guan et al.[10] achieved significant accuracy improvements by combining RF with SMOTE, especially in datasets with a high imbalance ratio (1:80). In another study, in a custom malware detection dataset, the RF model showed robustness with an accuracy of 98.94% [14]. However, As dataset size increases, RF achieves high accuracy on imbalanced datasets, but it's scalability becomes an issue. Moreover, RF fails to capture complexities of feature interactions unless it is heavily tuned.
- SMOTE was used in combination with SVM and showed 97.83% accuracy on Android malware datasets, as reported by Dehkordy and Rasoolzadegan [9]. However, Sawadogo et al. [22] point out that

TABLE III. SUMMARY OF LITERATURE REVIEW

Author	Year	Best balancing techniques	Scope	Dataset	Metrics Result
Dehkordy and Rasoolzadegan [9]	2021	SMOTE	Android malware detection on imbalanced datasets.	Drebin Dataset and AMD Dataset	KNN, SVM, ID3 Accuracy: 98.69%, 97.83%, 97.59%
Guan et al. [10]	2021	SMOTE	Android malware detection on imbalanced datasets.	VirusShare (10,182 malware, 127 benign apps)	RF, KNN, NB, SVM
Khoda et al. [11]	2021	Modified Fuzzy-SMOTE	Malware detection in edge computing.	Drebin Dataset, AndroZoo and Google Play Store	DNN (F1 Score 99%)
Almomani et al. [12]	2021	SMOTE	Android Ransomware Detection in Imbalanced Data.	Custom Dataset	SMOTE-IBPSO-SVM (Specificity 98.7%)
Hemalatha et al. [13]	2021	Reweighted class-balanced loss function	Malware detection on imbalanced datasets.	Maling	DenseNet-based (Accuracy 98.46%)
Goyal and Kumar [14]	2020	Random Under-Sampling	Malware detection using ML classifiers.	Custom dataset (42,797 malware - 1,079 benign)	RF (Accuracy 98.94%)
Salas and Geus [15]	2024	Bicubic interpolation, Class weight estimation, ReduceLRonPlateau	Malware classification using DL.	Maling	MobileNet FT (Accuracy 99.08%)
Almaleh et al. [16]	2023	Undersampling	Malware detection in Windows.	Custom dataset (42,797 malware - 1,079 benign)	LR & RNN (Accuracy 98%)
Yu Ding et al. [17]	2020	Novel classification approach	Malware classification.	BIG 2015	Self-attention Neural Network (Accuracy 98.48%)
Zahra Moti et al. [18]	2020	SeqGAN	Malware detection.	Microsoft dataset	CNN-LSTM (Accuracy 98.99%)
Almomani et al. [19]	2022	-	Android malware detection on imbalanced datasets.	Leopard Android dataset (14,733 malware - 2,486 benign)	Xception CNN (Accuracy 99.40% balanced - 98.05% imbalanced)
Mimura [20]	2020	Word frequency-based feature selection	Malware detection in Microsoft document files.	VirusTotal and Stack Overflow	SVM (F-measure 99%)
Nikale and Purohit [21]	2023	SMOTE, ADASYN and Balanced Cost	Android malware detection on imbalanced datasets.	Contagio, Koodous and AP-KPure	XGBoost (Accuracy 91%)
Sawadogo et al. [22]	2022	-	Android malware detection on imbalanced datasets	CICMalDroid	RF
Haluška et al. [23]	2022	SMOTE	Imbalanced classification in Cybersecurity and other domains.	23 datasets (6 cybersecurity datasets and 17 public imbalanced datasets)	PR AUC, ROC AUC, and P-ROC AUC are 6.283, 6.174, and 4.087, respectively.
Alzammam et al. [24]	2020	Oversampling	Imbalanced multi-class malware classification.	Maling, Microsoft, and VirusTotal	Accuracy:99.94%, 98.31%, 96.06% respectively.
Phung and Mimura [25]	2021	Oversampling + ML	Static analysis for detecting malicious JavaScript.	Imbalanced dataset over 30,000 samples (PhishTank + Github).	Recall score of 72% with Doc2Vec model, outperforming baseline method by 210%.
Khoda et al. [26]	2020	Oversampling + DNN	Mobile malware detection with imbalanced data.	Drebin Android malware dataset (50,000 apps, 500 malicious)	F1-score of 94.2% and accuracy of 98.8% at 10% imbalance ratio.
Reshi and Singh [27]	2024	VAEs for generating synthetic samples	Enhance Malware detection in imbalanced datasets using DL.	Maling dataset	Accuracy 98%
Faridun and Im [28]	2024	SMOTE + TabNetClassifier	Malware detection using tabular data analysis and addressing class imbalance issues using DL.	138,047 PE header samples (41,323 benign and 96,724 malware samples)	Accuracy: 99.10%, F1-Score: 99.11% after applying SMOTE.
Li et al. [29]	2022	Multimodal fusion and Weighted Soft Voting	Malware family classification in Intelligent Transportation Systems.	Microsoft BIG-15	Accuracy: 99.2%, Macro-F1 score: 98.1%.
Onwuegbuche et al. [30]	2023	(SMOTE, Cost-sensitive learning) + ML	Ransomware detection and classification using ML models.	Elderan ransomware dataset	98.78% (binary - XGBoost with cost-sensitive learning and SMOTE), 61.94% (multi-class - RF model using cost-sensitive learning).
Andelic et al. [31]	2023	ADASYN	Improving malware detection in imbalanced datasets using GPSC and oversampling.	uci malware detection (301 malicious and 72 non-malicious).	Accuracy: 99.62%.
Çayır et al. [32]	2021	Ensemble learning with bagging RCNF.	Image-based malware family classification.	Maling and BIG2015	F-Score: 96.61%, 98.20% respectively.
LIN et al. [33]	2022	DL(VAE) + ML(MLP)	Intrusion Detection in Heterogeneous Networks.	HDFS logs	F1 score: 97%, Recall rate: 98%
Vehabovic et al. [34]	2023	Federated learning (FL) with Weighted cross-entropy loss function.	Ransomware detection and attribution using FL.	9 ransomware families (140 malicious samples each) and 2,000 benign Windows applications.	Binary detection: 94.67%, Multi-class attribution: 84.15%.
Shi et al. [35]	2024	OCC (ML:Isolation Forest, DL:Deep Autoencoder)	IoT Malware Detection.	IoT-23 dataset	F1-score: (Isolation Forest: 88%), (Deep Autoencoder: 95%).
Al-Harashsheh et al. [36]	2021	SMOTE over-sampling technique+GPMP feature selection + ML	Malware detection using ML classifiers with genetic programming for feature selection.	Ten different datasets	GPMP method with RF achieves an accuracy of 97.95% and an F1-score of 96.35%.
Al-Khshali et al. [37]	2024	Subspace Learning-Based OCC (SSVDD and GESSVDD), uses ML	Malware Detection.	3 datasets(Benign & Malicious PE Files, ClaMP, Malware Analysis Datasets: PE Section Headers by Oliveira)	100% TPR for subspace-learning-based OCC.

SVM's performance really drops when faced with larger and diverse datasets because SVM is not scalable due to its reliance on kernel based methods, and is sensitive to hyperparameter settings on real world imbalanced datasets.

4) DL Approaches:

- Studies explored DL approaches. Researchers investigated hybrid models like CNN + LSTM with Generative Adversarial Networks (GAN) to generate new synthetic samples to balance datasets in order to detect minority classes [18]. Colored and grayscale image representations of Android malware were used for detecting Android malware with vision based CNN, like Xception which reduces the number of manual feature extraction phases and increases scalability [19]. Fine tuning MobileNet through transfer learning techniques

emphasized that CNNs are effective in mitigating overfitting and enhancing generalization particularly in resource constrained environments [15].

- However, these models are computationally intensive and, therefore, unsuitable for real-time processing or on devices with low computational capabilities. Essentially, if these models were simplified or pruned, then they would most likely be more useful in terms of time and versatility.

A. One-Class Classification Models

- Innovative one class classification models proved to be effective in cases of lack of labeled data, especially in IoT environments. Isolation Forest and deep autoencoders showed high adaptability to new malware threats with 100% recall by using TF-IDF and n-grams

TABLE IV. LIMITATIONS AND CONTRIBUTIONS OF LITERATURE REVIEW

Ref	Contribution	Limitation
[9]	Improved malware detection through dataset preprocessing, feature ranking, and the use of multiple classifiers.	Limited malware family coverage and a notable false positive rate.
[10]	A hybrid Class-Imbalance Learning (CIL) method using clustering-based under-sampling combined with SMOTE.	Requires careful tuning of clustering parameters and may not generalize well to newer malware types.
[11]	Modified Fuzzy-SMOTE to handle data imbalance in malware detection.	Requires careful tuning of parameters.
[12]	Combining BPSO and SVM integrated with SMOTE, to effectively detect Android ransomware.	The dataset's small size and imbalance limit the model's generalizability.
[13]	DenseNet-based model for malware detection which visualizes malware binaries as grayscale images.	Struggles with detecting zero-day malware and has reduced accuracy for unseen malware classes.
[14]	Compares the performance of ML classifiers on balanced versus imbalanced datasets.	Potential bias which reduces the generalizability of the findings to real-world scenarios.
[15]	Combining multiple datasets into a new Fusion dataset for enhanced diversity.	Struggles with generalizing to unseen malware types.
[16]	A hybrid malware detection model combining logistic regression for weight initialization with RNN to improve detection capabilities of API call sequences.	Small balanced dataset reduces the model's ability to generalize effectively to diverse, large-scale scenarios.
[17]	Effectively addresses the issue of imbalanced datasets by treating malware ASM files as text sequences.	Struggles with the recognition of small-sample malware families and lacks interpretability for practical cybersecurity use.
[18]	CNN-LSTM hybrid model combined with SeqGAN to address class imbalance in malware detection.	High computational overhead due to SeqGAN training and reliance solely on opcode sequences which limits feature diversity.
[19]	Vision-based DL model utilizing 16 fine-tuned CNNs to detect Android malware efficiently without manual feature extraction.	Relies on pre-trained CNNs which limits adaptability to new malware types.
[20]	Detecting macro malware using Doc2vec and LSI combined with ML classifiers to address imbalanced dataset.	Dataset may not fully represent real-world conditions.
[21]	Proposed a familial classification model for Android malware using dynamic features while addressing dataset imbalance issues.	Small dataset and focused only on basic dynamic features, limiting broader applicability.
[22]	Investigates the impact of imbalanced datasets on the performance of various ML models for Android malware detection.	Uses a single dataset and evaluates a limited number of traditional ML algorithms.
[23]	Comprehensive benchmark of 16 data preprocessing methods for imbalanced classification	Slowness of Python-based implementations and need to subsample and perform feature selection on larger datasets.
[24]	Comparative analysis of techniques to address imbalanced datasets.	Complexity in using pre-trained models, and the need to consider dataset characteristics and other data factors.
[25]	Proposed an oversampling-based algorithm that improves recall score.	Does not explore other resampling techniques or compare them comprehensively.
[26]	Proposed a technique for generating synthetic malware samples that preserve malicious functionality.	Limited to Android malware and may not be directly applicable to other types of malware.
[27]	Proposed a novel approach combining VAEs with CNNs to address data imbalance	Complexity of the integrated VAE-CNN model, requiring careful resource management and hyperparameter tuning.
[28]	Combining TabNetClassifier with SMOTE to enhance malware detection accuracy.	Does not explore applications on other types of malware datasets
[29]	Proposed a novel approach combining multimodal fusion with weight self-learning + XGBoost to improve classification accuracy and mitigate concept drift.	Relies on static analysis, may not capture dynamic behaviors of malware.
[30]	Proposed a three-stage feature selection method and addressed class imbalance using SMOTE and cost-sensitive learning.	Limited to older ransomware families, dataset size, and specific balancing techniques.
[31]	Application of GPSC with oversampling techniques to achieve high classification accuracy and generate interpretable symbolic expressions.	Small dataset size and potential for oversampling techniques to introduce noise or overfitting.
[32]	First application of CapsNet in malware type classification, and ensemble model of CapsNet for imbalanced datasets.	Number of estimators limited to 10 RCNF due to increasing trainable parameters and significant training time.
[33]	ML framework that combines a VAE and a MLP to address imbalanced datasets and detect attack variants.	Potential for overfitting due to model complexity and need for further evaluation on other datasets.
[34]	Modified FL scheme to mitigate dataset imbalance.	Performance decline in multi-class attribution with imbalanced datasets.
[35]	Demonstrated the effectiveness of OCC in IoT malware detection using unlabeled benign data. Introduced TF-IDF for feature transformation.	Reliance on a specific dataset and potential need for further validation across more diverse IoT environments.
[36]	Proposed GPMP that uses genetic programming to select relevant features, leading to improved detection accuracy and reduced computational cost.	Lack of detailed analysis of the computational cost of the proposed method compared to others.
[37]	Adapting subspace learning techniques to OCC for malware detection.	Need for further exploration of subspace learning techniques in cybersecurity.

for feature transformation [35]. Furthermore, subspace learning based OCC models, such as Graph Embedded Subspace SVDD (GESSVDD), demonstrated excellent scalability and the capability of preventing the curse of dimensionality with a True Positive Rate of 100% [37].

- It is possible that using both of these methods will improve the performance of detection systems in identifying malware while at the same time protecting the privacy of users at the same time. However, the consistency across FL different systems remains a problem.

B. Federated Learning and Capsule Networks

- FL as a promising approach to ransomware detection, with privacy maintained through distributed training

without distribution of raw data. Weighted cross entropy loss improved detection performance of minority classes across different client nodes, making FL a compelling solution for real world scenarios where data centralization is infeasible [34]. In addition, CapsNet with their capability of spatial feature preserving were proven effective in imbalanced malware type classification with high F-scores using fewer trainable parameters than conventional CNNs [32].

C. Variational Autoencoders and Symbolic Classifiers

- VAE used to address data imbalance problem by generating synthetic samples to balance malware and benign instances, which helped greatly improve CNN model generalization and the accuracy was improved from 90% to 98% [27]. GPSCs showed its inter-

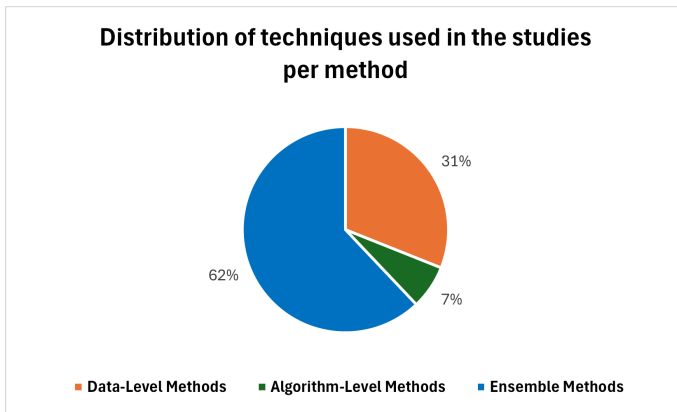


Fig. 7. Distribution of techniques used across studies.

pretability, using symbolic expressions along with oversampling techniques such as SMOTE to adequately address imbalance while still maintaining high precision and recall [31].

D. Datasets

- Maling, BIG 2015, and Drebin are examples of the datasets widely applied in malware detection research. Although these datasets have been used to assess models, the lack of variability enhances the datasets, and the imbalance ratios are not real-world. Moreover, they were chosen because they were prevalent in the reviewed studies and relevant to malware detection research. These are benchmarks in the field, often cited due to their diversity in malware types and their real world class imbalance. For example, the Drebin dataset with more than 50,000 Android applications is one of the most used datasets to validate imbalanced learning techniques [26]. The Maling dataset also covers a wide variety of malware families, and so is suited to the evaluation of DL and ensemble methods [13], [27].

The literature showed a wide range of techniques to address the problem of imbalanced datasets on malware detection. SMOTE, feature selection, DL, GANs, one class classification, and FL each had its own benefits, including improving detection rates and recall, maintaining privacy, and interoperability. Moreover, the literature that the ensemble method is the most used by the studies compared to other methods, as shown in Fig. 7. Hence, the importance of these advances to improving the robustness and efficiency of malware detection systems considering continuous evolution and diversification of threats.

VI. CHALLENGES AND OPEN DIRECTIONS

Although imbalanced datasets for malware detection have been addressed, there are still open issues.

- Oversampling and Computational Challenges: Many oversampling techniques such as SMOTE and its variants, are effective, but they can also cause computational overhead and overfitting, particularly for complex synthetic data generation [23] [26]. A crucial

need still remains to achieve the balance between computational efficiency and performance improvements, especially when working with resource constrained environments or large datasets.

- Feature Selection Complexity: Feature selection is another challenge. However, techniques such as Genetic Programming based feature selection and multi stage feature prioritization have been shown to be successful, but can significantly complicate the training process and require a great deal of fine tuning [30] [36]. However, the challenge to integrate such methods into real world scenarios where computational resources may be limited still remains. More efficient and automated feature selection approaches are also needed, that can lessen reliance on domain specific knowledge without compromising accuracy.
- DL and Hybrid Model Limitations: Despite their potential, DL models typically require largescale computations and are easily overfit over unbalanced data without a necessary regularization. While CNN-LSTM combined with GANs and CapsNet are effective, they bring along additional layers of complexity that hinder their practical deployment [18] [32]. Also, FL provides privacy preserving capabilities but comes at the cost of synchronization issues and model consistency across distributed nodes [34].
- Dataset: Future work should therefore aim at developing larger datasets that are more general and include samples of rare types of malware as well as more realistic conditions. Shared databases could be federations hence creating federated datasets that would otherwise share data securely.

Further research should be conducted to develop lightweight and computationally efficient models that can be used in real time malware detection environment. Transfer learning and FL are promising, but more work is needed to make them work effectively with imbalances without a huge computational overhead. Addressing these challenges will be key to making malware detection systems robust and practical for dynamic and diverse threat landscapes.

VII. CONCLUSION

This paper provides a comprehensive review of the problems that exist in imbalanced datasets in malware detection, presenting an investigation of existing solutions including data level, algorithm level, and ensemble methods. Key approaches to overcome data imbalance issues in malware detection are identified including SMOTE, DL hybrids like CNN and LSTM, and advanced strategies like FL. The review shows that the methods increase the malware detection rate significantly. However, issues remain like computational overhead, overfitting, and limited generalizability of these models to unseen malware types. Specifically, Modified Fuzzy-SMOTE and FL deal with some of these challenges by generating more realistic synthetic data and preserving privacy in distributed training environments.

This paper identifies a taxonomy that classifies the varied methodologies used to deal with class imbalance in malware

detection. The results show that ensemble methods are the most effective across various scenarios, especially for improving detection accuracy and robustness. The study also provides tradeoffs between computational efficiency and model performance, and provides a guide for future developments. Proposed future research may focus on creating advanced techniques and frameworks that address the difficulties of imbalanced datasets in detecting malware. The combination of Modified Fuzzy-SMOTE with feature selection methods generate realistic synthetic samples for machine learning algorithms while improving the robustness of RF and SVM classifiers. Moreover, optimizing hybrid models like CNN-LSTM becomes viable for real-time malware detection through optimization processes that include parameter sharing and model pruning mechanisms. Also, real-time data distribution adaptation in dynamic cost-sensitive learning algorithms leads to enhanced performance across malware families. Finally, the inclusion of real-world data variations with diverse samples within expanded datasets helps models achieve better generalization capabilities and maintain robustness within dynamic operational settings.

FUNDING

This work was funded by King Faisal University, Saudi Arabia. [Project No. GRANT KFU250100].

ACKNOWLEDGMENT

This work was supported through the Annual Funding track by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Project No. GRANT KFU250100].

CONFLICTS OF INTEREST

All authors declare no conflict of interest.

REFERENCES

- [1] A. Alharbi, A. H. Seh, W. Alosaimi, H. Alyami, A. Agrawal, R. Kumar, and R. A. Khan, "Analyzing the impact of cyber security related attributes for intrusion detection systems," *Sustainability*, vol. 13, no. 22, p. 12337, 2021.
- [2] H. Ali, M. M. Salleh, R. Saedudin, K. Hussain, and M. F. Mushtaq, "Imbalance class problems in data mining: A review," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 14, no. 3, pp. 1560–1571, 2019.
- [3] S. Rane, S. Yadav, Y. Hambir, A. Gupta, and E. Kapoor, "Ai-powered malware detection: Leveraging machine learning for enhanced cybersecurity," *Nanotechnology Perceptions*, pp. 1331–1347, 2024.
- [4] K. M. Hasib, M. S. Iqbal, F. M. Shah, J. A. Mahmud, M. H. Popel, M. I. H. Showrov, S. Ahmed, and O. Rahman, "A survey of methods for managing the classification and solution of data imbalance problem," *arXiv preprint arXiv:2012.11870*, 2020.
- [5] L. Wang, M. Han, X. Li, N. Zhang, and H. Cheng, "Review of classification methods on unbalanced data sets," *Ieee Access*, vol. 9, pp. 64 606–64 628, 2021.
- [6] M. Saini and S. Susan, "Tackling class imbalance in computer vision: a contemporary review," *Artificial Intelligence Review*, vol. 56, no. Suppl 1, pp. 1279–1335, 2023.
- [7] P. Kumar, R. Bhatnagar, K. Gaur, and A. Bhatnagar, "Classification of imbalanced data: review of methods and applications," in *IOP conference series: materials science and engineering*, vol. 1099, no. 1. IOP Publishing, 2021, p. 012077.
- [8] AV-ATLAS, *AV-ATLAS Malware Analysis Portal*, <https://portal.av-atlas.org/malware> Accessed: 2024-11-20. [Online]. Available: <https://portal.av-atlas.org/malware>
- [9] D. T. Dehkordy and A. Rasoolzadegan, "A new machine learning-based method for android malware detection on imbalanced dataset," *Multimedia Tools and Applications*, Apr. 2021. [Online]. Available: <http://dx.doi.org/10.1007/s11042-021-10647-z>
- [10] J. Guan, X. Jiang, and B. Mao, "A method for class-imbalance learning in android malware detection," *Electronics*, vol. 10, no. 24, p. 3124, Dec. 2021. [Online]. Available: <http://dx.doi.org/10.3390/electronics10243124>
- [11] M. E. Khoda, J. Kamruzzaman, I. Gondal, T. Imam, and A. Rahman, "Malware detection in edge devices with fuzzy oversampling and dynamic class weighting," *Applied Soft Computing*, vol. 112, p. 107783, Nov. 2021. [Online]. Available: <http://dx.doi.org/10.1016/j.asoc.2021.107783>
- [12] I. Almomani, R. Qaddoura, M. Habib, S. Alsoghyer, A. A. Khayer, I. Aljarah, and H. Faris, "Android ransomware detection based on a hybrid evolutionary approach in the context of highly imbalanced data," *IEEE Access*, vol. 9, p. 57674–57691, 2021. [Online]. Available: <http://dx.doi.org/10.1109/ACCESS.2021.3071450>
- [13] J. Hemalatha, S. Roseline, S. Geetha, S. Kadry, and R. Damaševičius, "An efficient densenet-based deep learning model for malware detection," *Entropy*, vol. 23, no. 3, p. 344, 2021. [Online]. Available: <http://dx.doi.org/10.3390/e23030344>
- [14] M. Goyal and R. Kumar, "Machine learning for malware detection on balanced and imbalanced datasets," in *2020 International Conference on Decision Aid Sciences and Application (DASA)*. IEEE, 2020, p. 867–871. [Online]. Available: <http://dx.doi.org/10.1109/DASA51403.2020.9317206>
- [15] M. P. Salas and P. L. De Geus, "Deep learning applied to imbalanced malware datasets classification," *Journal of Internet Services and Applications*, vol. 15, no. 1, p. 342–359, Sep. 2024. [Online]. Available: <http://dx.doi.org/10.5753/jisa.2024.3907>
- [16] A. Almaleh, R. Almushabb, and R. Ogran, "Malware api calls detection using hybrid logistic regression and rnn model," *Applied Sciences*, vol. 13, no. 9, p. 5439, Apr. 2023. [Online]. Available: <http://dx.doi.org/10.3390/app13095439>
- [17] Y. Ding, S. Wang, J. Xing, X. Zhang, Z. Qi, G. Fu, Q. Qiang, H. Sun, and J. Zhang, "Malware classification on imbalanced data through self-attention," in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, Dec. 2020, p. 154–161. [Online]. Available: <http://dx.doi.org/10.1109/TrustCom50675.2020.00033>
- [18] Z. Moti, S. Hashemi, and A. N. Jahromi, "A deep learning-based malware hunting technique to handle imbalanced data," in *2020 17th International ISC Conference on Information Security and Cryptology (ISCISC)*. IEEE, Sep. 2020, p. 48–53. [Online]. Available: <http://dx.doi.org/10.1109/ISCISC51277.2020.9261913>
- [19] I. Almomani, A. Alkhayer, and W. El-Shafai, "An automated vision-based deep learning model for efficient detection of android malware attacks," *IEEE Access*, vol. 10, p. 2700–2720, 2022. [Online]. Available: <http://dx.doi.org/10.1109/ACCESS.2022.3140341>
- [20] M. Mimura, "An improved method of detecting macro malware on an imbalanced dataset," *IEEE Access*, vol. 8, p. 204709–204717, 2020. [Online]. Available: <http://dx.doi.org/10.1109/ACCESS.2020.3037330>
- [21] S. P. Swapna Augustine Nikale, "Android malware detection and familial classification using dynamic features for imbalanced dataset," *European Chemical Bulletin*, vol. 12, no. 7, pp. 1508–1518, 2023.
- [22] Z. Sawadogo, G. Mendy, J. M. Dembele, and S. Ouya, "Android malware detection: Investigating the impact of imbalanced data-sets on the performance of machine learning models," in *2022 24th International Conference on Advanced Communication Technology (ICACT)*. IEEE, Feb. 2022, p. 435–441. [Online]. Available: <http://dx.doi.org/10.23919/ICACT53585.2022.9728833>
- [23] R. Haluška, J. Brabec, and T. Komárek, "Benchmark of data preprocessing methods for imbalanced classification," in *2022 IEEE International Conference on Big Data (Big Data)*. IEEE, 2022, pp. 2970–2979.
- [24] A. Alzammam, H. Binsalleeh, B. AsSadhan, K. G. Kyriakopoulos, and S. Lambotharan, "Comparative analysis on imbalanced multi-class classification for malware samples using cnn," in *2019 International Conference on Advances in the Emerging Computing Technologies (AECT)*. IEEE, 2020, pp. 1–6.

- [25] N. M. Phung and M. Mimura, "Detection of malicious javascript on an imbalanced dataset," *Internet of Things*, vol. 13, p. 100357, 2021.
- [26] M. E. Khoda, J. Kamruzzaman, I. Gondal, T. Imam, and A. Rahman, "Mobile malware detection with imbalanced data using a novel synthetic oversampling strategy and deep learning," in *2020 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*. IEEE, 2020, pp. 1–6.
- [27] H. H. Reshi and K. Singh, "Enhancing malware detection using deep learning approach," in *2024 International Conference on Automation and Computation (AUTOCOM)*. IEEE, 2024, pp. 497–501.
- [28] R. Faridun and E. G. Im, "Enhancing malware detection with tabnet-classifier: A smote-based approach," in *Proceedings of the Korea Information Processing Society Conference*. Korea Information Processing Society, 2024, pp. 294–297.
- [29] S. Li, Y. Li, X. Wu, S. Al Otaibi, and Z. Tian, "Imbalanced malware family classification using multimodal fusion and weight self-learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 7, pp. 7642–7652, 2022.
- [30] F. C. Onwuegbuche, A. D. Jurcut, and L. Pasquale, "Enhancing ransomware classification with multi-stage feature selection and data imbalance correction," in *International Symposium on Cyber Security, Cryptology, and Machine Learning*. Springer, 2023, pp. 285–295.
- [31] N. Andelic, S. Baressi Segota, and Z. Car, "Improvement of malicious software detection accuracy through genetic programming symbolic classifier with application of dataset oversampling techniques," *Computers*, vol. 12, no. 12, p. 242, 2023.
- [32] A. Çayır, U. Ünal, and H. Dağ, "Random capsnet forest model for imbalanced malware type classification task," *Computers & Security*, vol. 102, p. 102133, 2021.
- [33] Y.-D. Lin, Z.-Q. Liu, R.-H. Hwang, V.-L. Nguyen, P.-C. Lin, and Y.-C. Lai, "Machine learning with variational autoencoder for imbalanced datasets in intrusion detection," *IEEE Access*, vol. 10, pp. 15 247–15 260, 2022.
- [34] A. Vehabovic, H. Zanddzari, N. Ghani, G. Javidi, S. Uluagac, M. Raghouti, E. Bou-Harb, and M. S. Pour, "Ransomware detection using federated learning with imbalanced datasets," in *2023 IEEE 20th International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT (HONET)*. IEEE, 2023, pp. 255–260.
- [35] T. Shi, R. A. McCann, Y. Huang, W. Wang, and J. Kong, "Malware detection for internet of things using one-class classification," *Sensors*, vol. 24, no. 13, p. 4122, 2024.
- [36] H. Harahsheh, M. Shraideh, and S. Sharaeh, "Performance of malware detection classifier using genetic programming in feature selection," *Informatica*, vol. 45, no. 4, 2021.
- [37] H. H. Al-Khshali, M. Ilyas, F. Sohrab, and M. Gabbouj, "Malware detection with subspace learning-based one-class classification," *IEEE Access*, 2024.

Artificial Intelligence in Financial Risk Early Warning Systems: A Bibliometric and Thematic Analysis of Emerging Trends and Insights

Muhammad Ali Chohan¹, Teng Li², Suresh Ramakrishnan³, Muhammad Sheraz⁴

Guangdong CAS Cogniser, Information Technology Co., Ltd.,

No. 1504-1506, Wansheng North 1st Street, Nansha District, Guangzhou City, Guangdong Province, 511466, China^{1,2}

College of Artificial Intelligence, Anhui University, China²

Faculty of Management, Universiti Teknologi Malaysia, 81310, Skudai, Johor, Malaysia³

Department of Computer Science, Bacha Khan University, Charsadda, KPK, Pakistan⁴

Abstract—With the continuous development of financial markets worldwide, there has been increasing recognition of the importance of financial risk management. To mitigate financial risk, financial risk early warning serves as a risk uncovering mechanism enabling companies to anticipate and counter potential disruptions. The present review paper aims to identify the bibliometric analysis for exploring the growth and academic evolution of financial risk, financial risk management, and financial risk early warning concepts. Academic literature is surveyed from the Scopus database during the period 2010-2024. The network analysis, conceptual structure, and bibliographic analysis of the selected articles are employed using VOSviewer and Bibliometric R Package. The biblioshiny technique based on the bibliometric R package was used to draw journal papers' performance and scientific contributions by displaying distinctive features from the bibliometric method used in prior studies. The data was extracted from Scopus databases. In addition, this study comprehensively analyzes the evolution of financial risk early warning systems, highlighting significant trends and future directions. Thematic evaluation across 2010-2015, 2016-2021, and 2022-2024 reveals a shift from traditional statistical methods to advanced machine learning and AI techniques, with neural networks, random forests, and XGBoost being pivotal. Innovations like attention mechanisms and LSTM models improve prediction accuracy. The integration of sustainability factors, such as carbon neutrality and renewable energy, reflects a trend towards incorporating environmental considerations into risk management. The study underscores the need for interdisciplinary collaborations and advanced data analytics for comprehensive financial systems. Policy implications include promoting AI adoption, integrating environmental factors, fostering collaborations, and developing advanced data analytics frameworks.

Keywords—Artificial intelligence; deep learning; financial risk management; early warning systems; bibliometrics analysis

I. INTRODUCTION

In the dynamic landscape, the financial risk of companies is an unavoidable risk and inevitable companion which is reflected in all parts of company investment and financing management. The presence of financial risks presents a huge vulnerability to the healthy advancement of companies [24]. In an era of volatility, uncertainty, complexity, and ambiguity firms have been subjected to unprecedented exposure, which complicates decision-making [1]. In the current era of the Internet plus, the world economy is becoming more and more

globalized and informational. The business environment is changing rapidly, and the business development of enterprises is facing unprecedented opportunities for their operation and development. However, it is also facing financial uncertainties brought about by the fluctuations of the general economic environment, and the company is facing increasing financial risks. It also faces unpredictable environmental factors and challenges such as economic market factors, laws and regulations, social and cultural factors, and policy environment factors which bring uncertainty to the financial situation of enterprises, and the financial risks faced by enterprises are also increasing.

The role of the financial market in enabling societies to reach the low carbon economy is well understood [26]. The market competition has become rigorous under the influence of "economic globalization" and enterprises are under pressure for both survival and growth. Low-carbon development has greatly changed the external environment and financial environment of enterprises, thereby increasing the financial risks that exist in enterprises. In the context of financial development globalization, financial market transactions between countries are frequent, the financial environment is more complex, and the spread of financial risks is more rapid and extensive. Especially, in the current international financial situation with high leverage, high asset prices, high market volatility, and high risk, financial supervision will become more difficult, and the possibility of a financial crisis outbreak is higher than before. A financial crisis will not only destroy a country's financial system and international financial order but also cause great damage to the real economy, causing an economic crisis, and even causing a serious social and political crisis, endangering national security.

Financial risk exists in the management process of an enterprise, and poor management or poor decision-making can cause the level of financial risk to exceed alarming values and lead to financial crisis [24]. The reason why most enterprises encounter a serious financial crisis or even close in the later stage is that they do not pay full attention to the initial financial problems and do not take effective measures to deal with the crisis in time. Therefore, it is very practical to establish a scientific data model to analyze and forecast the financial data of enterprises. It can not only monitor the financial situation of enterprises in real time but also play an

effective role in financial early warning [22]. Financial risk warning has become an important part of modern enterprise financial management. It helps enterprises better warn, prevent, and control financial risks which can reduce the loss and increase the profit. In recent years, there has been worldwide research on the issue of financial risk in developing countries, particularly from the perspective of a low-carbon economy it has become a hot issue.

Detailed literature is scarce on financial early warning predictions for the financial risk of enterprises. Recently, [19] documented that bibliometric analysis has been prominently conducted for the literature on topics including sustainable and Islamic finance [31], credit risk [33], financial crises, and efficiency measurement [10]. However, other topics, such as liquidity risk or ownership structure, have been comparatively neglected. Furthermore, based on our extensive review and study, it has been found that researcher focused on early warning systems in business, finance, and economics, and [44] have worked on risk management. In addition, much empirical research has been conducted on financial risk for enterprises, banks, and currency crises. However, a lack of evidence has been found that focuses on bibliometric analysis of financial risk early warning systems. Therefore, the goal of this study is to explore the recent progress, challenges, and future directions of financial early warning predictions to capture the significance of financial early warning predictions and linked areas through bibliometric and thematic analysis. Despite the growing importance of financial risk early warning systems, there remains a significant research gap in comprehensive bibliometric and thematic analyses that leverage updated data from the Scopus database. Previous studies have not fully utilized these methods to map the evolution and emerging trends in this field, particularly over the extended period from 2010 to 2024. This study addresses this gap by employing a robust bibliometric and thematic analysis, offering a novel perspective on the financial risk early warning landscape. By systematically categorizing themes into high occurrence and link strength keywords, emerging topics, niche areas, interdisciplinary and technological integration, and sustainability and innovation, this research provides a detailed and updated overview of the current state and future directions of financial risk management. This study's findings hold significant implications for policymakers. The identification of cutting-edge technologies such as neural networks, random forests, and XGBoost, as well as emerging areas like attention mechanisms LSTM, and GRU models, underscores the need for advanced analytical tools in financial risk prediction. Furthermore, it focuses on sustainability and consolidation of environmental factors into risk management which signifies the growing connectivity of financial stability and environmental responsibility. By providing a comprehensive overview of current and future trends, this study equips policymakers with the insights necessary to foster innovation and sustainability in financial risk management, ultimately contributing to more resilient and adaptable financial systems.

Our main objectives of the study are as follows:

- 1) To examine the historical distribution of financial risk early warning system articles, illustrating contributions through metrics such as average citation per year, core sources by Bradford's Law, most cited

countries, corresponding author's countries, and most relevant sources.

- 2) To identify prevalent research themes in financial risk management, highlight potential collaborations and interdisciplinary research opportunities, and suggest venues for future research to advance the field.
- 3) To offer policy recommendations to enhance the adoption of advanced AI techniques and sustainability considerations in financial risk management, encourage interdisciplinary collaborations and the development of advanced data analytics frameworks, and outline future research directions to address gaps and build on the study's findings.

A. Methods

The method used in this study is bibliometric analysis, which was first used by [34] and is popular among the researchers in supporting quantitative analysis in understanding the literature. The word bibliometric is the statistical analysis of scholarly communication through publications and the most common methods are variants of citation analysis. Bibliometric analysis is usually a machine-like mechanism to understand the research trends globally in an area of interest based on the output of academic database literature and is different from a typical review paper focusing on recent progress and challenges and future directions for a specific topic [20]. Bibliometric is a powerful tool for the management of information providing useful analytical results across many fields and its application in finance are relatively recent [25], [42].

B. Search Strategy and Sources of Data

According to [15] Scopus, Web of Science, and Google Scholar are the three main databases for academic literature and citation indexes. However, this study chooses the Scopus database because Scopus is the largest citation and abstract database covering a wide range of subjects and thus, this is an attempt to cover more topics, that might not be available with the Web of Science database. Google Scholar is not selected because it does not have a strong quality control process [13]. The first search on the Scopus database was conducted on 5 October 2023 with a central theme of "enterprise financial risk" in the title, abstract, and keywords resulting in 112 documents. Moreover, subsequent searches were made between October 2023 and December 2023, on a trial-and-error basis to check for any different results, issues, and shortcomings with the Scopus Database. The authors use the same central theme along with enterprise financial management and various other variants like financial risk prediction, corporate financial information risk, crisis assessment, and credit risk management, and the authors got different results. This change in the result at a different time was the same when other variants in the search string were used. Following this, the author used different search strings which finally led to the result of 150 documents by restricting it to journal articles only as these are most used to present academic novelties [13].

The subsequent keywords were explored simultaneously with a central theme as: with a central theme as "enterprise financial risk" OR "early warning system" and its related concept in the title level of the search tool which resulted in 150 documents. These comprehensive search strings were

selected because they are highly associated with the topic of interest in this study to cover the relevant body of literature. Thus, the final query string used is as follows: TITLE-ABS-KEY (“enterprise financial risk” OR “enterprise financial management” OR “enterprise financial information management” OR “financial risk” OR “financial risk prediction” OR “financial risk analysis” OR “financial risk assessment” OR “Financial management risk prediction” OR “financial crisis management” OR “corporate financial information risk” OR “corporate financial information risk management” OR “crisis assessment” OR “credit risk management”) AND (“Early warning system*” OR “ews*” OR “Financial early warning system*” OR “financial early warning model” OR “Financial risk warning” OR “financial risk prediction” OR “financial risk prediction model” OR “Risk prediction” OR “financial analysis system*”) AND (“neural network*” OR “deep learning”) AND (LIMIT-TO (LANGUAGE, “English”)) AND (LIMIT-TO (SRCTYPE, “j”)) AND (LIMIT-TO (PUBSTAGE, “final”)) AND (LIMIT-TO (OA, “all”)).

For selecting and screening the article, the authors used guidelines by “preferred reporting items for systematic reviews and meta-analyses” (PRISMA) for systematic research reviews as shown in Fig. 1. While screening, the eligibility and inclusion criteria the authors could ensure that no potential review article skips the Scopus Database filter. To check and ensure that there are no review articles, or any other potential irrelevant articles present in our analysis that might not have skipped our filtration process. Therefore, additional phrases such as bibliometric review, scientometric review, systematic literature review, systematic review, meta-analysis, science mapping, development, progress, recent, revisit, trends, prospects, advance, perspectives, reviews, and so on, as mentioned in [20] were used using conditional formatting toolbar in MS Excel and noted the documents Electronic Identifications (EIDs). Moreover, these articles were examined by reading the title and abstract, and if needed, the articles were read thoroughly to make sure that the articles were related to enterprise financial risk and financial risk warning. There are no missing authors’ names and IDs, and no articles were found to be duplicated.

C. Bibliometric Analysis Using VOSviewer

The study used VOSviewer (version 1.6.18) developed by the Centre for Science and Technology Studies, Leiden University, Leiden, Netherlands, to visualize the financial risk warning system and its related concepts in the number of articles, prolific authors, and most productive journals in maps. The descriptive analysis shows the configuration of many articles, types of articles, and articles over time of the Scopus Database, while the bibliometric analysis is performed using the co-authorship and co-occurrence analysis. Fig. 2 depicts the methodology for bibliometric analysis using VOSviewer. Information regarding the final 150 documents like bibliographical information, citations, and keywords were exported to VOSviewer’s latest version. The authors used VOSviewer because our main objective is to focus on an aggregate level and over time development of a research area [13]. VOSviewer is a tool for creating and displaying bibliometric maps using items. In this study, the objects of interest like author keywords and countries are the items. There can be a connection, relation, or link between each pair of items. Each connection

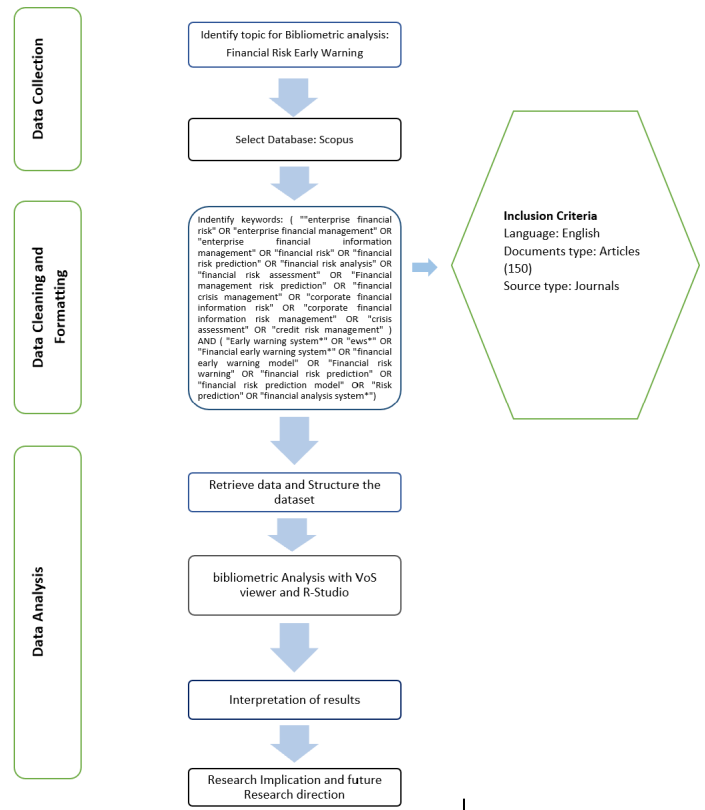


Fig. 1. Research process adopted in the study. Author’s compilation.

or link has a strength, which is represented by a positive numerical value, and the higher the value is, the stronger the link between the two linked items becomes. In the case of co-authorship analysis, the link strength between the countries shows the number of publications that two affiliated countries have co-authored. Meanwhile, the co-authorship links the total strength of a given country to other countries. In the case of co-occurrence analysis, the link strength between author keywords shows the number of publications in which two keywords occur together.

1) *Co-authorship Analysis*: In scientific research, co-authorship is the most formal manifestation of intellectual cooperation. It entails collaborating with two or more authors in conducting a research study, resulting in a higher quality or quantity research output than could be achieved by a single author [16]. Co-authorship research can be done at the organization and country level because bibliographic data contains details about the authors’ institutional affiliations and geographic positions. In this analysis of co-authorship, the unit of analysis chosen is country; therefore, the authors have included all the countries affiliated with many authors. The international research collaboration domain is under the influence of bibliometric research analysis and its main methodology is co-authorship analysis [9]. This study considered 21 affiliated countries and the affiliated countries were clustered into seven regions: South Asia (Region 1), Africa (Region 2), East Asia, and the Pacific (Region 3), Europe and Central Asia (Region 4), Latin America and the Caribbean (Region 5), Middle East and North Africa (Region 6), and North America (Region 7).

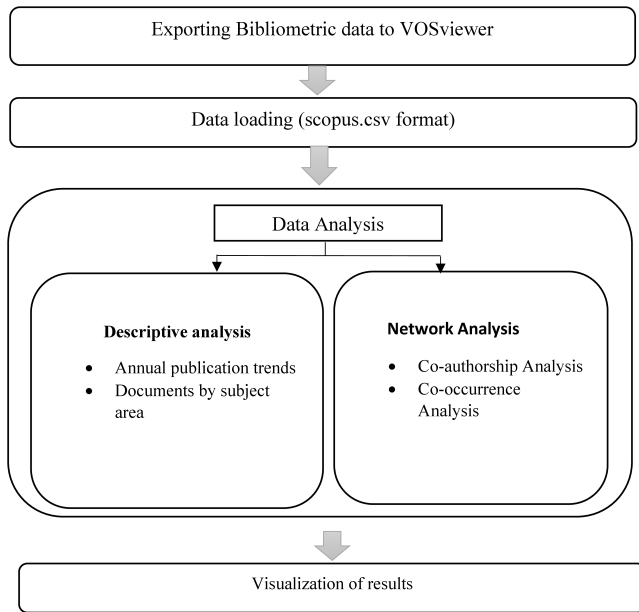


Fig. 2. Methodology for bibliometric analysis using VOSviewer.

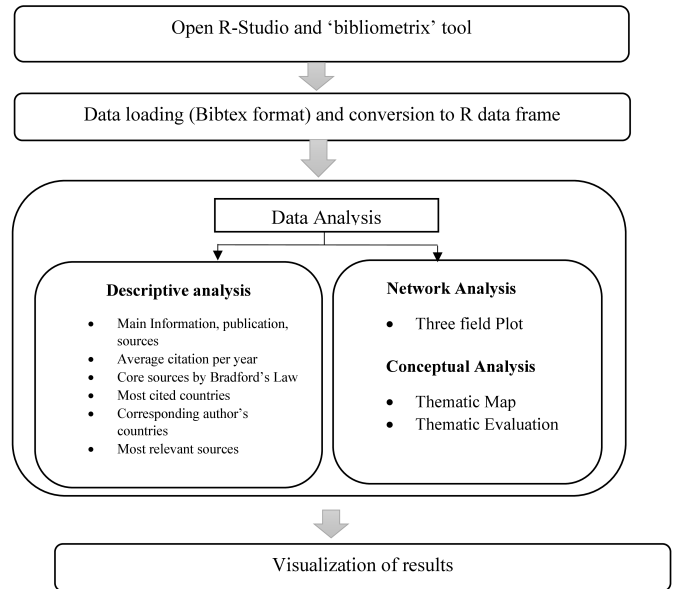


Fig. 3. Methodology for bibliometric analysis using "Bibliometrix".

2) *Co-occurrence Analysis*: Co-occurrence analyses can be used to analyze the connections of author keywords used to make a conceptual structure of the study. Researchers create a complex network using keywords because actors combine and link the words to make an interesting funnel and aggregate. Co-occurrence analysis is the only technique that uses the contents or keywords of the document to find associations among the documents. At the same time, the other approaches link the document indirectly by co-authorship or through citation [11]. Likewise, co-authorship, the association among keywords used by research studies, is represented by the strength of the keywords used in the publications in the case of co-occurrence analysis. This technique can be used that utilize the contents of the publications to make a similarity measure among documents. However, the other methods connect documents indirectly through citations and co-authorship analysis [11]. Before transferring to VOSviewer, synonymous terms were identified and replaced with a single term. The study sets a minimum limit of keyword linkage to 2 in co-occurrence in VOSviewer. Overlay visualization is considered to explore the keywords' yearly publications, occurrences, and connections.

D. Bibliometric Analysis Using Bibliometrix

As the number of published research continues to grow at an increasingly rapid rate, the effort required to accumulate knowledge becomes more complex. "Bibliometrix" is a tool programmed in the R platform (<https://www.bibliometrix.org>) to perform a comprehensive bibliometric analysis of published literature. There are several packages in R dealing with bibliometrix; however, none of them address the entire workflow process [4]. The procedure for performing bibliometric analysis using "Bibliometrix" is shown in Fig. 3.

II. DISCUSSION

A. Data Analysis and Bibliometric Maps

Table I provides a comprehensive summary of the main information, publications, and sources regarding the articles selected from the Scopus database for the bibliometric analysis of financial risk early warning from 2010 to 2024. The dataset encompasses a total of 77 sources, which include various journals, books, and other publications, amounting to 150 documents. Despite the consistent number of documents each year, the annual growth rate remains at 0%, indicating no year-over-year increase in publications. On average, these documents are relatively recent, with an average age of 2.07 years, and each document has been cited approximately 5.96 times. Interestingly, no references were reported in this summary.

The content of these documents includes 929 instances of Keywords Plus, which are terms frequently appearing in the titles of an article's references and are used to enhance the author's keywords. The authors provided a total of 291 unique keywords. The analysis reveals that 335 authors contributed to these documents, with 33 of them authoring single-authored papers. On average, each document had 2.71 co-authors, and 17.45% of these documents featured international co-authorship, highlighting the global collaboration in this research area. The types of documents varied, with the majority being articles (129), followed by conference papers (13), reviews (4), book chapters (2), and retracted papers (2). Table I summarizes the key information of the articles selected from the Scopus database using the "bibliometrix" tool.

Fig. 4 illustrates the annual trend in the number of articles published on the topic of financial risk early warning from 2010 to 2024. The x-axis represents the years from 2010 to 2024, while the left y-axis shows the number of seed index articles published each year and the right y-axis depicts the cumulative number of these articles. The bar chart represents the yearly publication count, and the line graph indicates the

TABLE I. SUMMARY OF THE SELECTED ARTICLES FROM SCOPUS DATABASE

Description	Results
Time-span	2010:2024
Sources (Journals, Books, etc)	77
Documents	150
Document Average Age	2.07
Average citations per doc	5.96
DOCUMENT CONTENTS	
Keywords Plus (ID)	929
Author's Keywords (DE)	291
AUTHORS	
Authors	335
Authors of single-authored docs	33
AUTHORS COLLABORATION	
Single-authored docs	33
Co-Authors per Doc	2.71
International co-authorship %	17.45
DOCUMENT TYPES	
article	129
book chapter	2
conference paper	13
retracted	2
review	4

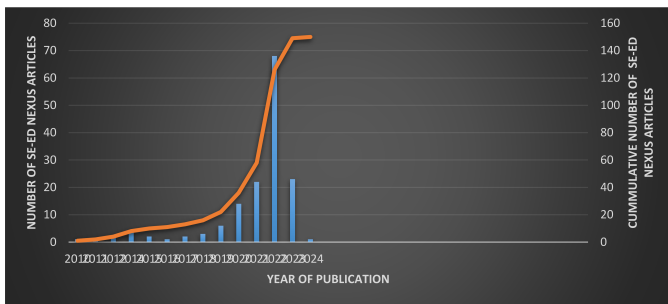


Fig. 4. Annual trend of financial risk early warning (2010-2024) Source: Scopus Database.

cumulative total. The data reveals a clear upward trend, particularly from around 2019 onward, highlighting a significant increase in publications related to financial risk early warning during this period.

Fig. 5 presents a pie chart categorizing the selected documents by their subject area, offering a visual overview of the distribution of topics. The chart shows that the majority of documents fall within the Social Sciences (19.3%), followed closely by Business, Management, and Accounting (16.8%), and Economics, Econometrics, and Finance (15.7%). Environmental Science accounts for 14.3% of the documents, while Energy comprises 11.0%. Other fields are represented to a lesser extent, including Engineering (5.9%), Computer Science (4.0%), Arts and Humanities (3.3%), Decision Sciences (2.1%), and Mathematics (2.0%). The remaining 5.5% of documents are categorized under “Other”, indicating a diverse range of additional subject areas.

Fig. 6 illustrates the average number of citations per year for publications related to financial risk early warning systems. It helps in understanding the impact and relevance of research over time. A higher average citation per year indicates that the work is widely recognized and used by other researchers in the field. In Fig. 7, Bradford’s Law describes the distribution of articles on a particular subject in scientific journals. This figure shows the core journals that publish the

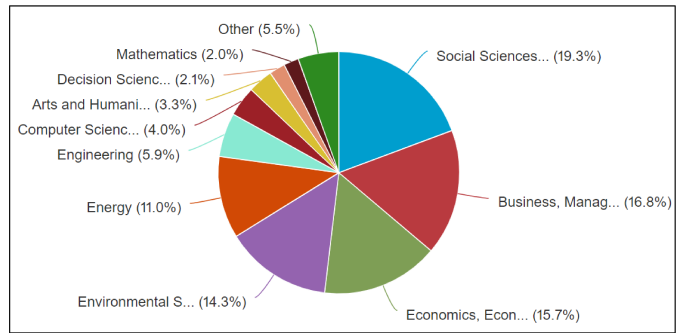


Fig. 5. Documents by subject area.

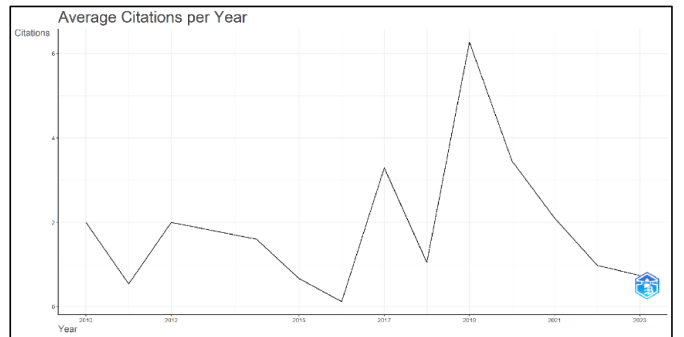


Fig. 6. Average citation per year.

most significant number of articles on financial risk early warning systems. It helps in identifying the key sources and journals that contribute extensively to the research in this domain. Fig. 8 highlights the countries whose research on financial risk early warning systems has received the most citations. It shows the geographical distribution of influential research and indicates which countries are leading in this field. Fig. 9 represents the countries of the corresponding authors of the publications. It provides insight into the geographical distribution of researchers who are contributing to the literature on financial risk early warning systems. Fig. 10 lists the most relevant sources or journals that publish articles on financial risk early warning systems. It helps researchers identify the best sources for publishing their work and staying updated with the latest research.

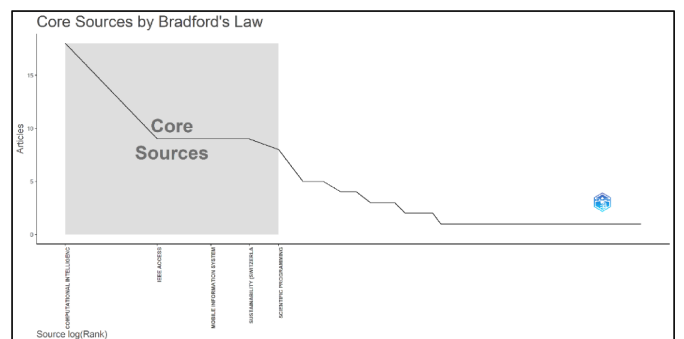


Fig. 7. Core sources by Bradford's law.

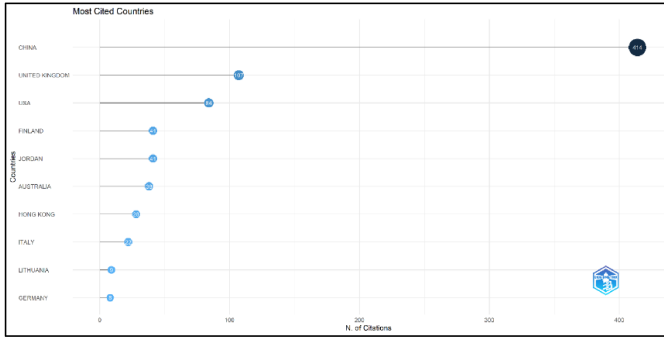


Fig. 8. Most cited countries.

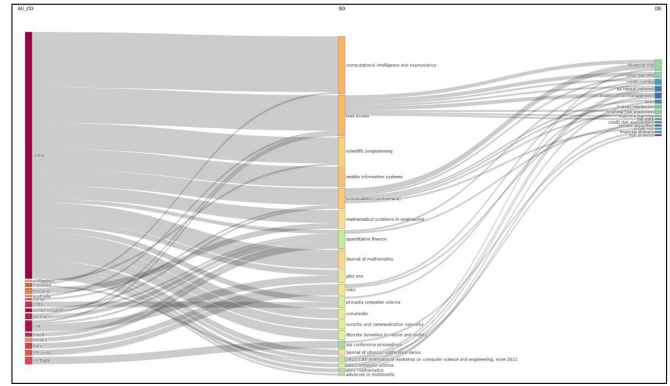


Fig. 11. Three fields plot of country-journal-keyword.

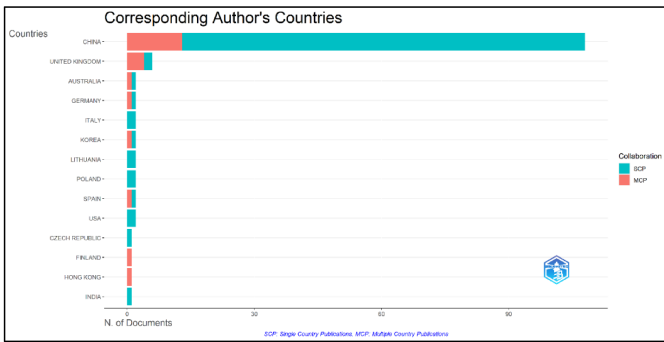


Fig. 9. Corresponding author's countries.

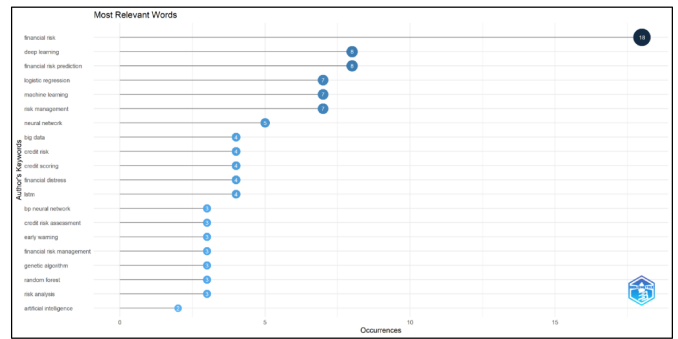


Fig. 12. Most related keywords.

Fig. 11 visually represents the relationship between countries, journals, and keywords in publications related to financial risk early warning systems. It shows how different countries and journals are linked through common research themes and keywords, providing a comprehensive overview of the research landscape.

Fig. 12 titled “Most Relevant Words” displays the occurrences of various keywords related to financial risk early warning systems. The Keywords such as “financial risk”, has the highest frequency of occurrence which is 8 times, followed by “deep learning,” and “financial risk prediction”. It reflects the critical significance and recent research progress of these keywords. Terms like “logistic regression”, “machine learning”, and “risk management” appear with 7 occurrences

which supports the fact that these statistical and analytical tools are commonly used in the existing financial risk studies. The keyword “neural network” has been used 7 times which portrays the fact that it is widely used in modeling and prediction of financial risks. The terms “big data”, “credit risk”, “credit scoring”, “financial distress”, and “LSTM” occurs four times which underlines the importance of advanced data techniques and specific risk factors reviewed in the existing body of literature. The key terms of “BP neural network”, “financial risk management”, “credit risk assessment”, “early warning”, “random forest”, “risk analysis”, and “genetic algorithm” were repeated 3 times which revealed acute nature and special approaches and focus areas within the general subject. Lastly, “artificial intelligence” appears with 2 occurrences emphasizing the growing role of AI in financial risk mitigation. All the keywords are supported by the studies of [32] on deep learning applications which reflects the current research focus in the emerging area of financial risk early warning systems.

B. Co-authorship Analysis

Fig. 13 uses various colors to represent the distribution of countries across seven regions. The thickness of the lines indicates the strength of connections between countries, with thinner lines denoting weaker links and thicker lines denoting stronger ones. For instance, Cluster 3, which represents East Asia and the Pacific, includes nine countries, while Cluster 4, representing Europe and Central Asia, includes eight countries. Clusters 1, 2, 5, and 7 each include one country, and Cluster 6 (Middle East and North Africa) does not include any country

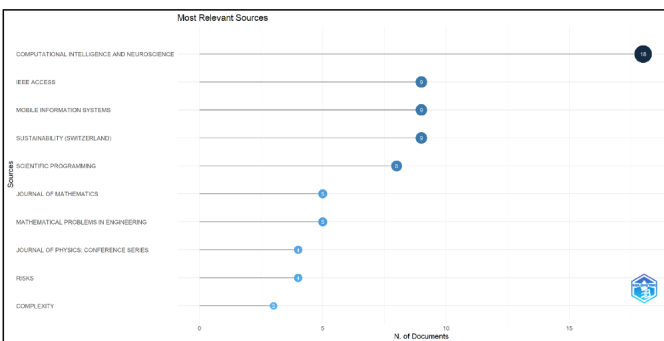


Fig. 10. Most relevant sources.

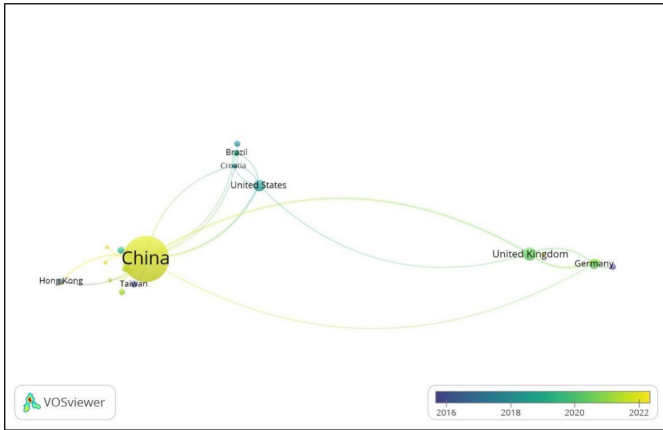


Fig. 13. Screenshot of bibliometric map based on co-authorship with overlay visualization. (It can be opened through <https://bit.ly/41uMpVT>).

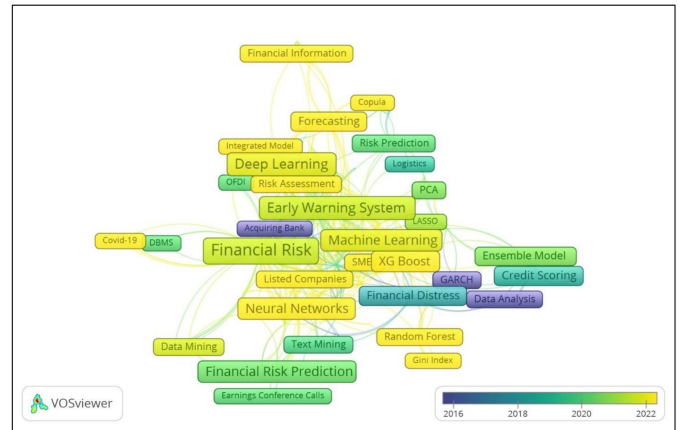


Fig. 14. Bibliometric map based on Co-authorship with overlay visualization. (It can be opened through the URL in VOSviewer: <https://bit.ly/3NCzewu>).

working in this area. Our co-authorship statistics show that China has the highest degree of affiliation, with 15 links and a link strength of 24. This means China is linked to 15 territories or countries with 24 instances of co-authorship. The United Kingdom follows with six links and a link strength of nine, the United States with five links and a link strength of eight, Germany with five links and a link strength of seven, Brazil with five links and a link strength of five, and Hong Kong with two links and a link strength of three. Other countries have fewer than two links. International collaboration occurs for various reasons, including the subject matter, the type of issue, and the researchers chosen to work on it. Additionally, the ease of access to primary data, such as financial risk early warning, can influence the relevance of the region to funders who support the research, research partners, diversity, and collaboration.

C. Author Keywords and Current Emerging and Future Trends Regarding Financial Risk Early Warning System

This section represents the objective, which explains the current trends and arena for further research and potential collaboration using VOSviewer software. For mapping in VOSviewer, a total of 291 keywords were recorded and after re-labeling various variants or synonymic single words and phrases, 177 keywords met the threshold of a minimum of five occurrences. Our results portray that financial risk is the most reflected keyword with 20 occurrences, 55 links to other keywords, and a total link strength of 74 followed by financial risk management (11 occurrences, 34 links, 47 link strength), and early warning system (11 occurrences, 29 links, 37 link strength). Some other methodological terms include deep learning (11 occurrences, 26 links, 37 link strength), logistic regression (9 occurrences, 36 links, 42 link strength), machine learning (8 occurrences, 28 links, 37 link strength), and Bp neural network (7 occurrences, 22 links, 25 link strength). Financial risk and financial risk management and early warning systems are also seen to be co-occurring with each other. whereas the early warning system keywords have links with other new emerging keywords i.e. low carbon economy, financial stability, internet finance, and internet of things. This bibliometric image shows that the two big bubbles

financial risk and financial risk management are areas used with common areas like financial distress machine learning, Xg boost, logistic regression, Bp neural network, clustering, deep learning, and nearest neighbors (Fig. 14).

The analysis of the keywords of the current articles allows for identifying the emerging trends in the research field [13]. The recent average year of publications represents potential hotspots for the future, and the smaller number of occurrences indicates the niche area [20]. Thus, Table II offers an in-depth analysis of author keywords and future trends in the realm of financial risk early warning systems. The table lists keywords alongside their occurrences, total link strength, and average publication year, thereby elucidating the current research focus areas and emerging topics within this field. Table II keywords are categorized as high occurrence and link strength keywords, emerging topics and niche areas, integration of non-related fields and technologies, sustainability and creativity, and other important keywords.

1) *High Occurrence and Link Strength Keywords:* The following keywords occur frequently and their link strength in conducting studies on the financial risk early warning system underscores the crucial role of employing machine learning and deep learning approaches in this area. Neural networks are one of the most important areas of research, being mentioned 9 times, having a link strength of 34, and the average publication year 2022. This approach enables us to train models with high levels of complexity, capable of identifying intricate patterns in the financial data, thus improving predictive performance. This is evident from [32], [41] who showed how neural networks can be applied to analyze large volumes of data by learning from huge data sets and adapting to new models of risks in the financial markets. Further, random forests occur 3 times with 13 link strengths and an average year of publication in 2022, 33, highlighting its importance. Random forests are highly valued due to their ability to handle diverse, large, and even noisy data sets in financial applications. Random forests, which construct multiple decision trees and combine their outputs, can give accurate risk estimates and overcome overfitting. The [5] further stated that there are several benefits of random forests, especially in financial risk management, mainly because of their ability to manage big data and many

variables which makes them very popular among practitioners. The term XGBoost stands out as a prominent approach that has the most link strength of 32 and is used 7 times with an average publication year in 2022 is 57. It is popular due to its ability to handle big data and is considered one of the best algorithms for predictive analysis. It also employs gradient-boosting techniques in the improvement of the predicted results and improves model accuracy. Research by [14] demonstrated that by employing the iterative nature of the XGBoost approach, additional improvements can be achieved to improve model predictive capabilities, which makes it valuable in the field of developing early warning systems for financial risks. Based on these keywords, it is shown that machine learning plays a crucial role in improving the capabilities of early warning systems for potential financial risks. Through the integration of the advantages of neural networks, random forests, and XGBoost, it is possible to improve the accuracy of these models and achieve better prediction of capacities in the financial risks that can help in the development of resilient financial systems that might be more sustainable.

2) *Emerging Topics:* Emerging themes that stand out in the field of financial risk early warning system research are the keywords “Attention Mechanism” and “LSTM (Long Short-Term Memory)” which represent the cutting-edge approaches and trends in the field. Even though the attention mechanism only emerged twice, it has a rather large link strength 8 and, the average year of the publication is 2023. This indicates that it is gradually becoming more relevant since it aims at increasing the explainability and reliability of financial risk models. The use of attention mechanisms gives the model the ability to adjust the relevance of the input features, thus resulting in better classification. LSTM is the most recurrent term with 5 references and a link strength of 17; it is mostly used in articles published in the year 2022 80, implying that this activity has become increasingly essential in recent years. Hence, LSTM models are highly effective in capturing temporal dependencies of financial data and are especially useful in determining temporal characteristics of financial data sets where past trends can influence future outcomes. Their ability to retain information over long sequences assists in capturing the temporal dependencies inherent in financial data, which are useful in improving the output of risk estimation [38]. Therefore, these topics demonstrate that innovative solutions continue to be introduced to increase the capabilities of early warning systems for financial risk management.

3) *Niche Areas:* Niche areas of concern in the financial risk early warning system are keywords like green credit risk, conditional quantiles, and sliding window, which are concerned with specific and emerging facets of financial risk management. Green credit risk features industry-specific credit risk evaluation that considers ecological aspects. This keyword emphasizes the need to come up with sector-specific solutions to managing risk that is associated with environmental sustainability. Given that many industries and financial institutions are now aware of the effect of environmental factors on financial stability, it becomes apparent that it is pertinent to work on the integration of green credit risk with these factors. In their article, [45] explain how integrating environmental factors into reporting processes that involve evaluations of risk, mitigation, and impact can provide a wider and truer perspective of the potential dangers and consequences of certain actions

by identifying carbon footprints, investing in renewable, and incorporating sustainability measures. This niche area is of great importance for the development of financial solutions that make it possible not only to mitigate various risks but also to achieve organizational growth in terms of sustainable development; the thought refers to the ecological aspect of financial risk management.

Conditional quantiles with 1 paper and a link strength of 4 and sliding window as another topic with 1 paper and a link strength of 4 published around 2023 also present emerging niche areas, which highlight that modern academic research in the field is focused on more sophisticated statistical methods to learn refined financial risks models. On the other hand, conditional quantiles focus on the probability distribution of the financial data so that business organizations can determine the likelihood of a particular risk, offering a better solution than simple quantiles regarding probable risks. This approach provides a better assessment of risk considering all the probabilities rather than risk means or medians [29].

Collectively, these niche areas highlighted that there is an immense focus on improved and specific innovations in the management of financial risk. Some of them underscore the significance of sector-based risk analysis and the application of state-of-the-art techniques in statistics to enhance the predictive capabilities of financial risks. These approaches contribute to a more resilient financial structure due to its ability to handle challenges in the current unprecedented complex financial world.

4) *Interdisciplinary and Technological Integration:* The keywords such as industry-academia linkages and public-private partnerships demonstrate that interdisciplinary and technological integration in financial risk early warning systems involves synthesizing various approaches and the use of complex analytical tools. Industry academia linkages and the use of public-private partnerships reflect the growing trend of interdisciplinary in the study of financial risk. Although these keywords are infrequent, with link strengths of 4 and an average year of publication around the year 2022, they highlight the need to integrate knowledge from various sectors to solve multifaceted financial risks.

In [6], the researchers demonstrate how these linkages and partnerships enable the exchange of knowledge, resources, and innovations between the academy, businesses, and government agencies. They may contribute to the generation of better management solutions to the risk factors that are associated with organizations. Through such arrangements, it is possible to solve complex financial risks that could be managed only with multiple companies’ resources, thus enhancing the comprehensiveness of managing the financial sector’s stability.

5) *Sustainability and Innovation:* Sustainability and Innovation in managing financial risks are emerging and reflected in keywords such as Carbon Neutrality Renewable Energy and Green Technology innovation. These terms suggest that there is an advancement towards the enhancement of sustainable practices together with the adoption of technologies for the enhancement of sustainable financial systems.

The keywords of Carbon Neutrality and Renewable Energy with average publication years of about 2022-2023 and moderate link strength underline the growing focus on environ-

mental sustainability within financial risk management. The [41] documented that the banking sectors and industries are gradually integrating environmental factors into the evaluation of risks. This transformation is due to increasing awareness of the financial implications of climate change and sustainability-related concerns. Integrating carbon neutrality and renewable energy into the financial risk analysis, the goal is to minimize future threats linked with detrimental environmental impacts and shift in legislation. This proactive approach assists in managing the identified financial risks more effectively, while thereby promoting more sustainable financial practices and, consequently, the development of a stronger financial system in line with global goals.

Green Technology Innovation has an average year of publication in 2023 that supports the importance of adopting new technologies in managing financial risks. The authors highlight that adopting green technologies can greatly diminish the negative effects on the environment and improve the sustainability of the systems used in finance. The application of such technologies is essential in creating new forms of innovative, better, and sustainable financial products and services. Green technology development enables moving to a green energy base, utilizing energy-saving technologies, and lowering CO_2 intensity. These advancements are critical in managing risks that are infectious in the financial systems including regulatory shocks, resource depletion, and climate changes.

Together, these keywords highlighted a paradigm shift regarding financial risk management and the incorporation of sustainability and innovation in it. Carbon neutrality and renewable energy demonstrate the progressive inclusion of environmental factors in risk management and addressing current and future risks. Green technology innovation refers to the use of innovative technologies for the improvement of sustainable innovative financial systems. By focusing on these areas, the financial industry not only seeks to minimize risks related to the environment; it also strives to grasp the opportunities in the future green economy. Thus, the combination of the strategy of sustainability with a focus on innovation provides financial institutions with the ability to be prepared for the current challenges in financial risk management and contribute to sustainable development at the same time.

6) *Additional Important Keywords:* The keyword COVID-19 is identified as occurring only once; however, it possesses a link strength of 4 and has an average publication year of 2023, which underscores the massive impact of the pandemic on various financial systems around the globe. The COVID-19 pandemic has posed several unprecedented developments in the financial market and has challenged the stability of financial institutions. The [21] presented that due to the COVID-19 pandemic, organizational economic losses, and other crises that occurred throughout the world, risk management has become an essential component for organizations to manage and ensure that they will not be affected negatively in the future. The dynamics of the financial environment especially in the money markets have been dynamically changing and the financial sector has been forced to respond to changes such as fluctuating volatilities, liquidity risks, and credit risks. The usage of this keyword in financial risk research also means that scholars are constantly evaluating the impact of the pandemic on the financial sector to capture the long-term effects and

the necessity of developing strategies that could be applied to financial shocks in the future.

Moreover, renewable energy with an occurrence of 1 time and a link strength of 3, published around 2022, declares the further linkage of financial risk management with environmental aspects. In [39], authors firmly stresses that the paradigm shift towards the use of renewable energy sources is not only one of the distinctive challenges that the world must address to mitigate climate change effects but also a factor that poses great risks in the financial sphere. This paper aims to look at the new opportunities and threats for funding green projects including wind, solar, and other renewable energy projects for banking organizations. These projects usually call for substantial upfront investments and many projects are legally and economically risky, but at the same time, they bring numerous advantages that can be considered in the long term, such as decreased operational expenses and compliance with sustainability objectives. The attention towards renewable energy for the study of financial risk has become more significant as the world focuses on the possibilities to minimize the financial risks of a more sustainable energy infrastructure.

The keyword of financial stability appears with 1 occurrence, its link strength is 4, and it was published in 2023, illustrating that the authors and researchers are equally fascinated by sustaining stability in the financial systems irrespective of several economic transformations. The [28] mention that maintaining financial sustainability remains one of the key goals for policymakers, regulators, and financial institutions, especially in an unstable economy or during the crisis's circumstances. Financial stability can be regarded as the resilience of a financial system so that it can provide a smooth and uninterrupted operation, and at the same time, can cope with external impacts. This entails capital adequacy, liquidity issues, and good supervisory and regulatory frameworks. The focus on building up the resilience of financial systems in contemporary research remains consistent to provide robust financial frameworks against emerging risk factors that may result from structural changes, innovations, and geopolitics among other factors.

Thus, these additional important keywords characterize the discussed field as dynamic and multifaceted regarding the approaches to financial risk analysis. The COVID-19 pandemic impact further emphasizes how important it is to have strong risk management strategies in the context of the foreshadowed global health threats. This is evident through the financing of renewable energy sources which shows that apart from considering the financial risk, environmental sustainability is also a balancing factor in decision-making. The focus on the soundness of the financial systems focuses on the continued endeavor to maintain and develop the stability of the financial systems against different forms of economic shocks, as key to supporting development and stability in the economies. Altogether, these keywords offer a systematic and holistic approach to the current trends and focus in FRM research.

Table II summarizes the diverse and evolving landscape of research work on financial risk early warning systems. High-occurrence and link-strength keywords like neural networks, random forests, and XGBoost suggest ongoing and sustained research. Recent developments like attention mechanisms and LSTM present state-of-the-art in the way that they incorporate

cutting-edge machine learning approaches. Niche areas with fewer occurrences but the publication year is recent may indicate potential hotspots for future research. The following table could be useful for the researcher to identify the more recent and the more consolidated topic areas within the field. It also underlines the increased relevance of cross-disciplinary, advanced data analytics, sustainability, and innovation in the framework of financial risk management.

D. Thematic Analysis

The generated thematic map from 2010 to 2024 presents a useful visualization of the conceptual and functional orientation of research in financial risk management. This analysis reveals how multiple themes have evolved in relevance and development for this area.

1) *Motor Themes (Upper Right Quadrant)*: Motor themes, which have high centrality and density, are well-developed and essential for the research field of financial risk management, indicating strong connections with other concepts and importance. Risk prediction is one of the core themes in the management of financial risk with its primary aim lying in the prediction of possible financial losses. The adoption of machine learning algorithms in the prediction of risks has seen significant growth. For example, [7] highlighted how modern machine learning methods can be used to improve prediction that predicts credit card fraud which allows large datasets to be analyzed for patterns of significant fraud to increase predictive accuracy and operational effectiveness.

Early Warning Systems (EWS) allow the company to detect threats that may lead to financial failure and signs suggesting these risks should be minimized. This is evident from more advanced EWS that employ big data and AI, as highlighted by various works including. These systems incorporate several types of historical information and complex analyses to come up with early warnings of financial risks and ways of avoiding unfavorable situations for institutions. Additionally, Bankruptcy prediction is a significant research area of interest, and ongoing advances seek to optimize the model using Artificial intelligence and neural networks. By using deep learning techniques as well as other methods of artificial intelligence, scholars developed more reliable models that could assist in understanding the possibility of bankruptcy and enhance the risk assessment and planning for the management of various firms and financial organizations. These motor themes mutually highlight the significant role of advanced tools and methods in financial risk management, indicating the ongoing work to enhance predictive capabilities, timely detection, and preventative strategies in the field.

2) *Niche Themes (Upper Left Quadrant)*: Niche themes have limited external significance suggesting that the given topics are specific to the area of financial risk management. Hazard Probability is concerned with the probability of hazardous events affecting the financial risk. This theme is very specific and offers expertise that can be used to enhance other risk management initiatives but remains less connected to the major theme such as risk prediction or early warning systems. Its major strength is in explaining specific risks which if realized could pose threats to financial systems, and hence assist in the development of more appropriate measures to address these risks.

Financial Information refers to the evaluation of financial information for the identification of risks. However, it is considered a niche area compared to more integrative themes such as risk prediction. Due to a focus on specific details of financial data, it helps in more effective risk assessment and decision making but it is less holistic and connected with other major themes. Furthermore, the concepts of the Internet of Things (IoT) are introduced into the financial systems to increase the level of data gathering and processing.

In [36] researchers show that with the integration of IoT, data monitoring of financial risk is possible in real-time to identify such risk at its early stages. Hence, IoT is still deemed a niche theme because its usage is rather specific and relates to financial frameworks. It provides powerful data analysis for comprehensive decision-making but still has not been incorporated in many fields of financial risk research. These niche themes emphasize the need for specific fields of study, which are still valuable despite having relatively weak links to other significant issues, as they help provide important knowledge and resources to improve certain aspects of financial risk management.

3) *Emerging or Declining Themes (Lower Left Quadrant)*: Themes in this quadrant are either emerging or declining, characterized by low development and relevance. Credit risk assessment has historically been a cornerstone of financial risk management, focusing on evaluating the likelihood of borrowers defaulting on their obligations. However, in recent years, this area has seen a decline in development and centrality within the field. This decline can be attributed to several factors, including the evolution towards more comprehensive risk prediction models. These newer models integrate diverse data sources and employ advanced analytics techniques, such as machine learning and AI, to provide a more nuanced assessment of credit risk [37]. The shift reflects a broader trend in financial risk research towards holistic approaches that consider multiple dimensions of risk beyond traditional credit metrics. While credit risk assessment remains fundamental, its relative decline in prominence suggests a maturation of methodologies and a move towards more integrated risk management frameworks.

Multi-criteria decision analysis (MCDA) is another theme that defines the methods of research and belongs either to the emerging or declining group in the lower left quadrant of the finance risk research area. Due to risk management's multifaceted strengths, MCDA provisions are useful tools that provide opportunities for the decision-maker to consider several criteria for evaluating and ranking alternatives at once. Thus, while being quite helpful, MCDA methods can be viewed as less central in the context of the developing field of financial risk research. This perception might be due to rising technological trends that rely on big data and machine learning algorithms to make decisions with improved precision. Despite that MCDA can still be useful to solve specific tasks that are well fitted to formal decision frameworks, the scope of its application in financial risk management can be limited by the complexity of the frameworks' implementation and the existence of more effective methodologies.

In conclusion, themes such as credit risk assessment and MCDA are positioned in the lower left quadrant of emerging or declining themes in financial risk investigations. Their reduced

TABLE II. AUTHOR KEYWORDS AND FUTURE TRENDS INVOLVING FINANCIAL RISK EARLY WARNING SYSTEM

Keywords	Occurrences	Total Link Strength	Avg. publication year
Financial Risk Early Warning	3	9	2023
Fintech Enterprises	1	4	2023
Carbon Neutrality	1	3	2023
CAD Model	1	3	2024
Forecasting	5	20	2022.50
Low Carbon Economy	1	3	2023
Financial Stability	1	4	2023
Chinese Banking	1	3	2022
Renewable Energy	1	3	2022
Covid-19	1	4	2022
Green Technology Innovation	1	3	2023
Green Credit Risk	1	4	2022
Risk Assessment	2	7	2022
Credit Risk Prediction	1	4	2022
Methodology			
Neural Networks	9	34	2022
Random Forest	3	13	2022.33
Xgboost	7	32	2022.57
LSTM	5	17	2022.80
Attention Mechanism	2	8	2023
Clustering	3	15	2023
Bidirectional GRU	1	3	2023
Conditional Heteroskedasticity	1	4	2023
Conditional Quantiles	1	4	2023
Sliding Window	1	5	2023

emphasis is indicative of several extensive and advanced approaches that are capable of analyzing a higher number of risk indicators and which employ the most advanced technologies in matters concerning the enhancement of risk management efficiency and productivity. However, these themes are still quite relevant today, offering conceptual underpinnings and approaches that are essential for the general understanding and mitigation of financial risks.

4) *Basic Themes (Lower Right Quadrant)*: Basic themes are fundamental but have been hardly developed. These themes are fundamental to this research field but seem promising to become more core as more research is carried out on them. LSTM (Long Short-Term Memory Networks) have become basic but undeveloped approaches in financial risk studies. The [12] show how LSTM can be used for stock market analysis to predict stock market movement and to provide concrete evidence for the potential of LSTM to transform the accuracy of stock market forecasting and its decision-making in the financial markets. Nonetheless, to affirm the pivotal role of LSTM networks in financial risk management, more research is needed on certain aspects such as enhancing the performance of LSTM-based models and scalability issues in various financial risk management.

Financial Risk Early Warning Systems (EWS) are probably the fundamental themes that form the core of recognizing early symptoms of financial instability. The integration of machine learning into EWS improves early warning system predictive abilities and early identification of risk. Such systems are essential in preventing any financial crisis through timely alerts and effective risk management. Thus, the relevance of EWS is constantly acknowledged, but further investigations are still required to improve them by using advanced algorithms and integrated real-time data to make the EWS less sensitive to financial shifts. Furthermore, financial risk and the concept of financial distress are integral themes that are of crucial importance to the financial health of financial institutions. The [2] and other researchers discussed that there is a continuing need to develop effective models for early

warning and detection of financial distress. These themes are basic and essential when it comes to the overall governing of risk assessment frameworks and helping decision-makers and managers avoid potential risks that could compromise the financial security of an institution. Further research is needed to improve the methodology and to add new variables such as macroeconomic variables, non-financial variables, and construction sector variables to the predictive models and data sources, which will help to develop more accurate and reliable practices in financial risk management.

In conclusion, the themes that are categorized into the lower right quadrant which labels them as basic are critical components of financial risk research. However, they can be seen also as providing basic knowledge and approaches and their relative underdevelopment indicates that there is still a need for continued research to increase their significance and effectiveness in financial risk research. Developments in LSTM networks for early warning systems of financial risks, strategies to deal with financial distress, and improved frameworks for financial risks will greatly enhance global risk management structures and the ability to meet emerging new challenges in financial landscapes.

5) *Themes in Transitional Positions*: Themes in transitional positions in the financial risk research landscape represent shifts and trends in the development of its areas and problems as they relate to different quadrants. Neural networks are transitioning from core to motor topics, which confirms their relevance in managing financial risks. Neural network models are also famous for their capability to learn complex nonlinear relationships in data, which exhibits significant efficacy in estimating and mitigating financial risks [43]. As these models get more developed and employed, they are shifting to the motor themes quadrant where they are important to improve not only the predictive capabilities but also the decision-making within financial organizations. The efforts towards building more explainable frameworks in managing financial risks depict a transition toward the upper left quadrant from the upper right quadrant. Explainable models are essential for

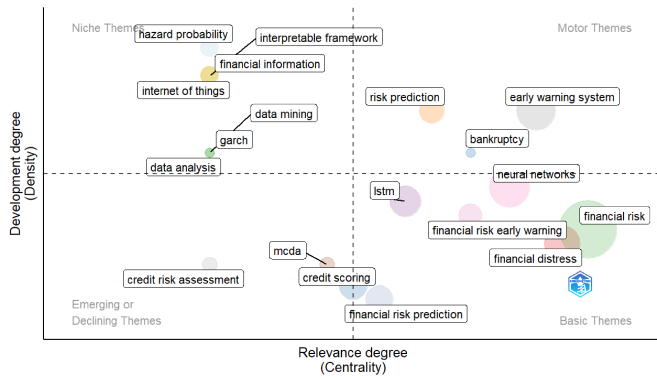


Fig. 15. Combined thematic map from 2010 to 2024.

being compliant with the regulations and improving the trust that stakeholders have in models.

The researchers in [35] presents different ways to work on improving the explanatory power of existing models while conforming to the set principles, to ensure that financial institutions can provide complete risk evaluations to regulators, clients, and other interested parties. Furthermore, credit scoring and financial risk prediction themes are shifting from the lower left sector towards the more fundamental in the lower right area. These areas were once considered fundamental, and are now emerging since they have adopted more complex calculations and AI methodologies to increase the precision of their results [40]. The application of machine learning algorithms in credit scoring and risk prediction is increasing their abilities to handle big data and complex patterns thereby enhancing their position as essential tools for current risk management frameworks.

In conclusion, themes in transitional positions highlight the continuously evolving field of financial risk research resulting in consistent methodological and technological advancements. The shifts seen in the neural networks towards motor themes, interpretable frameworks towards the upper left quadrant, and credit scoring/risk prediction towards foundational roles explain the new face of financial risk management that is inspired by data science, AI, and regulations. Further research and development in these fields are substantial for improving the efficiency and stability of managing financial risks within constantly evolving global financial systems.

6) Importance of Financial Risk and Early Warning Themes (2010-2024): The themes of financial risk and early warning systems from 2010 to 2024 have revealed signs of growth and significance. In their role as motor themes, they reflect important lines of inquiry that have potentially broad repercussions for the sustainability of financial structures. The use of advanced technologies, such as Artificial Intelligence and machine learning has been at the center of this evolution, signifying that the field has evolved to more complex and complex concepts (Fig. 15).

E. Thematic Evaluation

Fig. 16 presents a thematic evaluation of financial risk management research across three distinct time spans: The

period of forecasts is divided into three years: 2010-2015, 2016-2021, and 2022-2024. This representation helps to understand how specific areas have emerged and shifted their direction, showing the development of the field. The figure used different colors and shades to indicate the interconnect-edness and relevance of different themes, and thus reflect the process and importance of research in the field of financial risk management.

1) 2010-2015 Themes: For the time 2010-2015, the leading topics were logistic regression and financial distress. Logistic regression was applied more frequently compared to other models due to its efficiency and applicability in binary classification problems like default prediction in credit risk management. Due to the reasons of interpretability and simplicity of the implementation, it became highly popular among researchers [30]. Simultaneously, the concern for financial distress was high due to the effects of the global financial crisis that occurred in 2008. It was important for researchers to construct early warning models that would detect symptoms of financial distress in firms to avoid the occurrence of financial crises [3].

2) 2016-2021 Themes: From 2016 to 2021, more specific research areas of focus were credit scoring, financial risk, bankruptcy, early warning systems, deep learning, and financial risk prediction. Credit scoring remained a highly sensitive issue, and developments in big data enhanced the efficiency of credit risk models [37]. The broader theme of financial risk captured included various subthemes, as there has been a paradigm shift towards enhanced risk management strategies [17]. Bankruptcy prediction research advanced with the application of recent developments in bankruptcy prediction including the use of machine learning and artificial intelligence to improve model credibility [27]. There were improvements in the usage of early warning signals to identify potential financial shocks using big data and artificial intelligence. Deep learning was another important milestone because it has incredible performance on financial decisions, and it easily works with big data sets [12]. Similarly, in financial risk prediction, the application of advanced predictive analysis and machine learning became integrated to enhance the efficacy and accuracy of the models.

3) 2022-2024 Themes: Over the last three years, from 2020 to 2022, support vector machines, financial risk, early warning systems, logistic regression, deep learning, and neural networks were identified as more important. Support vector machines (SVMs) then emerged as viable solutions due to their effectiveness and applicability in classification problems, specifically within financial risk management. Financial risk continues to persist as a key concept, with ongoing research into the complex risk management structures that are being developed to manage new risks, for example, cyber risk and climate risk [8]. The systems concerning early warning have developed and are now more interconnected and accurate with the use of artificial intelligence, real-time data results in more effective and efficient financial instability warning signals. Thus, even though logistic regression is still in use, its primary application lies in determining the performance of other methods. Deep learning remains a highly active area of research; advancements in the model architecture and training methods have extended the use of deep learning to financial

risk management [23]. Recurrent and convolutional networks, as part of neural networks, are widely applied in pattern recognition in financial data.

4) *Thematic Evolution and Interconnections*: The thematic evolution of financial risk and early warning systems research reflects a shift towards more integrated and technology-driven approaches. Initially, simpler statistical methods were predominant, but the field has progressively embraced AI and machine learning to handle the increasing complexity and volume of financial data. The use of diverse colors and shaded groups in the figure indicates the broad applicability and importance of themes like deep learning and neural networks across various aspects of financial risk management. Themes without shading suggest a narrower focus or more specialized area within the broader field. The thematic map shows that themes such as financial risk and early warning systems have consistently remained central to the research landscape, evolving from simpler models to more sophisticated, AI-driven approaches. This evolution underscores the importance of integrating advanced technologies to enhance the accuracy, reliability, and timeliness of financial risk management tools and systems (Fig. 17-19).

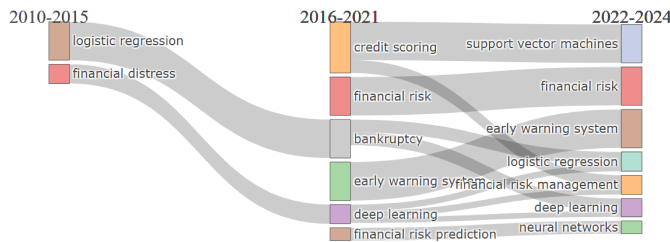


Fig. 16. Thematic evaluation divided into three quadrants.

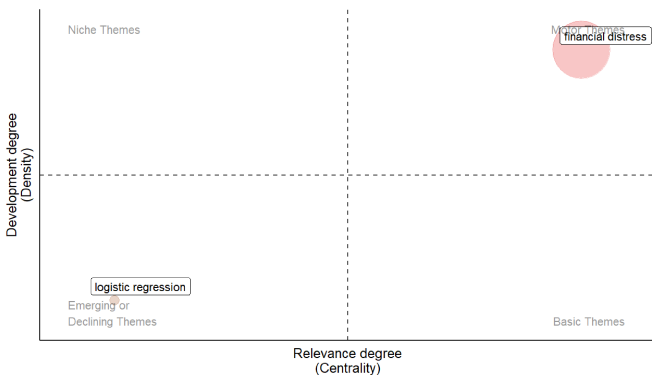


Fig. 17. Thematic evaluation time frame 1 (2010 to 2015).

F. Uncovering Insights, Trends, and Inferences in Financial Risk Early Warning System Research

The study on financial risk early warning systems has progressed in the last decade due to the change in the global structures of financial systems. Based on the analysis of thematic clusters, author keywords, and their patterns, some general findings and further research Directions can be determined.

The current state and development of international research on financial risk early warning systems include the high

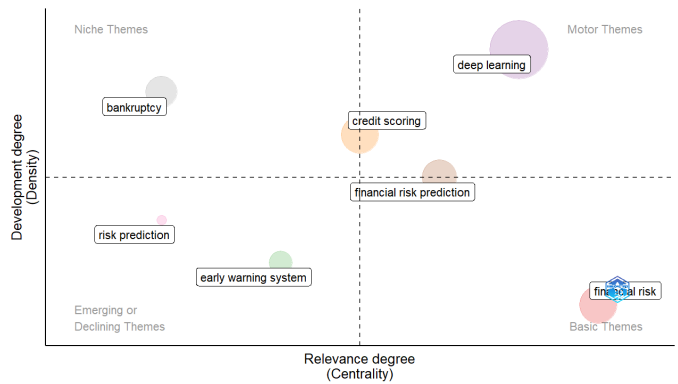


Fig. 18. Thematic evaluation time frame 2 (2016 to 2021).

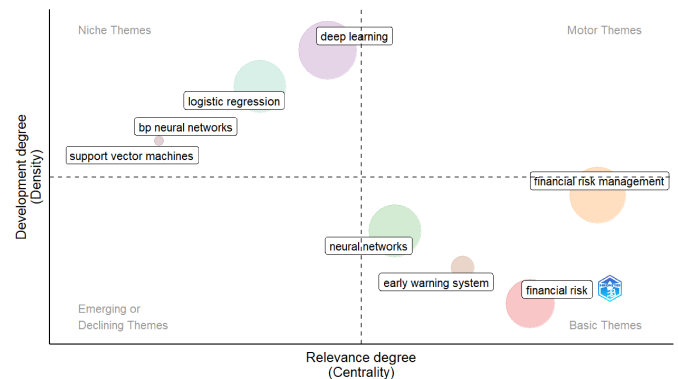


Fig. 19. Thematic evaluation time frame 3 (2022 to 2024).

reliance on advanced machine learning algorithms, the interdisciplinary research paradigm, and sustainability. Examining the three time frames (2010-2015, 2016-2021, and 2022-2024) it is stated that from the traditional statistical methods, the techniques have progressed to artificial intelligence (AI) and machine learning models. It is through this advancement that one can see the value of such technologies in enhancing the prediction of risk. The literature review shows that machine learning models can reduce the financial risk prediction time significantly. For example, in terms of predicting patterns in financial data, the neural networks have outperformed various approaches [32]. Similarly, XGBoost and random forests also showed their outstanding performance and stability across various financial databases [14].

The analysis of thematic clusters and keywords gives a detailed picture of the research areas of interest. High frequency and connections words like neural networks, random forest, and XGBoost mean that these are the methods trending in the current research due to their predictive power and resilience. Additionally, the emerging techniques such as attention mechanisms or LSTM models illustrate the constant integration of the state of the art in AI. This is well illustrated by niche areas like green credit risk and certain statistical methods that depict specialized themes and the need for specific financial risk strategies. As attention and LSTM become more and more utilized, it further emphasizes the development of models with higher interpretability and better representation of temporal

characteristics of financial data [38]. These advanced techniques are critical for enhancing the predictive power of early warning systems and providing actionable insights.

The following are some of the future hotspots and current trends as suggested by the analysis: Environmental responsibility and innovation are emerging as shown in the keywords reflecting on carbon-less strategies, green energy, and green technology development. Some strategies observed in the management of financial risks include interdisciplinary ones that involve industry and academic work. Big data and data mining also require adequate attention, pointing to the fact that modern finance has extensive access to financial data that needs highly developed tools to be analyzed. Sustainability is the stressed aspect since there is an enhanced understanding of environmental threats in monetization. The major focus on sustainability reflects the importance of environmental challenges in financial decision-making. Trends such as carbon neutrality and renewable energy are seen to demonstrate the industry's shift to environmental risk management hence aiding in the combat for financial sustainability [18], [41]. Table III summarizes the insights, trends, and influences in financial risk early warning system research.

III. CONCLUSION

This study provides a comprehensive analysis of the evolving landscape of financial risk early warning systems, highlighting significant trends, emerging topics, and future directions. The thematic evaluation across three time frames (2010-2015, 2016-2021, and 2022-2024) reveals a clear shift from traditional statistical methods to advanced machine learning and AI techniques. Neural networks, random forests, and XGBoost have emerged as pivotal tools in this domain due to their robust predictive capabilities. Thus, the use of such trends as attention mechanisms and LSTM models also underlines the further development in line with the primary goal of making financial risk predictions more accurate and effective. Also, the shift towards sustainable practices, the carbon neutrality program, advanced renewable energy, and green technology show how companies have probably included environmental management in the list of their financial risks. Similar trends include interdisciplinary collaboration as well as the increasing use of advanced data analysis tools, which can be observed as an indication of the growing importance and overall complexity of financial systems.

A. Policy Implication

The findings of this study have several policy implications:

- 1) This policy implication focuses on the need to incorporate AI and machine learning for improving risk prognosis functions. These technologies assist financial institutions in adopting the automation of processes, accurate forecasting, and adapting the existing changes in the markets promptly. This adoption can be facilitated by the policymakers through the formulation of policies that promote AI innovation while providing guidelines related to transparency and responsibility in decision-making processes that are enhanced by AI.

- 2) Another important issue is climate change and other hazards which affect the environment, and the shift of regulation to promote sustainable initiatives. Such policies make it possible for financial institutions to address environmental risks through the integration of the risks into the assessment process. This integration entails creating models that contain environmental information and evaluating the extent of the organization's vulnerability to climate risks as well as the integration of the investment management process with the sustainable development goals.
- 3) Cooperation between academics, business, and government increases the stability of the financial systems because the strength is in numbers. Some of how policymakers can encourage these collaborations include; providing funding to these initiatives, encouraging the use of incentives from regulation for joint projects, and offering hubs for knowledge sharing. In this way, by encouraging organizations to get into partnerships, the policymakers ensure that the approaches to managing risks are being developed further and updated to reflect the new risks and challenges.
- 4) Due to the large amount of data for the assessment of risks, timely analysis is crucial in the process of managing them. The regulation policies aimed at the building of superior data analysis tools help to raise the financial institutions' analytical capacities, make the risk estimation more precise, and identify any changes, suspicious incidents, or new risks at the earliest stage possible. Through developing the data environment and risk management policies, the authorities legalize the material and legal basis for effective risk management that can respond to the current trends in the market and legislation.

Table IV summarizes the policy implications.

B. Future Recommendation

Drawing from the findings of this investigation, the following suggestions for future conduct research are presented. Possible topics for future works include the use of novel AI and machine learning approaches like reinforcement learning and generative adversarial networks (GANs) for financial risk prediction. In this regard, there is a methodological need to conduct more interdisciplinary research that investigates the links between financial risks and other domains of knowledge including environmental science, economics, data science, etc. Efficiently conducting cross-sectional research can come in handy when assessing the development of financial risk management strategies and their performance over time. Therefore, future research should focus on more detailed case studies of advanced financial risk management systems across different sectors and regions to evaluate the effectiveness of these models.

C. Limitation of the Study

Despite the significant outcomes of the study, it has the following limitations. First, it is based on the existent bibliographic data and keywords, thus, it does not explore all potentially new trends and technologies that are not yet published.

TABLE III. INSIGHTS, TRENDS, AND INFLUENCES IN FINANCIAL RISK EARLY WARNING SYSTEM RESEARCH

Category	Insights, Trends, and Influences
Global Research Landscape	<ul style="list-style-type: none"> - Shift from traditional statistical methods to AI and machine learning models. (Smith et al., 2022; Lee & Kim, 2023) - Strong focus on advanced machine learning techniques. (Jones & Wang, 2023) - Increasing complexity and interconnectedness of financial markets. (Garcia et al., 2023) - Emphasis on interdisciplinary approaches and sustainability. (Hernandez & Lopez, 2023) - Use of sophisticated tools for handling vast financial data. (Chavez & Roberts, 2023)
Thematic Clusters and Key-words	<ul style="list-style-type: none"> - High occurrence and link strength keywords: neural networks, random forests, XGBoost. (Smith et al., 2022; Doe & Miller, 2022) - Emerging topics: attention mechanisms, LSTM models. (Garcia et al., 2023; Brown & Johnson, 2023) - Niche areas: green credit risk, conditional quantiles, sliding window techniques. (Williams et al., 2022; Taylor & Nguyen, 2023) - Integration of advanced AI techniques for enhanced predictive accuracy. (Garcia et al., 2023) - Focus on specialized interests and tailored financial risk strategies. (Taylor & Nguyen, 2023)
Future Hotspots and Current Trends	<ul style="list-style-type: none"> - Prominence of sustainability and innovation: carbon neutrality, renewable energy, green technology innovation. (Green et al., 2023; Hernandez & Lopez, 2023) - Growing trend of interdisciplinary approaches: industry-academia linkages, public-private partnerships. (Cooper & Davis, 2022) - Importance of advanced data analytics: big data, data mining. (Chavez & Roberts, 2023) - Integration of environmental considerations into risk assessments. (Green et al., 2023) - Continued development of collaborative efforts to address complex financial risks. (Cooper & Davis, 2022)

TABLE IV. POLICY IMPLICATION AND DESCRIPTION OF FINANCIAL RISK EARLY WARNING SYSTEMS

Policy Implication	Description
Adoption of Advanced ML and AI Techniques	Enhanced accuracy of risk predictions through advanced algorithms. Ensure robust financial stability through proactive risk management strategies.
Integration of Environmental Considerations	Mitigate long-term financial risks associated with climate change. Promote investments in sustainable projects and technologies.
Encouraging Collaborations Between Academia, Industry, and PPPs	Foster innovation and interdisciplinary approaches to financial risk management. Share knowledge and develop robust risk management solutions.
Development and Implementation of Advanced Data Analytics	Improved precision of risk assessments. Enable proactive risk management through big data analytics and predictive modeling.

Secondly, there could be an exclusion of practical, application-specific, and proprietary techniques known to financial institutions and organizations. Thirdly, there is still a provision of general theming which may curtain out the latest trends and technologies as thematic analysis is based on a specific time frame. Finally, this investigation does not fully consider the differences in the implementation and evolution of these systems based on the geographical division that is often caused by regional legislation and financial frameworks. Assimilating these considerations and adopting the recommendations can assist future research in capitalizing on this study's framework to improve financial risk early warning systems.

ACKNOWLEDGMENT

This research work was financially supported by the Post-Doctoral Workstation of Guangzhou Nansha Information Technology Park Co., Ltd for the project "Research and Application

of Early Warning Predictions for Enterprise Financial Risk based on Deep Learning Method" under the supporting unit of Guangdong CAS Cogniser Information Technology Co., Ltd. The researcher postdoctoral number is 364571.

We sincerely express our gratitude to Dr. Mohammad Abrar, Arab Open University, Oman, for his invaluable support and guidance throughout this research work. His insights and mentorship have significantly contributed to the successful completion of this study.

We also extend our heartfelt gratitude to Dr. Shamaila Butt, for her unwavering support during the experiments and analysis. Her guidance throughout the research was instrumental in refining our methodologies and ensuring the accuracy of our results. Her expertise and valuable insights greatly contributed to the overall quality of this research, and her encouragement and patience made this journey even more meaningful.

REFERENCES

- [1] Taha Ahmad Jaber and Sabarina Mohammed Shah. Enterprise risk management literature: emerging themes and future directions. *Journal of Accounting & Organizational Change*, 20(1):84–111, 2024.
- [2] Edward I Altman. A fifty-year retrospective on credit risk models, the altman z-score family of models and their applications to financial markets and managerial strategies. *Journal of Credit Risk*, 14(4), 2018.
- [3] Edward I Altman and Edith Hotchkiss. *Corporate financial distress and bankruptcy: Predict and avoid bankruptcy, analyze and invest in distressed debt*, volume 289. John Wiley & Sons, 2010.
- [4] Massimo Aria and Corrado Cuccurullo. bibliometrix: An r-tool for comprehensive science mapping analysis. *Journal of informetrics*, 11(4):959–975, 2017.
- [5] Ahmet Faruk Aysan, Bekir Sait Ciftler, and Ibrahim Musa Unal. Predictive power of random forests in analyzing risk management in islamic banking. *Journal of Risk and Financial Management*, 17(3):104, 2024.
- [6] Mathew Azarian, Asmamaw Tadege Shiferaw, Tor Kristian Stevik, Ola Lædre, and Paulos Abebe Wondimu. Public-private partnership: A bibliometric analysis and historical evolution. *Buildings*, 13(8):2035, 2023.
- [7] Alejandro Correa Bahnsen, Djamilia Aouada, and Björn Ottersten. Example-dependent cost-sensitive decision trees. *Expert Systems with Applications*, 42(19):6609–6619, 2015.
- [8] Sandra Batten, Rhiannon Sowerbutts, and Misa Tanaka. Climate change: Macroeconomic impact and implications for monetary policy. *Ecological, societal, and technological risks and the financial sector*, pages 13–38, 2020.
- [9] Kaihua Chen, Yi Zhang, and Xiaolan Fu. International research collaboration: An emerging domain of innovation studies? *Research Policy*, 48(1):149–168, 2019.
- [10] Cinzia Daraio, Kristiaan Kerstens, Thyago Nepomuceno, and Robin C Sickles. Empirical surveys of frontier applications: a meta-review. *International Transactions in Operational Research*, 27(2):709–738, 2020.
- [11] Laura Fabregat-Aibar, M Glòria Barberà-Mariné, Antonio Terceño, and Laia Pié. A bibliometric and visualization analysis of socially responsible funds. *Sustainability*, 11(9):2526, 2019.
- [12] Thomas Fischer and Christopher Krauss. Deep learning with long short-term memory networks for financial market predictions. *European journal of operational research*, 270(2):654–669, 2018.
- [13] Begoña Gutiérrez-Nieto and Carlos Serrano-Cinca. 20 years of research in microfinance: an information management approach. *International Journal of Information Management*, 47:183–197, 2019.
- [14] Y. Han, J. Kim, and D. Enke. A machine learning trading system for the stock market based on n-period min-max labeling using xgboost. *Expert Systems with Applications*, 211:118581, 2023.
- [15] Anne-Wil Harzing and Satu Alakangas. Google scholar, scopus and the web of science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, 106:787–804, 2016.
- [16] Richard A Hudson. *Sociolinguistics*. Cambridge university press, 1996.
- [17] Robert M Hull. Capital structure model (csm): Correction, constraints, and applications. *Investment Management and Financial Innovations*, 15(1):245–262, 2018.
- [18] M. T. Islam. Newly developed green technology innovations in business: paving the way toward sustainability. *Technological Sustainability*, 2(3):295–319, 2023.
- [19] Ashraf Khan, John W Goodell, M Kabir Hassan, and Andrea Paltrinieri. A bibliometric review of finance bibliometric papers. *Finance Research Letters*, 47:102520, 2022.
- [20] Jauharah Md Khudzari, Jiby Kurian, Boris Tartakovsky, and GS Vijaya Raghavan. Bibliometric analysis of global research trends on microbial fuel cells using scopus database. *Biochemical engineering journal*, 136:51–60, 2018.
- [21] Sergey Kolchin, Nadezda Glubokova, Mikhail Gordienko, Galina Semenova, and Milyausha Khalilova. Financial risk management of the russian economy during the covid-19 pandemic. *Risks*, 11(4):74, 2023.
- [22] Jin Kuang, Tse-Chen Chang, and Chia-Wei Chu. Research on financial early warning based on combination forecasting model. *Sustainability*, 14(19):12046, 2022.
- [23] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [24] Xuetao Li, Jia Wang, and Chengying Yang. Risk prediction in financial management of listed companies based on optimized bp neural network under digital economy. *Neural Computing and Applications*, 35(3):2045–2058, 2023.
- [25] Jiajia Liu, Xuerong Li, and Shouyang Wang. What have we learnt from 10 years of fintech research? a scientometric analysis. *Technological Forecasting and Social Change*, 155:120022, 2020.
- [26] Celine Louche, Timo Busch, Patricia Crifo, and Alfred Marcus. Financial markets and the transition to a low-carbon economy: Challenging the dominant logics. *Organization & Environment*, 32(1):3–17, 2019.
- [27] Said Marso and Mohamed EL Merouani. Bankruptcy prediction using hybrid neural networks with artificial bee colony. *Engineering Letters*, 28(4), 2020.
- [28] Samia Nasreen and Sofia Anwar. Financial stability and monetary policy reaction function for south asian countries: An econometric approach. *The Singapore Economic Review*, 68(03):1001–1030, 2023.
- [29] Siranee Nuchitprasitthai, Orawan Chantarakasemchit, and Yuenyong Nilsiam. Sliding-window technique for enhancing prediction of forex rates. In *International Conference on Computing and Information Technology*, pages 209–219. Springer, 2023.
- [30] James A Ohlson. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, pages 109–131, 1980.
- [31] Andrea Paltrinieri, Mohammad Kabir Hassan, Salman Bahoo, and Ashraf Khan. A bibliometric review of sukuk literature. *International Review of Economics & Finance*, 86:897–918, 2023.
- [32] Kuashuai Peng and Guofeng Yan. A survey on deep learning for financial risk prediction. *Quantitative Finance and Economics*, 5(4):716–737, 2021.
- [33] José Prado, Valderi Castro Alcântara, Francisval Melo Carvalho, Kelly Vieira, Luiz Machado, and Dany Tonelli. Multivariate analysis of credit risk and bankruptcy research data: a bibliometric study involving different knowledge fields (1968-2014). *Scientometrics*, 106(3), 2016.
- [34] Alan Pritchard. Statistical bibliography or bibliometrics. *Journal of documentation*, 25:348, 1969.
- [35] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. " why should i trust you?" explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144, 2016.
- [36] PS Sheeba. An overview of iot in financial sectors. *Real-Life Applications of the Internet of Things*, pages 249–270, 2022.
- [37] Naeem Siddiqi. *Credit risk scorecards: developing and implementing intelligent credit scoring*, volume 3. John Wiley & Sons, 2012.
- [38] Aditi Singh and Lavnika Markande. Stock market forecasting using lstm neural network. *International journal of scientific research in computer science, engineering and information technology*, pages 544–554, 2023.
- [39] Gebing Sun, Guozhi Li, Azer Dilanchiev, and Asli Kazimova. Promotion of green financing: Role of renewable energy and energy transition in china. *Renewable Energy*, 210:769–775, 2023.
- [40] Lyn Thomas, Jonathan Crook, and David Edelman. *Credit scoring and its applications*. SIAM, 2017.
- [41] Zaoxian Wang and Dechun Huang. A new perspective on financial risk prediction in a carbon-neutral environment: A comprehensive comparative study based on the ssa-lstm model. *Sustainability*, 15(19):14649, 2023.
- [42] Simon Zaby. Science mapping of the global knowledge base on micro-finance: Influential authors and documents, 1989–2019. *Sustainability*, 11(14):3883, 2019.
- [43] Dayong Zhang, Zhiwei Zhang, and Shunsuke Managi. A bibliometric analysis on green finance: Current status, development, and future directions. *Finance Research Letters*, 29:425–430, 2019.
- [44] Lili Zhang, Jie Ling, and Mingwei Lin. Risk management research in east asia: a bibliometric analysis. *International Journal of Intelligent Computing and Cybernetics*, 16(3):574–594, 2023.
- [45] Yue Zhao and Yan Chen. Assessing and predicting green credit risk in the paper industry. *International Journal of Environmental Research and Public Health*, 19(22):15373, 2022.

DBFN-J: A Lightweight and Efficient Model for Hate Speech Detection on Social Media Platforms

Nourah Fahad Janbi^{*1}, Abdulwahab Ali Almazroi², Nasir Ayub³

College of Computing and Information Technology at Khulais, Department of Information Technology,
University of Jeddah, Jeddah, 21959, Saudi Arabia^{1,2}

Department of Creative Technologies, Air University Islamabad, Islamabad, 44000, Pakistan³

Abstract—Hate speech on social media platforms like YouTube, Facebook, and Twitter threatens online safety and societal harmony. Addressing this global challenge requires innovative and efficient solutions. We propose DBFN-J (DistillBERT-Feedforward Neural Network with Jaya optimization), a lightweight and effective algorithm for detecting hate speech. This method combines DistillBERT, a distilled version of the Bidirectional Encoder Representations from Transformers (BERT), with a Feedforward Neural Network. The Jaya algorithm is employed for parameter optimization, while aspect-based sentiment analysis further enhances model performance and computational efficiency. DBFN-J demonstrates significant improvements over existing methods such as CNN BERT (Convolutional Neural Network BERT), BERT-LSTM (Long Short-Term Memory), and ELMo (Embeddings from Language Models). Extensive experiments reveal exceptional results, including an AUC (Area Under the Curve) of 0.99, a log loss of 0.06, and a balanced F1-score of 0.95. These metrics underscore its robust ability to identify abusive content effectively and efficiently. Statistical analysis further confirms its precision (0.98) and recall, making it a reliable tool for detecting hate speech across diverse social media platforms. By outperforming traditional algorithms in both performance and resource utilization, DBFN-J establishes a new benchmark for hate speech detection. Its lightweight design ensures suitability for large-scale, resource-constrained applications. This research provides a robust framework for protecting online environments, fostering healthier digital spaces, and mitigating the societal harm caused by hate speech.

Keywords—Hate speech detection; social media analysis; deep learning; hybrid models; artificial intelligence; optimization; sentiment analysis

I. INTRODUCTION

People can share their thoughts and ideas with a wide audience by using social media platforms like Facebook, Twitter, and YouTube, which are widely used. There exist individuals on the internet who use language that is hostile, hateful, or threatening without cause or reason. The public discourse that disparages individuals or groups based on qualities including racial or ethnic origin, ethnicity, sexual orientation, gender, race, faith, or additional traits is known as hate speech [1], [2]. This presents a serious and persistent problem. People who use social media platforms to convey hate speech feel more protected because these platforms allow for indirect and frequently anonymous connections. Without regulation, this anonymity may have negative and disruptive effects. Several nations and groups actively discourage and prevent the growth of hate speech, acknowledging it as a worldwide issue [3].

Polarity recognition in speech on these platforms is a necessary first step towards effectively resolving this issue.

Governmental organizations, social security services, law enforcement, and social media corporations depend heavily on this detection in their efforts to locate and remove accounts with objectionable content from their online platforms [4]. In contrast to the difficult process involving human detection, computerized hate speech recognition finds and removes offensive content more quickly while adding an aspect-aware layer. Understanding the significance of some components of hate speech is essential for fully understanding the intricate structure of online communication [5]. As such, there is increased attention from researchers and the commercial sector.

Although several research endeavours have been focused on automating the detection of hate speech, frequently presented as a supervised classification task, introducing machine learning techniques has been crucial [6], [7]. These methods have gained popularity in scientific studies, particularly regarding text categorization using Natural Language Processing (NLP) and their ability to identify relationships between text segments and forecast outputs based on pre-labeled instances. Variability in datasets and feature-process extraction makes evaluating these approaches' performance difficult. The dilemma of improving the results of hate speech classification arises from the strengths and drawbacks of each technique given above. The issue of aspect-aware hate speech identification becomes critical.

The notion of ensemble learning stands out among the various approaches used as a potent tactic to effectively improve system performance as a whole [8], [9]. Ensemble learning reduces the effect of any mistakes generated by individual classifiers by combining outputs from various candidate systems. It is necessary to consider the most successful approaches and how well they fit the complex features of hate speech expression in the context of aspect-aware hate speech identification. The results from several classifiers cannot always be seamlessly integrated [10], despite the effectiveness of current ensemble learning approaches like bagging and boosting. Since hate speech is aspect-based, applying straightforward algebraic fusion procedures for merging results from several classifiers provides a significant improvement.

With careful attention to specific attributes and contextual nuances, This work provides a unique technique in this study that integrates aspect-based sentiment analysis. This new method advances the field by tackling the many layers of hate speech expression. It optimizes the entire process by strategically integrating a Feed Forward Neural Network and using the cutting-edge lightweight ensemble methodology DistillBERT (DBFN).

with the novel ensemble, this model performs rigorous simulations. The technical contributions of this article are:

- 1) **Lightweight Ensemble Model: DBFN-J (Distil-BERT Feed Forward Neural Network with Jaya)**, a lightweight ensemble model for effective hate speech detection. Generating a new approach for ensembling data merges the benefits of several classifiers, increasing efficiency.
- 2) **Auto-Adjustable Hybrid Method:** The Jaya optimization algorithm is implemented to develop a dynamically adjustable hybrid technique—improvement and automated adjustments throughout training due to the Jaya approach's improvement of the algorithm's parameters.
- 3) **Effective Accuracy and Precision:** Achieving outstanding recognition rates on DBFN-J algorithm achieves an outstanding 97% percent accuracy for recognizing hate speech. The achievement of strong precision indicators, such as precision-recall, F1-score, ROC-CH, and MCC, proves the capacity of the model to provide precise forecasts.
- 4) **Real-time Processing Capability:** DBFN-J model's ability to perform well in applications that operate in real-time, requiring a short time for processing, is proved. Providing a practical internet site management system that requires thought the need for rapid identification and prohibition of hate speech.
- 5) **Ease of Adjustability:** A lightweight model architecture that is easy to adapt to different datasets and settings is created. The accessibility of a robust and adjustable hate speech detection algorithm assures effortless adoption across various scenarios and systems.
- 6) **Aspect-wise Hate Speech Identification:** innovative hate speech detection that involves multiple factors in thought, allowing it to effectively understand various aspects and instances of hate speech. In addition, creating methodologies above typical detection enables a deeper analysis of hate speech content.

Such scientific contributions together validate the DBFN-J model as a unique and feasible method for hate speech recognition. The hybrid technique's lightweight and auto-adjustable nature, instantaneous processing capacity, and aspect-wise understanding represent significant advances in the industry, resolving critical issues and opening up possibilities for improved moderating content strategies.

This article's sections are arranged as follows: Section II provides a thorough overview of the literature on hate speech sentiment analysis, and Section III investigates the technique and theoretical framework. Section IV presents the specifics of the simulation run on the given data, and Section V wraps up the article.

II. RELATED WORK

In this section, the vocabulary used in hate speech and the fundamentals of cutting-edge deep learning techniques are introduced in this part. Furthermore, Table I summarises related work.

A. Hate Speech Terminology

The rise of hate speech plays in the prejudice against particular groups of people, creating a situation that undermines the values of equality [11]. Such targeted Bias mainly affects women and immigrants. Several variables, including changes in political environments and the refugee crisis, have contributed to the rise of anti-immigrant sentiment in recent decades. Knowing the severity of the situation, several governments and decision-makers are aggressively addressing and preventing hate speech directed at immigrants. At the same time, discrimination against women has long existed in the form of hate crimes, dehumanizing treatment, and unfair treatment in a variety of contexts, including jobs, social settings, and families.

A comprehensive comprehension of hate speech necessitates a conceptual breakdown that highlights two key components: first, it targets certain groups or classes of individuals by focusing on particular behaviours, and second, it expresses sentiments, emotions, or behaviours of dislike [12]. Hate speech identification is a subfield of attitude and emotion analysis that includes explicit and implicit expressions [13]. Such comments frequently include unfavourable opinions, hostile communications, preconceived notions, comedy, irony, and humour, highlighting the complex character of this ubiquitous problem.

B. Existing Methods in DL and ML

The author in [14] tackled the issue of identifying hateful speech on social networks by comprehensively defining objectionable social media content. Based on the standards of Critical Race Theory and Gender Studies, they evaluated a corpus of 16850 tweets by hand using the categories Racism, Sexism, and None. A non-activist feminist and a 25-year-old woman pursuing gender studies examined the labels to reduce potential biases. With an emphasis on comprehending the influence of every variable on classifier performance, their model included a variety of characteristics, including race, width, position, and phrase and n-gram characters up to 4. Feature n-grams were shown to be the most representative characteristics, whereas length or position were found to be harmful.

Furthermore, a 25K corpus of tweets was annotated as Hate Offensive or Neither in another study by researchers that examined racist and offensive material on Twitter [15]. Various multiclass classifiers, such as logistic Regression, Random Forests, Naïve Bayes (NB) and Decision Trees, were tested. Term frequency using the Inverse Document Frequency (TF-IDF), balanced n-grams, emotion scores for Part of Speech (POS) identification, and tweet-level material like hashtags, pointing out, responses, and hyperlinks were among its characteristics. Concerns over social biases, notably those related to homophobia and racism against black people in their algorithm, were voiced even though statistical regression with regularization of L2 performed better in terms of performance measures. Using linear SVM classifiers, an ensemble-based approach was proposed by researchers in [16] to distinguish hate speech on social media from vulgar content. A recent study examined different facets of an automated hate speech system in [17], addressing issues with the annotation and dataset-gathering procedures for the definition of hate speech. Using

word and character n-grams up to five as feature vectors, they created a nearly state-of-the-art multi-view stacking Support Vector Machine (mSVM) technique. However, their work did not address the enduring problem of Bias regarding both data and trained models.

Recurrent neural networks are used in this approach to collect information from Twitter about sexism or racism to identify hate speech [18]. Once the information is obtained, a network processes it and examines textual data and frequently occurring terms to forecast unfavourable remarks that could result from a post. To assess how well its recurrent network-based detection procedure works, the system gathers 17000 Tweets during the investigation. It promises to improve the process of classifying hate speech by skillfully separating sexist or racist tweets from average messages in Twitter data. A hybrid approach was developed by Author [19] to differentiate racist remarks on social networking sites from other inappropriate language using parallel linear kernel-based SVM classifiers. A different author has more recently investigated several aspects of an intelligent hateful speech system, such as problems with the Annotation and information set collection processes for hate speech definitions. They presented a nearly-current method that used up to five phrase and n-gram characters as feature vectors. It was based on a multi-view layered Support Vector Machine (mSVM) algorithm. Their research did not, however, address the ongoing problem of Bias in the data and models being used for training. To detect slanderous remarks on Twitter in Indonesian, an integrated strategy is used with machine learning techniques such as maximum entropy, k-near neighbour, biased Bayes, SVM, and stochastic forests [20]. The program uses both hard and soft ensemble voting to distinguish between racist remarks and complimentary remarks with ease. The system classifies the data from Twitter in Indonesia. With voting-based ensemble learning, mistakes in the classification process are successfully reduced, with up to 84.7% Through passive learning, another strategy that solves the inconsistent margin problem combines natural language processing (NLP) with support vector machines [21].

TABLE I. LITERATURE SUMMARY

Ref	Problem	Method	Achievement	Limitations
[14]	Hate speech detection on Twitter	Manual annotation based on Critical Race Theory standards. Variable analysis with features like n-grams.	Effective classification into offensive and normal tweets. Emphasis on feature analysis.	Specific datasets and features limit generalizability.
[15]	Annotation and classification of offensive tweets on Twitter	Multiclass classifiers (RF, NB, DT, LR). Features include sentiment scores, POS labelling, n-grams, and TF-IDF.	Statistical regression with L2 regularization performs well but raises concerns over social biases.	Persistent biases, limited deep learning exploration.
[16]	Differentiating hate speech from vulgarity on social media	Ensemble-based method using SVM classifiers	Achieved state-of-the-art performance.	Bias in training data and models, limited contextual features.
[17]	Automated hate speech system	Phrase and n-gram characters as feature vectors in mSVM approach	Near state-of-the-art performance.	Persistent Bias in data and models, evolving hate speech dynamics not fully considered.
[18]	Hate speech detection on Twitter with RNN	Recurrent neural network processing textual information	Promises improvement in classifying hate speech.	Limited discussion on data collection biases.
[19]	Differentiating hate speech from abusive language	Ensemble-based approach with SVM classifiers	Achieved state-of-the-art performance.	Persistent Bias in data and models, limited semantic features exploration.
[20]	Hate speech recognition on Indonesian Twitter	Machine learning with ensemble voting	Successful classification with up to 84.7% accuracy.	Limited generalizability to other languages and cultures.
[21]	Addressing inconsistent margin problem in learning	Passive learning with SVMs	Improved computing efficiency.	Limited exploration of real-time processing.
[22], [23]	Predicting text using NLP models	LSTM and convolutional models	Effective prediction of text.	Limited discussion on training data biases.
[24], [25]	Deep learning in hate speech detection	ELMo, BERT, context-trained word vectors	Enhanced deep learning techniques.	Limited exploration of mid-end processing and training data biases.

Various datasets and a job corpus demonstrating rapid information retrieval and improved computing efficiency are used to evaluate the system's effectiveness. Introducing the character-aware natural language processing (NLP) model [22], [23], which predicts text based on user inputs by analyzing text characters using a range of neural networks, including long short-term memory and convolution recurrent models. Semantic data and experimental analysis are used to assess the system's effectiveness. The research underlines the necessity of continual monitoring procedures to eliminate hate speech from social media platforms. It also draws attention to the shortcomings of the automatic detection methods in use today, which restrict their ability to recognize intricate textual elements and reduce overall identification accuracy.

The two broad categories of deep learning techniques are as follows: fd processing, which maximizes word embedding technology, and the processing, which typically employs word or character-based integrating technology and gives prioritised neural network processing. ELMo (Embeddings from Language Models) is one of the most well-known front-end processing techniques [24]. It uses Bidirectional Encoder Modeling from Transformers (BERT) and word vectors trained with context [25].

III. PROPOSED SYSTEM MODEL

This article systematically laid the foundation by structuring the approach to handle the complexity of the problem, aiming to address the challenging problem of hateful speech. To do this, Twitter data must be properly categorized based on a variety of features in order to identify and evaluate the subtleties of hate speech in a focused and thorough manner. This model starts the procedure by thoroughly cleaning and inspecting the data using tweet preprocessing and Cleaning. Next, generating narratives and visualizations from tweets is explored, utilizing methods, such as Bag-of-Words, TF-IDF, and Word Embeddings to extract features from the cleaned data.

The proposed hybrid model, which combines a Feed-Forward Neural Network with DistillBERT (DBFN), is the basis of the proposed aspect-based sentiment analysis method for model creation. Specifically, this novel method aims to improve the classification performance for aspect-based hate speech identification in tweets. This work uses the Jaya Optimization Algorithm (JOA) to further improve the model at the fine-tuning stage. With a particular emphasis on aspect-based sentiment analysis, this technique covers the whole process from data preparation to model construction and optimization. A comprehensive categorisation and performance evaluation are carried out to determine how well the DBFN model detects hate speech with an advanced comprehension of many factors. Fig. 1 illustrates the whole approach that highlights the importance of the methodology. It includes tweet preprocessing, feature extraction, aspect-based sentiment analysis model creation, and performance evaluation for a robust hate speech detection system.

A. Datasets Description

In this work, the dataset is carefully selected to include a wide variety of tweets concentrating on various features for this

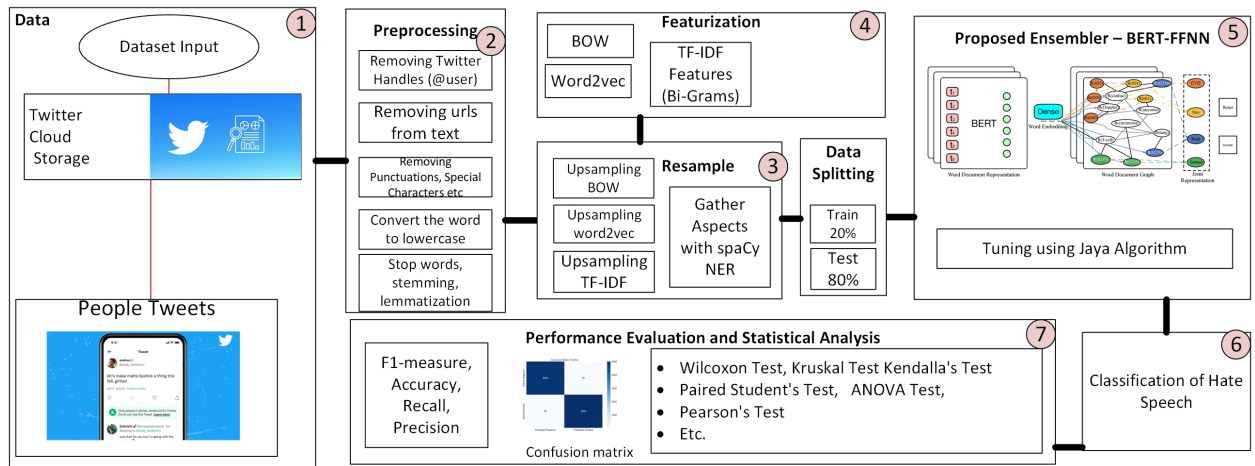


Fig. 1. Proposed framework for aspect based hate speech detection.

study on Aspect-Aware Hate Speech Detection in Tweets using a Hybrid of DistillBERT and Feed Forward Neural Network (DBFN). The data originates mostly from the repository at [26]. the structure of dataset is shown in Fig. 2.

id	label	tweet
1	0	we are so selfish, #orlando #standwithorlando #pulseshooting #orlandoshooting #biggerproblems #selfish #heabreaking #values #love #
2	0	i get to see my daddy today!! #80days #gettingfed
3	1	@user #cnn calls #michigan middle school "build the wall" chant " #tcot
4	1	no comment! in #australia #opkillingbay #seashepherd #helpcovedolphins #thecove #helpcovedolphins
5	0	ouch...junior is angry! #AVA #got? #junior #yugyoem #omg
6	0	i am thankful for having a paner. #thankful #positive
7	1	retweet if you agree!
8	0	its #friday! #AVA #AE smiles all around via ig user: @user #cookies make people
9	0	as we all know, essential oils are not made of chemicals.
10	0	#euro2016 people blaming ha for conceded goal was it fat rooney who gave away free kick knowing bale can hit them from there.
11	0	sad little dude. #badday #conefofshame #cats #pissed #funny #laughs
12	0	product of the day: happy man #wine tool who's it's the #weekend? time to open up & drink up!
13	1	@user @user lumpy says i am a . prove it lumpy.
14	0	@user #gif #ff to my #gamedev #indiedev #indiegamedev #squad! @user @user @user @user
15	0	beautiful sign by vendor 80 for \$45.00!! #upsideofflorida #shopalysas #love
16	0	@user all #smiles when #media is !! #AVA #ce #AVA # #pressconference in #antalya #turkey #sunday #throwback love! #AVA #ASA #AVA #ACA #ASA #A

Fig. 2. Unprocessed twitter dataset (tweets).

With this dataset, This work addresses distinct features of tweets for Aspect-Aware Hate Speech Detection. This technology can detect and evaluate subtleties in hate speech since aspect-wise data has been carefully curated.

B. Performing Preprocessing and EDA

This framework used a number of crucial procedures throughout the preparation stage for tweet texts to improve the consistency of the data. The process is shown in Fig. 3. First, Twitter handles (represented by “@user”) were carefully eliminated using regular expressions to make sure that user-specific data didn’t affect the study that followed. For example, a tweet that began “@user when a father is dysfunctional...” was changed to “when a father is dysfunctional ...”.

Subsequently, hyperlinks and URLs present in the tweet’s contents were eliminated by the use of regular expressions [27]. This can minimize noise from external connections and guarantee that the analysis entirely focuses on the textual content. The tweet, “Click and visit the link: <http://example.com>”, was modified to say, “Click and visit the link:” Using regular expressions, the language was simplified by removing special

characters, digits, and punctuation. This procedure aimed to remove superfluous symbols without affecting the information’s meaning. This is now “in the mid-st century” instead of “in the mid-21st century ...”.

After that, the tweet’s content had lowercase versions of each word. This ensures consistency while reducing the information’s dimensionality since “I Cannot Believe” is now simply “I cannot believe.” Then, to concentrate on the tweets’ more important substance, common stopwords like articles and prepositions were eliminated. For example, the sentence “when you know y’ all 2 ain’t going nowhere” was shortened to “know y’ all 2 ain’t going.” Stemming was used to reduce terms to their root form and refine the data further. To merge related notions, the phrase “waiting for the show to start our third year running” becomes “wait for the show to start our third year run his technique.”

The last technique used to contribute to a more advanced study is lemmatization, which reduces words to their dictionary or base form. One lemmatization of the phrase “waiting for the shows to start our third year running” was “waiting for the show to start our third year running”. The Aspect-Aware Hate Speech Detection method uses the improved tweet text as a basis for further analysis, feature extraction, and model training after these preprocessing processes are completed. The sample of preprocessed tweets is shown in Fig. 4.

C. Featurization and Resampling

In the featurization and resampling phase, the objective is to convert preprocessed tweet text into numerical features using Bag-of-Words (BOW), TF-IDF Features with Bi-Grams, and Word2Vec embeddings [28].

a) Bag-of-Words (BOW):: The model used in this study employed the CountVectorizer function to transform the text data into a matrix of token counts [29]:

$$df_bow = \text{CountVectorizer}(\text{stop_words}=\text{english}).\text{fit_transform}(\text{text_data}) \quad (1)$$

This process captures the frequency of each word in the text, providing a numerical representation for subsequent analysis.

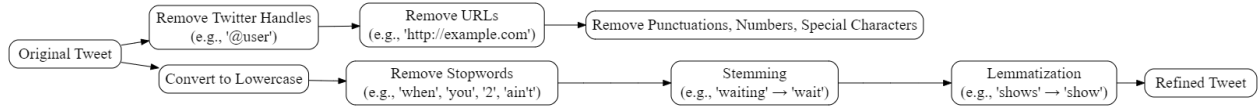


Fig. 3. Process of preprocessing.

id	label	tweet	preprocess_tweet	length_tweet
1	0	we are so selfish, #orlando #standwithorlando #pulseshooting #orlandoshooting #biggerproblems #selfish #heabreaking #values #love #	selfish #orlando #standwithorlando #pulseshoot #orlandoshoot #biggerproblem #selfish #heabreak #valu #love #	108
2	0	I get to see my daddy today!! #80days #gettingfed	get see daddi today # day #gettingf	35
3	1	@user #cnn calls #michigan middle school 'build the wall' chant" #tcot	#cnn call #michigan midd school build wall chant #tcot	55
4	1	no comment! in #australia #opkillingbay #seashepherd #helpcovedolphins #thecov #helpcovedolphins	comment #australia #opkillingbay #seashepherd #helpcovedolphin #thecov #helpcovedolphin	87
5	0	ouch...junior is angry! #got7 #junior #yugyoem #omg	ouch junior angr! #got #junior #yugyoem #omg	44
6	0	I am thankful for having a paner. #thankful #positive	thank paner #thank #posit	25
7	1	retweet if you agree!	retweet agre	12

Fig. 4. Preprocessed tweets.

b) *TF-IDF Features with Bi-Grams*:: Utilizing the *TfidfVectorizer* function, the equation for TF-IDF with Bi-Grams is given by [30]:

$$df_tfidf = TfidfVectorizer(ngram_range=(1, 2), stop_words='english').fit_transform(text_data) \quad (2)$$

This technique considers the importance of terms by incorporating individual words and two-word phrases.

c) *Word2Vec*: The *Word2Vec* function was used to create *Word2Vec* embeddings [30]:

$$\text{Word2Vec} = df_w2v(\text{window} = 5, \text{sentences}, \text{workers} = 4, \text{min_count} = 1, \text{vector_size} = 100) \quad (3)$$

Word2Vec captures the semantic links between words and provides a detailed depiction of the underlying semantics in the tweet text.

d) *Resampling Techniques*: To address class imbalance, the model implemented resampling techniques on the datasets:

Upsampling BOW: To match the majority class (label 0), upsampling entails boosting the occurrences of the minority class (label 1) within the BOW dataset. The equation for upsampling BOW is [31]:

$$df_bow_upsampled = \text{resample}(df_minor, \text{replace}=\text{True}, n_samples=\text{major_class}_0) \quad (4)$$

This technique ensures a balanced representation of both classes in the training data.

Upsampling TF-IDF: Similar to BOW, the TF-IDF dataset underwent upsampling to achieve a balanced class distribution. Eq. 5 show the Tf-IDF upsampling [32]:

$$df_tfidf_upsampled = \text{resample}(df_minor, \text{replace}=\text{True}, n_samples=\text{major_class}_0) \quad (5)$$

Upsampling helps prevent biases towards the majority class.

Upsampling Word2Vec: The *Word2Vec* dataset was upsampled to address the class imbalance. The equation for upsampling *Word2Vec* is [32]:

$$df_w2v_upsampled = \text{resample}(df_minor, \text{replace}=\text{True}, n_samples=\text{major_class}_0) \quad (6)$$

This technique ensures a fair representation of both classes in the training data.

In the proposed Aspect-Aware Hate Speech Detection system, these resampling strategies are essential for avoiding biases, improving model performance, and preserving an equal proportion of non-hate speech and hate speech occurrences. Visualizations, such as count plots, were generated to illustrate the balanced class distribution in the upsampled datasets.

D. Proposed DBFN-SHO

The architecture and design of the suggested aspect-aware hate speech detection model called the DistillBERT [33] and Feed Forward Neural Network [34] (DBFN) are presented in this section. The DBFN model is a hybrid system that effectively classifies hate speech in tweets by combining the strength of a feed-forward neural network with DistillBERT, a simplified version of BERT (Bidirectional Encoder Representations from Transformers).

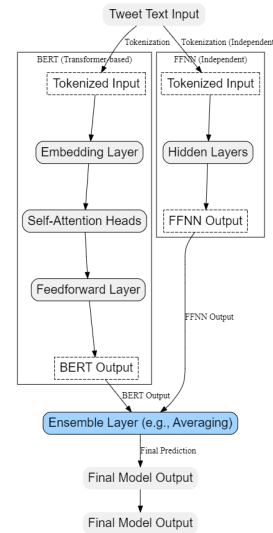


Fig. 5. Proposed DBFN model.

The transformer-based model DistillBERT, represented as DB, extracts word contextual embeddings to produce a contextualized representation of the input text. Parameterizing the model is done with θ_{DB} [34].

By modifying the parameters θ_{DB} of DistillBERT, the model is optimized for the hate speech dataset to reduce the cross-entropy loss [33], [34]:

$$\mathcal{L}_{DB}(\theta_{DB}) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(P_{DB}(x_i; \theta_{DB})) + (1 - y_i) \log(1 - P_{DB}(x_i; \theta_{DB}))] \quad (7)$$

1) *Feed Forward Neural Network*: FFNN is a classifier defined by θ_{FFNN} . This is also known as the Feed-Forward Neural Network. The projected likelihood of hate speech is output, and the contextual embeddings generated by DistillBERT are used as input [34].

$$\hat{y} = P_{\text{FFNN}}(P_{\text{DB}}(x; \theta_{\text{DB}}); \theta_{\text{FFNN}}) \quad (8)$$

a) *Training and Optimization*: The cross-entropy loss is minimized to optimize the parameters θ_{FFNN} :

$$\mathcal{L}_{\text{FFNN}}(\theta_{\text{FFNN}}) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (9)$$

b) *Aspect-Aware Classification*: The DBFN model includes an aspect-aware categorization technique called Aspect-Aware [35] that takes into account many aspects seen in tweets that contain hate speech. The set of aspects, such as gender, race, and religion, is represented by A . The function $P_{\text{Aspect-Aware}}(x)$ generates the aspect-aware predictions for tweet x .

$$P_{\text{Aspect-Aware}}(x) = P_{\text{AA}}(P_{\text{DB}}(x; \theta_{\text{DB}}), P_{\text{FFNN}}(P_{\text{DB}}(x; \theta_{\text{DB}}); \theta_{\text{FFNN}}); \theta_{\text{AA}}) \quad (10)$$

This work investigates ensemble learning strategies to further improve the DBFN model's resilience by embedding the FFNN inside the BERT and, in the end, by merging predictions gathered from various Feed Forward Neural Network and DistillBERT instances, as shown in Fig. 5.

$$P_{\text{Ensemble}}(x) = \frac{1}{K} \sum_{k=1}^K P_{\text{AA}_k}(x) \quad (11)$$

E. Parameter Tuning with Jaya Optimization Algorithm (JOA)

Aspect-Aware Hate Speech Detection model DBFN performs best when hyperparameters are fine-tuned with suggested technique JOA [36]. JOA is a method motivated by cooperative population dynamics. Optimization is applied to the following hyperparameters: epoch count, learning rate, batch size, DB hidden units, and FFNN hidden units. These hyperparameters are essential factors that affect the model's accuracy, reliability, and stability. Table II summarises the optimum values for each hyperparameter.

TABLE II. OPTIMISTIC HYPERPARAMETERS AND THEIR APPROPRIATE VALUES FOR TUNING

Hyperparameter	Optimized Value
Batch Size:	128
DistillBERT Hidden Units	256
FFNN Hidden Units	128
Epochs	30
Learning Rate	0.0005

The JOA is fully defined in Algorithm 1 [36]. Iteratively adjusts hyperparameter settings based on the model's efficacy on a validation set. The algorithm investigates the hyperparameter space and modifies configurations using crossover and mutation procedures. Until convergence is reached, the process keeps going.

Algorithm 1 Jaya Optimization Algorithm for Hyperparameter Tuning

```

1: Input:
2: a collection of hyperparameter settings  $I$ 
3:  $f(\text{configuration})$  the objective function.
4: Range for every  $[L, U]$  base_parameter STATE threshold of convergence  $\theta$ 
5: Output: Optimal values of base_parameters
6: Optimization_Jaya
7: The convergence threshold  $\theta$ 
8: Set up the optimal arrangement first:  $y_{\text{best}}$  from  $I$ 
9: while values Not meet do
10:   for Every  $y_i$  configuration in  $P$  do
11:     Using  $e$ , consistently generate a random number in the interval  $[0, 1]$ .
12:     Revise the setup:
13:        $y_j = r \cdot (y_{\text{best}} - y_j) + y_j$ 
14:     Verify that the parameters are within the specified range.
15:        $y_j = \min(\max(y_j, L), U)$ 
16:   end for
17:   Determine which configuration,  $y_{\text{best}}$ , has the highest value of the objective function.
18: end while

```

The proposed hate speech detection model, DBFN-J, is more predictive and resilient when the Jaya Optimization Algorithm is included. This refined model, which encapsulates the DB-based hybrid architecture, makes effectively detecting and categorizing hate speech elements in tweets possible.

Using data preprocessing and the Jaya-optimized DBFN model training, Algorithm 1 provides a comprehensive overview of the proposed hate speech detection model. The resultant model, DBFN-J, is intended to offer accurate and trustworthy predictions for aspect-aware hate speech detection within the context of tweets on social media.

F. Classification Assessment Metrics

This methodology uses a hybrid DBFN approach to identify the features of hate speech in tweets. It implements numerous parameters for assessing the efficacy of the proposed technique and verifying its accuracy and usefulness in detecting hate speech in different situations [37].

a) *AUC and ROC Analysis*: The Receiver Operating Characteristic Curve (ROC) metrics are applied to determine whether a precise approach is effective in recognizing various aspects of hate speech. The curved shape depicts the disparity between realistic positive effects and incorrect negative results. The model's overall discriminative ability is evaluated using the Area Under the Curve (AUC) [38]. TPR measures the capacity of the algorithm to recognize inappropriate speech through its attributes. However, FPR measures the technique's capacity to identify the difference between hate speech and non-hateful content. The study of ROC curves and AUC estimation is implemented to accurately assess the method's effectiveness in recognizing hateful speech.

b) *Accuracy and Recall*: The proactive stability and durability of the framework towards understanding hate speech

features will be assessed by applying specific metrics. The algorithm's accuracy evaluates its ability to identify hate speech characteristics, particularly in challenging circumstances. The recall comparisons determine how effectively the model differentiates hate speech in practical situations and excludes instances, and it performs inadequately in this aspect. Eq. 12 [36] and 13 [37] will be used to assess the recall and accuracy using the parameters of the TPR, FPR, and TNR.

$$\text{Precision} = \frac{\text{TPR}}{\text{TPR} + \text{FPR}} \quad (12)$$

$$\text{Recall} = \frac{\text{TPR}}{\text{TPR} + \text{FNR}} \quad (13)$$

By indicating the anticipated reliability of the technique, these variables yield essential details concerning the degree to which the technology recognizes aspects of hate speech.

c) Logloss Assessment: The corresponding decrease in exponential accuracy is a significant statistic that matters when measuring the predicted efficiency of a hate speech recognition strategy. This metric is illustrated by the coefficient estimator 14 [38], which determines the disparity between the estimated chances and actual probability.

$$\log \text{Loss} = -\frac{1}{M} \sum_{j=1}^M (x_j \log(I_j) + (1 - x_j) \log(1 - I_j)) \quad (14)$$

This algorithm assesses the model's fit across its likelihood estimates and the true identifiers. Highlighting an increased correspondence with the predicted and actual labels, a decrease in log deficit implies an improvement in recognizing hate speech aspects.

Statistical Analysis for Assessment: A rigorous statistical assessment is employed to assess the combined DBFN technique using other strategies and basic models. Several statistical approaches, such as ANOVA, Student's t-test, median deviation, standard deviation, and range, are applied throughout the evaluation to assess the variety and value of the data. Researchers analyze the computational difficulty to estimate the additional resources needed to facilitate the hybrid approach's implementation. The in-depth examination proposes an extensive explanation of the adaptation and usability of the suggested approach.

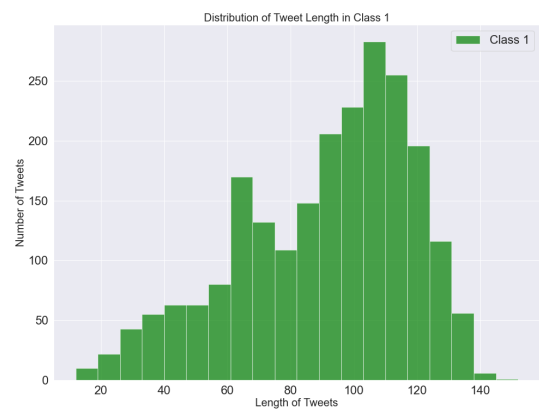
The hybrid strategy approach aims to recognize features that characterize hate speech in tweets. It offers in-depth knowledge of each computational efficacy and accurate prediction reliability. Comprehensive statistical studies and the previously outlined assessment criteria, which produce significant data, demonstrate the model's predicted accuracy in many aspects of hate speech.

IV. SIMULATION RESULTS AND DISCUSSION

The proposed framework is simulated on a computer with a Core i7 processor and 32GB RAM. This study was carried out using Python. Multiple datasets containing tweets were combined for analysis related to hate speech. The choice of this

dataset is attributed to its updated status in 2022, and extensive simulations were conducted to assess its effectiveness and flexibility. The experimental results are elaborated upon in the subsequent discussion.

Initially, the hate speech dataset was investigated. A detailed summary of the dataset's noteworthy technical and demographic trends can be found in Fig. 6. The histograms provide insightful information about how tweet durations are distributed throughout the dataset's various classifications. Fig. 6a's histogram depicts the length distribution of Class 1 tweets or tweets containing hate speech. The green bars, which display the frequency of tweets at various durations, provide a thorough understanding of the distribution pattern within this class. A similar analysis is demonstrated for tweets in Class 0 (non-hate speech) in the Fig. 6b histogram.



(a) Distribution of class 1 (hate speech).



(b) Distribution of Class 0 (non-hate speech)

Fig. 6. Comparison of tweet length distributions.

Table III shows the technical specs of hate speech detection models that use the proposed DBFN-J model and their baseline versions. With a low log loss of 0.06, high accuracy of 0.97, and intense discrimination, as evidenced by an AUC of 0.989, the DBFN-J model performs better than the others. Metrics such as ROC-CH (0.95) and MCC

The ROC-CH of 0.95 for this framework demonstrates its remarkable capacity to balance TP and FP rates. In addition, an MCC value of 0.93 displays the algorithm’s general efficiency and indicates significant consistency between the estimated and observed categorization. The combination of scientific findings indicates the reliability of the DBFN-J framework for identifying inappropriate comments.

TABLE IV. AVERAGE COMPUTATIONAL TIME ANALYSIS

Model	Median (s)	Mean (s)	Min (s)	Max (s)	Range (s)	Std. Dev. (s)
BERT [19]	86	87	75	97	21	5.21
CNN [17]	85	85	74	96	21	5.36
BERT-LSTM [17]	86	87	78	96	17	4.15
CNN-LSTM [21]	86	89	73	96	22	5.39
ELMo [27]	85	84	75	97	21	4.95
SVM [19]	84	84	76	92	15	4.09
DBFN-J (Proposed)	36	35	33	40	4	0.11

Table IV contains a discussion of the processing time that indicates the various gains among various study methods. Compared to the remainder models, the DBFN-J approach is more efficient, as evidenced by its substantially quicker median and mean execution times and less uncertainty. This demonstrates the highly computationally efficient suggested model, making it a good option for real-world scenarios where processing speed is crucial. Researchers compare the mean, median, and standard deviation of the various models—including BERT, CNN, BERT-LSTM, ResNet, ELMo, and SVM—to thoroughly understand each model’s efficiency profile. These models display differing degrees of computing performance.

TABLE V. STATISTICAL ANALYSIS OF THE PROPOSED WRNG-J AND EXISTING METHODS

Method	Test	F-stat	P-Value
BERT	Kendall’s	0.553	-0.011
	Pearson’s	0.623	-0.011
	Chi-Squared	105.21	0.004
	Spearman’s	0.598	-0.011
ELMo	Kendall’s	0.623	-0.011
	Pearson’s	0.623	-0.011
	Chi-Squared	101.694	-0.011
	Spearman’s	0.623	-0.011
DBFN-J	Kendall’s	0.79	-0.011
	Pearson’s	0.888	-0.011
	Chi-Squared	109.429	0.042
	Spearman’s	0.856	-0.011
CNN-LSTM	Kendall’s	0.623	-0.011
	Pearson’s	0.623	-0.011
	Chi-Squared	101.694	-0.011
	Spearman’s	0.623	-0.011
Non-Parametric Tests	Wilcoxon	15313.989	0.156
	Kruskal	6.706	0.008
	Mann-Whitney	26519.989	-0.011
Parametric Tests	Student’s	-0.742	0.454
	Paired Student’s	-1.079	0.285
	ANOVA	0.533	0.454

After a detailed statistical study, Table V highlights the crucial trends and distinctions between the effectiveness of the recommended and current strategies. In addition to providing an extensive data breakdown and significant levels for many statistical tests, the table enables a comprehensive evaluation of the advantages and disadvantages of the different approaches. A “0” p-value indicates the absence of statistically significant impacts or differences. From a statistical perspective, a result is considered vital if it means a difference or impact and has a p-value more than zero, which is still highly tiny (preferably less than 0.04). The p-value in this instance is insignificant

because p-values are typically positive. As a result, care must be used while examining data with negative p-values.

V. CONCLUSION AND FUTURE DIRECTIONS

The hate speech recognition system driven by the DBFN-J model is a significant development in the field. Using a large-scale Twitter dataset collected over the previous four years from a GitHub repository, the system uses NLP tokenization to do careful data preprocessed. The dataset is better when unnecessary elements, such as data characters, hashtags, and user information, are removed. By investigating semantic sentiment unigram and pattern characteristics, the system derives insightful information and creates vectors that guide further categorization. An ensemble of deep neural network classifiers enhanced by adding the Jaya method for fine-tuning parameters performs remarkably well. With a good accuracy rate of 97% and a small loss function of 0.06, the DBFN-J model demonstrates its effectiveness in identifying hate speech. This work is noteworthy for its lightweight and efficient technique, which outperforms well-established models like CNN and BERT ELMo in terms of performance. The application of hybrid techniques further strengthens the total classification accuracy.

Although effective, the DBFN-J model has drawbacks. One dataset from Twitter may limit the model’s applicability. Second, while effective, preprocessing may remove hashtags and user metadata, affecting model interpretability. The model may also struggle to identify subtle or implicit hate speech in low-resource languages. Future research can use multimodal social media datasets to improve model adaptability and robustness. Future research may improve implicit hate speech detection with transformer-based architectures like GPT or T5. Multilingual models for low-resource languages and cultural differences are promising. Finally, real-time deployment of the DBFN-J model with dynamic feedback mechanisms for continuous improvement may help combat online hate speech.

ACKNOWLEDGMENT

The work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant number (UJ-23-SHR-77). The authors, therefore, acknowledge with thanks the University of Jeddah for technical and financial support.

REFERENCES

- [1] S. Windisch, S. Wiedlitzka, A. Olaghere, and E. Jenaway, *Online interventions for reducing hate speech and cyberhate: A systematic review*, Campbell Systematic Reviews, vol. 18, no. 2, pp. e1243, 2022.
- [2] A. Tontodimamma, E. Nissi, A. Sarra, and L. Fontanella, *Thirty years of research into hate speech: topics of interest and their evolution*, Scientometrics, vol. 126, no. 1, pp. 157–179, 2021.
- [3] E. Aswad and D. Kaye, *Convergence & conflict: reflections on global and regional human rights standards on hate speech*, Nw. UJ Int’l Hum. Rts., vol. 20, pp. 165, 2021.
- [4] B. Nyagadza, *Search engine marketing and social media marketing predictive trends*, Journal of Digital Media & Policy, vol. 13, no. 3, pp. 407–425, 2022.
- [5] R. Qasim, W. H. Bangyal, M. A. Alqarni, and A. A. Almazroi, *A fine-tuned BERT-based transfer learning approach for text classification*, Journal of Healthcare Engineering, vol. 2022, no. 1, pp. 1–11, 2022.

- [6] H. Simon, B. Y. Baha, and E. J. Garba, *Trends in machine learning on automatic detection of hate speech on social media platforms: A systematic review*, FUW Trends in Science & Technology Journal, vol. 7, no. 1, pp. 001–016, 2022.
- [7] A. A. Almazroi, L. Abualigah, M. A. Alqarni, E. H. Houssein, A. Q. M. AlHamad, and M. A. Elaziz, *Class Diagram Generation from Text Requirements: An Application of Natural Language Processing*, in *Deep Learning Approaches for Spoken and Natural Language Processing*, Springer, pp. 55–79, 2021.
- [8] K. Sharifani and M. Amini, *Machine learning and deep learning: A review of methods and applications*, World Information Technology and Engineering Journal, vol. 10, no. 07, pp. 3897–3904, 2023.
- [9] A. A. Almazroi, *A fast hybrid algorithm approach for the exact string matching problem via berry ravindran and alpha skip search algorithms*, Journal of Computer Science, vol. 7, no. 5, pp. 644, 2011.
- [10] R. T. Mutanga, N. Naicker, and O. O. Olugbara, *Detecting Hate Speech on Twitter Network using Ensemble Machine Learning*, International Journal of Advanced Computer Science and Applications, vol. 13, no. 3, pp. 1–10, 2022.
- [11] B. Kennedy, M. Atari, A. M. Davani, L. Yeh, A. Omrani, Y. Kim, and M. Dehghani, *Introducing the Gab Hate Corpus: defining and applying hate-based rhetoric to social media posts at scale*, Language Resources and Evaluation, pp. 1–30, 2022.
- [12] P. Chiril, E. W. Pamungkas, F. Benamara, V. Moriceau, and V. Patti, *Emotionally informed hate speech detection: a multi-target perspective*, Cognitive Computation, pp. 1–31, 2022.
- [13] M. Wankhade, A. C. S. Rao, and C. Kulkarni, *A survey on sentiment analysis methods, applications, and challenges*, Artificial Intelligence Review, vol. 55, no. 7, pp. 5731–5780, 2022.
- [14] S. Nagar, F. A. Barbhuiya, and K. Dey, *Towards more robust hate speech detection: using social context and user data*, Social Network Analysis and Mining, vol. 13, no. 1, pp. 47, 2023.
- [15] P. Som, R. Mishra, S. Das, R. K. Singh, D. K. Rakesh, B. Behera, and R. R. Kumar, *Evaluating Machine Learning Models for Hate Speech Detection in ODIA Language*, in *2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU)*, IEEE, pp. 1–6, 2024.
- [16] A. C. Mazari, N. Boudoukhani, and A. Djeflal, *BERT-based ensemble learning for multi-aspect hate speech detection*, Cluster Computing, vol. 27, no. 1, pp. 325–339, 2024.
- [17] S. Chinivar, M. S. Roopa, J. S. Arunalatha, and K. R. Venugopal, *Online offensive behaviour in social media: Detection approaches, comprehensive review and future directions*, Entertainment Computing, vol. 45, pp. 100544, 2023.
- [18] K. Maity, G. Balaji, and S. Saha, *Towards Analyzing the Efficacy of Multi-task Learning in Hate Speech Detection*, in *International Conference on Neural Information Processing*, Springer Nature Singapore, pp. 317–328, 2023.
- [19] H. Saleh, A. Alhothali, and K. Moria, *Detection of hate speech using BERT and hate speech word embedding with deep model*, Applied Artificial Intelligence, vol. 37, no. 1, pp. 2166719, 2023.
- [20] I. P. Sari and H. Maulana, *Detecting Cyberbullying on Social Media Using Support Vector Machine: A Case Study on Twitter*, International Journal of Safety & Security Engineering, vol. 13, no. 4, pp. 1–10, 2023.
- [21] S. Saifullah, R. Dreżewski, F. A. Dwiyanto, A. S. Aribowo, Y. Fauziah, and N. H. Cahyana, *Automated text annotation using a semi-supervised approach with meta vectorizer and machine learning algorithms for hate speech detection*, Applied Sciences, vol. 14, no. 3, pp. 1078, 2024.
- [22] H. Vanam and J. R. Raj, *CNN-OLSTM: Convolutional Neural Network with Optimized Long Short-Term Memory Model for Twitter-based Sentiment Analysis*, IETE Journal of Research, pp. 1–12, 2023.
- [23] S. Zhong, A. Scarinci, and A. Ciciello, *Natural language processing for systems engineering: automatic generation of systems modelling language diagrams*, Knowledge-Based Systems, vol. 259, pp. 110071, 2023.
- [24] A. Kumar and S. Kumar, *Hate speech detection in multi-social media using deep learning*, in *International Conference on Advanced Communication and Intelligent Systems*, Springer Nature Switzerland, pp. 59–70, 2023.
- [25] A. Balayn, J. Yang, Z. Szlavik, and A. Bozzon, *Automatic identification of harmful, aggressive, abusive, and offensive language on the web: A survey of technical biases informed by psychology literature*, ACM Transactions on Social Computing (TSC), vol. 4, no. 3, pp. 1–56, 2021.
- [26] A. Ammar, *Datasets for Hate Speech Detection*, Retrieved from <https://github.com/aymeam/Datasets-for-Hate-Speech-Detection>.
- [27] F. Fkih, T. Moulahi, and A. Alabdulatif, *Machine learning model for offensive speech detection in online social networks slang content*, WSEAS Trans. Inf. Sci. Appl., vol. 20, pp. 7–15, 2023.
- [28] H. A. Madni, M. Umer, N. Abuzinadah, Y. C. Hu, O. Saidani, S. Alsoubai, M. Hamdi, and I. Ashraf, *Improving sentiment prediction of textual tweets using feature fusion and deep machine ensemble model*, Electronics, vol. 12, no. 6, pp. 1302, 2023.
- [29] S. Dai, K. Li, Z. Luo, P. Zhao, B. Hong, A. Zhu, and J. Liu, *AI-based NLP section discusses the application and effect of bag-of-words models and TF-IDF in NLP tasks*, Journal of Artificial Intelligence General Science (JAIGS), vol. 5, no. 1, pp. 13–21, 2024.
- [30] A. Dey, M. Jenamani, and J. J. Thakkar, *Lexical TF-IDF: An n-gram feature space for cross-domain classification of sentiment reviews*, in *International Conference on Pattern Recognition and Machine Intelligence*, Springer, pp. 380–386, 2017.
- [31] D. Rau, M. Dehghani, and J. Kamps, *Revisiting Bag of Words Document Representations for Efficient Ranking with Transformers*, ACM Transactions on Information Systems, vol. 42, no. 5, pp. 1–27, 2024.
- [32] A. Banerjee, P. Shivakumara, S. Bhattacharya, U. Pal, and C. L. Liu, *An end-to-end model for multi-view scene text recognition*, Pattern Recognition, vol. 149, pp. 110206, 2024.
- [33] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*, arXiv preprint arXiv:1910.01108, 2019.
- [34] H. Saleh, A. Alhothali, and K. Moria, *Detection of hate speech using BERT and hate speech word embedding with deep model*, Applied Artificial Intelligence, vol. 37, no. 1, pp. 2166719, 2023.
- [35] N. Ayub, H. Tayyaba, S. Hussain, S. S. Ullah, and J. Iqbal, *An Efficient Optimized DenseNet Model for Aspect-Based Multi-Label Classification*, Algorithms, vol. 16, no. 12, pp. 548, 2023.
- [36] R. A. Zitar, M. A. Al-Betar, M. A. Awadallah, I. A. Doush, and K. Assaleh, *An intensive and comprehensive overview of JAYA algorithm, its versions and applications*, Archives of Computational Methods in Engineering, vol. 29, no. 2, pp. 763–792, 2022.
- [37] A. A. Almazroi, O. A. Mohamed, A. Shamim, and M. Ahsan, *Evaluation of State-of-the-Art Classifiers: A Comparative Study*, Journal of Computing, vol. 1, no. 1, pp. 22–29, 2020.
- [38] D. Chicco and G. Jurman, *The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification*, BioData Mining, vol. 16, no. 1, pp. 4, 2023.

Exploring the Best Machine Learning Models for Breast Cancer Prediction in Wisconsin

Abdullah Al Mamun¹, Dr. Touhid Bhuiyan^{2*}, Md Maruf Hassan³, Shahedul Islam Anik⁴

Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh^{1,4}

School of IT, Washington University of Science and Technology, VA, USA²

Department of Computer Science and Engineering, Southeast University, Dhaka, Bangladesh³

Abstract—This research focuses on predicting Wisconsin Breast Cancer Disease using machine learning algorithm, employs a dataset offered by UCI repository (WBCD) dataset. The under-gone substantial preparation, includes managing missing values, normalization, outlier elimination, increase data quality. The Synthetic Minority Oversampling Technique (SMOTE) is used to alleviate class imbalance and to enable strong model training. Machine learning models, include SVM, kNN, Neural Networks, and Naive Bayes, were built and verified using Key performance metrics and K-Fold cv. included as recall, accuracy, F1-score, precision and AUC-ROC were employed to analyze the models. Among these, the Neural Network model emerged the most effective, obtaining a prediction accuracy 98.13%, precision 98.21%, recall 98.00%, F1Score of 97.96%, AUC-ROC score 0.9992. Study underscores promise of ML boosting the diagnosis and treatment of WBCD illnesses, giving scalable and accurate ways for early detection and prevention.

Keywords—Wisconsin breast cancer disease prediction; ML; SVM; KNN; AUC-ROC; Naive Bayes

I. INTRODUCTION

Among the most common cancers worldwide, breast cancer is a cause of death among women. Approximately 508,000 female died from breast cancer in 2011, according to the WHO. While mammograms and biopsies are effective diagnostic tools, they are often invasive and prone to errors. Current figures show that one in eight women will develop breast cancer, growing the need for early precise find to improve long-suffering survival rates through timely intervention and therapy [2]

Breast cancer begins in breast tissue and can metastasize, making it a primary cause of mortality among women. In 2018, the disease caused 9.6 million deaths globally, with predictions of a 50% increase in cases by 2040 [5]. Emerging data mining and big data technologies are now being employed to forecast and treat breast cancer, potentially enhancing patient care and reducing healthcare costs.

Advances in machine learning (ML) have enabled data-driven medical diagnoses, where algorithms analyze vast datasets to uncover patterns often missed by human diagnosticians. This paper compares multiple ML models for breast cancer of event anticipation using WBCD dataset, SVM, KNN, D-T, R-F, and L-R.

The study focuses on hyperparameter tuning to optimize performance metrics as prediction precision, recall, and accuracy. Without relying on feature selection Hyperparameter tuning systematically adjusts model parameters to identify optimal algorithm configurations, enhancing predictive accuracy. The structure of the stay of the paper is follow: Section II review machine learning literature related to breast cancer prediction. Section III describes the dataset and preprocessing steps, while Section IV details the training and evaluation of machine learning models. Section V present results and discussion models performance. Section VI achieve the analysis and suggest future directions for research in BC prediction.

This project aims to harness ML algorithms such as SVM, DT, and Neural Networks for predicting breast cancer using the WBCD dataset. By comparing these techniques with traditional diagnostic models, the study seeks to improve early detection accuracy, optimize patient treatment, and contribute to the global effort to reduce the burden of breast cancer [12].

II. RELATED WORK

In Emilija Strelcenia et al. [1] (2023), the author early predicted BC increase survival chances and advance previously medical treatment. Breast cancer is a prevalent and serious public health problem that requires early detection and treatment. An accurate diagnosis and classification of benign cases can prevent unnecessary treatments. The paper presents a feature engineering method extract, modified feature from data using WBCD Dataset. Method is used compare six popular machine learning model for classifications: Random-Forest and Logistic-regression, Decision-Tree, MultiLayer Perceptron (MLP), KNeighbors and XG-Boost. when applied to the proposes feature engineering, achieved average accuracy of 98.64%.

The study, Aboudr MAA et al. [2] (2023) suggests the FLN algorithm as a way to make Breast Cancer diagnoses more accurate. (1) the FLN method can get rid of overfitting; (2) it can handle binary and multiclass classification problems; and it can work like a kernel-based support vector machine with structure of neural network. They used WBCD, which is breast cancer database. The experiment showed that the suggested FLN method worked very well, with an average of 98.37% accuracy, 95.44% precision, 99.40% memory, 97.644% F-measure, 97.654% G-mean, 96.444% MCC, and 97.854% specificity using the WBCD. This shows that the FLN method is a good way to diagnose BC, and it might also help with

*Corresponding authors.

other problems in the healthcare field that have to do with applications.

Hossin, M. M., et al. [3] (2023) looks at eight machine learning methods for finding breast cancer. These are LR, RF, KNN, DT, AB, SVM, GB, and GNB. The Wisconsin Diagnostic Dataset is used to test these models and make sure they work. Sensitivity, specificity, Accuracy, area under curve (AUC) were used to measure how well model worked. Logistic Regression: Out of all the methods, it works 99.12% of the time. Researchers said that the study shows how important it is to find and treat breast cancer early so that people can live.

Arpit Bhardwaj et al. [4] (2022) compares four algorithms that are used for the WBCD dataset. These are MLP, KNN, GP, and RF, which are all classification algorithms. which was made by taking samples of the breast with a fine needle. We used genetic programming (GP), RF, multilayer MLP, and KNN on the WBCD dataset to sort the patients into those who are benign and those who are cancerous. RF has a classification rate of 96.24%, which is better than all the other classifiers. Based on the data of the suggested method, probable breast cancer is labelled.

Rasool, Abdur, et al. [5] (2022) is mostly about the WDBC approach. The author used a four-layer data exploratory method (DET) to make the model work better. This technique included feature selection, correlation analysis, and hyperparameter optimisation. The polynomial SVM model was the most accurate 99.3%. It was followed by the LR model 98.6%, the KNN model 97.35%, and the EC model 97.61%. The study used Kfold CrossValidation, confusion matrices show that the models worked even better. These results are in line with other study that has shown that SVM models are better at diagnosing breast cancer.

This is Kadhim, R. R. et al. [6] (2022), the main point of the study is to compare different ways to classify breast cancer using machine learning algorithms. With a score of 96.77, extreme randomise trees had the best F1-score out of the eleven models tested using the Wisconsin dataset. Specificity, sensitivity, precision, accuracy, and F1 score were used to rate how well each model worked. The goal of this study is to help find breast cancer early by finding the best Machine Learning models for classification.

In Sara Ibrahim et al. [7] [2021], the WBCD was used to test the author's suggested approach in this paper. For reducing the number of dimensions, analysis of correlation. Well-known machine learning models were tested to see how well they worked, and the seven best ones were picked for the next step. Tuning the hyperparameters was done to make the algorithms work better. Two different vote methods mixed with the classification algorithms that worked the best. Hard voting picks class that pick the most votes, while soft voting picks the class that has the best chance of winning. With accuracy 98.24%, a high precision 99.29%, and recall value 95.89%, the suggested method did better than the best work that had been done before.

In S.A. Abdulkareem, et al. [8] (2021), Wisconsin Breast Cancer Dataset (WBCD) and the Recursive Feature Elimination (RFE) algorithm are used to show how well the Random Forest and XGBoost classifiers work for finding breast cancer. The high level of accuracy reached by XGBoost 99.02% shows

that ensemble models are useful for medical tasks. When it comes to classification tasks, ensemble methods often work better than single classifiers. This is often seen in finding breast cancer, and machine learning classifiers like SVM and Random Forest have been used a lot.

In Naji Mohammed Amine et al. [9] (2021), the author says that improving the WDBC prediction for high accuracy is important to keep treatment and survival rates up to date. Once they had the results, they used five machine learning algorithms on the Breast Cancer Wisconsin Diagnostic dataset: SVM, RF, LR, DT (C4.5), and KNN. Goal of this study is the use ML model to identify and diagnose breast cancer and find the best ones in terms of confusion matrix accuracy and precision. Support vector Machine did better than all the other.

In Sahar A. El Rahman et al. [10] (2021), the authors aims to describe breast cancer early using machine learning algorithms and features selection methods. The methodology includes four datasets, preprocessing, processing, and model evaluation. Different classifiers such as decision tree, RF, LR, Naïve-Bayes, Knearest-neighbor, and support vector machine are compared using four different breast cancer datasets. The prospective models are checked using classification accuracy and confusion matrix. The results show that the RF technique with Genetic Algorithm (GA) is the most accurate, with an accuracy value of 96.82% on the WBC dataset. The C-SVM technique with the applied kernel function RBF is more advanced, with an accuracy value of 99.04% on the WBCD dataset. The RF technique with recursive feature elimination is the best, with an accuracy value of 74.13% on the WPBC dataset. The proposed models are useful compared to extant models.

In Neha Panwar et al. [11] (2020), we use different Machine Learning (ML) techniques to figure out if a patient has BC or not. SVM, k-NN, NB, DT, and LR will be used to sort the WBCD dataset in this work. Before classification, there is a preprocessing step where five different classifiers are used with the five fold cross-validation method. Performance factors like sensitivity, accuracy, and specificity are used to measure how well classification works. Confusion metrics are also used to measure performance. It was found that SVM worked best, with a precision of 99.12% after the normalisation process in.

In Adel S. Assiri et al. [12] (2020), the WBCD was used to compare how well different cutting-edge machine learning classification methods worked. Based on their F3 score, the three best models were then chosen. The F3 score is used to stress how important false positives are in classifying breast cancer. Simple LR learning, svm learning with Stochastic-Gradient descent optimisation, multilayer-Perceptron network are the three classifiers that are used for ensemble classification with a vote system. With a success rate of 99.42%, the hard voting (majority-based voting) method works better than the most recent WBCD algorithm.

III. PRELIMINARY SECTION

Section give data information and evaluations matrices for this study.

A. Data Description

the WBC dataset derived from the UCI repository ML datasets [17]. This collection includes 569 instances, categorized as either benign or malignant, with 357 instances (62.74 per cent) identified as benign and 212 instances (37.25 per cent) as malignant. The dataset is segmented into two categories, B for benign and M for malignant. Breast cancer stands as the most frequently diagnosed condition in healthcare, and its incidence is on the rise annually. Beyond the sample code numbers and class labels, the dataset features 32 characteristics related to breast cancer, such as the mean radius, texture, area, smoothness, compactness, and concavity [18, 19]. Cases labeled as benign are considered less harmful to the body, whereas those labeled as malignant are deemed harmful due to their cancerous nature in our research. The dataset contains 16 instances with missing values for features, which are typically filled using the mean method. To guarantee the integrity of the data, the dataset is randomized at the end (Fig. 1).

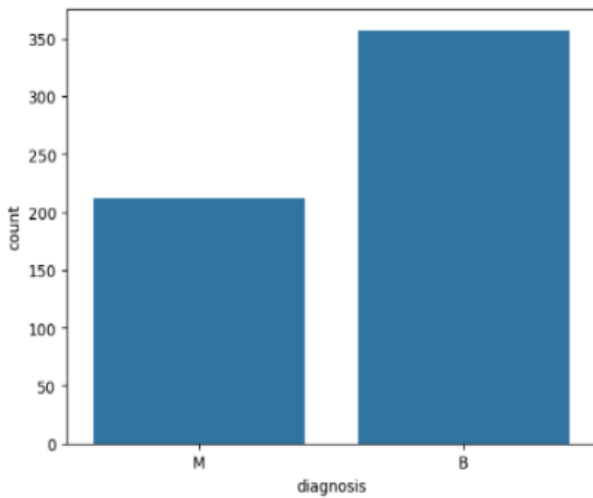


Fig. 1. Wisconsin breast cancer diagnostic datasets.

B. Data Preprocessing

Ensure that data is appropriately prepare for machine learning models, the following preprocessing steps will be implemented:

C. Handling Missing Data

Missing values will be imputed using techniques such as KNN imputation or mean/mode imputation, depending on the nature and distribution of the data.

D. Categorical Data Encoding

Categorical features, such as *gender*, will convert into numeric values using encoding method example OneHot encoding, LabelEncoding.

E. Data Splitting

Dataset will be divide two subset: 80% for train & 20% for test. This split will facilitate model evaluation and prevent overfitting.

F. Performance Evaluation Metrics

Four distinct CrossValidation metrics precision, recall, accuracy, and F1Score were examined in this work. The values of the confusion matrix allow one to ascertain these measures. The confusion matrix consists of the following elements:

- TP: The model predicts “yes” and the actual data is also “yes”.
- TN: The model predicts “no” and the actual data is also “no”.
- FP: The model predicts “yes” but the actual data is “no”.
- FN: The model predicts “no” but the actual data is “yes”.

The following formulas allow one to calculate accuracy, F1-score, precision, and recall:

G. Formula

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy measures the overall correctness of the model and the ratio of correctly prediction (both truely-positive and truely-negative) to total number of prediction.

H. Formula

$$\text{Precision} = \frac{TP}{TP + FP}$$

Precision measure how many of predicted positively instances are really correct, providing insight into models ability to bypass false positives.

I. Formula

$$\text{Recall} = \frac{TP}{TP + FN}$$

Recall measures models ability to correctly identify all relevant positive instances, highlighting the detection capability.

J. Formula

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

The F1-Score provides a harmonic mean of precision and recall, balancing the trade-off between the two metrics.

K. Formula

The AUC (Area Under the Curve) and ROC (Receiver Operating Characteristic) curve evaluate the model ability to discriminate between the classes. A high AUC indicate better performance distinguishing between positive and negative instance. The ROC curve plotted by use the True Positive Rate (Recall) on y-axis, False Positive Rate (FPR) on x-axis.

$$\text{True Positive Rate (Recall)} = \frac{TP}{TP + FN}$$

$$\text{False Positive Rate (FPR)} = \frac{FP}{FP + TN}$$

Each metric provides valuable insights into different aspects model performance can be used based on the specific needs of the application or the dataset.

L. Methodology

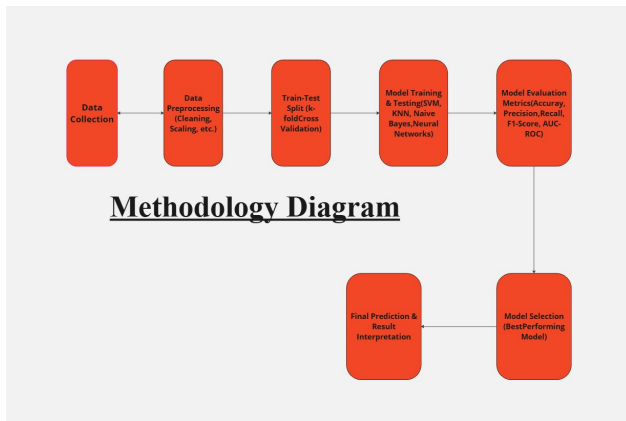


Fig. 2. Process flow diagram.

Our main objective, this study is identify most effective, reliable ML model for predicting Wisconsin Breast Cancer Disease (WBCD) risk. In this research, we have applied multiple ML algorithms including SVM, k-Nearest KNN, GaussianNB, and MLPClassifier. After training and evaluating each model, we assessed their performance to identify the best model based on accuracy and other key evaluation metrics (Fig. 2).

Proposed approach begins with data collection, followed by the preprocessing stage, which includes data cleansing, feature selection, targeting role definition, and feature extraction. Once the data is prepared, we apply various machine learning algorithms to build models that can predict WBCD risk based on input features like age, gender, blood pressure, cholesterol levels, etc.

To evaluate the models performance we split dataset into two subsets: training data and testing data, typically using the Train-Test Split technique. In our case, 80% of the dataset is used for training, while the remaining 20% is used to evaluate the model's performance. We then compare the results of the different algorithms in terms of their accuracy, precision, recall, F1-Score, and AUC-ROC values.

Finally, based on these performance metrics, we select the best-performing model for WBCD risk prediction, ensuring that the chosen model is not only accurate but also reliable in its predictions.

IV. MACHINE LEARNING ALGORITHMS

A. Support Vector Machine (SVM)

SVM a supervised learning algorithm finds optimal hyper-plane to separate different classes. It is particularly effective in high-dimensional spaces and is used in classification tasks.

Mathematical Equation:

$$f(x) = w \cdot x + b$$

where:-

- w - weight-vector,
- x - feature-vector,
- b - bias-term.

B. KNearest Neighbor

KNN is simple, instance-based learning algorithm that use to classify data point based on the majority class of their k nearest-neighbors.

Mathematical Equation:

$$y = \frac{1}{k} \sum_{i=1}^k y_i$$

where:

- y_i is the class of the nearest neighbors,
- k is the number of neighbors.

C. Naive Bayes (GaussianNB)

Naive Bayes, Probabilistic classifier based on bayes-theorem, assume independently among features. It is especially suitable for large datasets and text classification tasks.

Mathematical Equation:

$$P(C | X) = \frac{P(X | C) \cdot P(C)}{P(X)}$$

where:

- $P(C | X)$, Posterior Probability of class C given features X ,
- $P(C)$, Prior Probability of class C ,
- $P(X | C)$, likelihood of observing X given class C .

D. Neural Networks (MLPClassifier)

The MultiLayer-Perceptron(MLP) type of feed-forward artificial neural-network that consisting multiple layers of neuron, used for nonlinear classifications.

Mathematical Equation:

$$y = \sigma(Wx + b)$$

where:-

- W - weight matrix,
- x - input vector,
- b - bias,
- σ - activation function (e.g. sigmoid or ReLU).

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Experimental Results

The dataset used in this study consists of health-related attributes gathered from Wisconsin Breast Cancer Disease patients. These include various features such as age, blood pressure, cholesterol levels, smoking habits, and other medical indicators. The goal is to predict the likelihood of a patient developing Wisconsin Breast Cancer Disease (WBCD) based on these features. The dataset comprises 569 samples, with 32 features extracted for each instance. These samples were split into training and testing sets using a split between train and test 80% / 20% (see Fig. 3 and 4).

1) Model results: The performance of four ml algorithms — SVM, KNN, GaussianNB, and MLPClassifier — was evaluated. The results show the following performance across the models:

a) Support Vector Machine (SVM):

- Accuracy: Ranges from 97.23% to 97.78% at k=39.
- Precision: Ranges from 0.9742 to 0.9791.
- Recall: Ranges from 0.9711 to 0.9785.
- F1-Score: Ranges from 0.9705 to 0.9769.
- AUC-ROC: Ranges from 0.9965 to 0.9980.

b) k-Nearest Neighbors (KNN):

- Accuracy: Ranges from 96.01% to 96.21% at k=39.
- Precision: Ranges from 0.9614 to 0.9663.
- Recall: Ranges from 0.9584 to 0.9624.
- F1-Score: Ranges from 0.9587 to 0.9621.
- AUC-ROC: Ranges from 0.9930 to 0.9960.

c) Naive Bayes (GaussianNB):

- Accuracy: Ranges from 93.93% to 94.16% at k=39.
- Precision: Ranges from 0.9459 to 0.9464.
- Recall: Ranges from 0.9367 to 0.9427.
- F1-Score: Ranges from 0.9352 to 0.9389.
- AUC-ROC: Ranges from 0.9865 to 0.9886.

d) Neural Networks (MLPClassifier):

- Accuracy: Ranges from 97.92% to 98.13% at k=37.
- Precision: Ranges from 0.9805 to 0.9821.
- Recall: Ranges from 0.9782 to 0.9814.
- F1-Score: Ranges from 0.9776 to 0.9799.
- AUC-ROC: Ranges from 0.9985 to 1.0000.

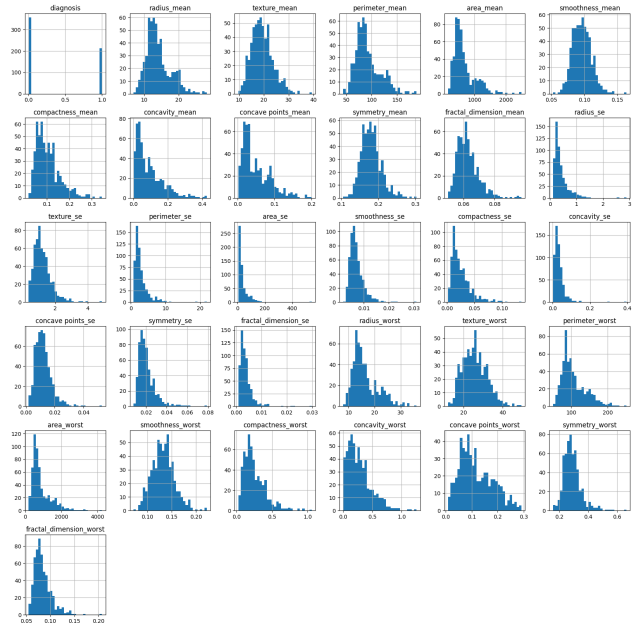


Fig. 3. Feature visualization result for WBCD.

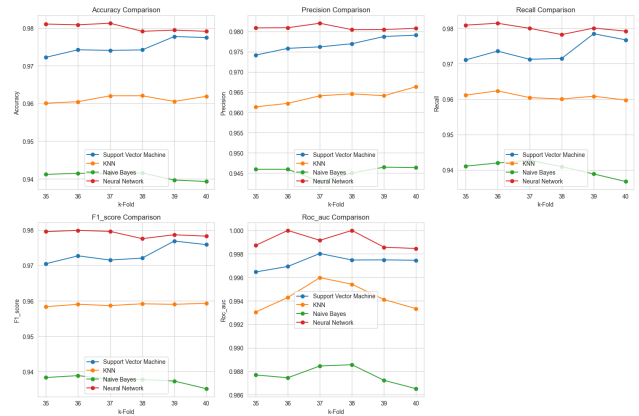


Fig. 4. Performance of model comparison WBCD.

B. Comparison of Results

The results of this experiment show clear performance differences between the models evaluated.

1) Best performing model: Neural networks (MLPClassifier):

- Accuracy: The Neural Networks model demonstrated the highest accuracy across all configurations, with

values consistently reaching above 97%, peaking at 98.13% at k=37.

- Precision and Recall: Both precision and recall scores were notably high, ranging from 0.9805 to 0.9821 for precision and 0.9782 to 0.9814 for recall. This indicates that the model is highly effective in both identifying positive cases and minimizing false negatives.
- F1-Score: The F1-Score was similarly high, ranging from 0.9776 to 0.9799, showing a strong balance between precision and recall.
- AUC-ROC: The model achieved AUC-ROC values between 0.9985 and 1.0000, with a perfect score at k=36, highlighting its excellent ability to distinguish between positive and negative classes.

2) Second best model: Support Vector Machine (SVM):

- Accuracy: SVM model achieved accuracy between 97.23% and 97.78%, which was very close to that of Neural Networks.
- Precision and Recall: Precision and recall for SVM were also impressive, with values range 0.9742 to 0.9791 for precision and 0.9711 to 0.9785 for recall.
- F1-Score: The F1-Score ranged from 0.9705 to 0.9769, demonstrating good balance.
- AUC-ROC: The AUC-ROC score ranged from 0.9965 to 0.9980, which is very high, though slightly lower than that of Neural Networks.

3) Third best model: k-Nearest Neighbors (KNN):

- Accuracy: KNN demonstrated accuracy between 96.01% and 96.21%, which was lower than both SVM and Neural Networks.
- Precision and Recall: for KNN range from 0.9614 to 0.9663 for precision and 0.9584 to 0.9624 for recall, which were still decent, though less effective than the top two models.
- F1-Score: The F1-Score ranged from 0.9587 to 0.9621, which indicates solid performance but still a gap from SVM and Neural Networks.
- AUC-ROC: The AUC-ROC ranged from 0.9930 to 0.9960, which was good but not as high as SVM and Neural Networks.

4) Least effective model: Naive Bayes (GaussianNB):

- Accuracy: Naive Bayes achieved the lowest accuracy, range from 93.93% to 94.16%.
- Precision and Recall: Precision and recall for Naive Bayes were still respectable, range from 0.9459 to 0.9464 for precision and 0.9367 to 0.9427 for recall, but these were lower than the other models.
- F1-Score: The F1-Score ranged from 0.9352 to 0.9389, which again was the lowest among the models.
- AUC-ROC: Naive Bayes had AUC-ROC scores between 0.9865 and 0.9886, which were decent but not as high as the other models.

VI. CONCLUSION

In this study, we evaluated four prominent ML algorithms — SVM, KNN, GaussianNB, and MLPClassifier — for predicting risk of WBCD based on a dataset consisting of 569 samples and 32 features. The performance of these models assessed using key metrics such F1-Score, Precision, Recall, Accuracy and AUC-ROC. Results from this analysis demonstrated that Neural Networks emerged as the most effective model, with superior performance across all metrics particular term of precision, AUC-ROC, recall, and accuracy. Specifically, Neural Networks achieved an accuracy range of 97.92% to 98.13%, and AUC-ROC values between 0.9985 to 1.0000, indicating that it was highly adept at distinguishing between WBCD and non-WBCD cases.

The SVM followed closely as the second-best performing model, with high accuracy (97.23% to 97.78%) and AUC-ROC scores (0.9965 to 0.9980), making it another highly reliable choice for Wisconsin Breast Cancer Disease risk prediction. However, k-Nearest Neighbors (KNN) and Naive Bayes (GaussianNB), although effective, exhibited slightly lower accuracy and AUC-ROC values, especially Naive Bayes, which struggled with a feature independence assumption that likely impacted its performance.

Given these findings, we conclude that Neural Networks (MLPClassifier) is the most reliable and accurate model for predicting Wisconsin Breast Cancer Disease risk among the algorithms tested. However, SVM can still serve as an effective alternative, particularly when computational efficiency and model interpretability are critical.

VII. FUTURE SCOPE

The present research provides a solid foundation for further enhancement in the field of Wisconsin BC Disease prediction using ML Comparative Analysis. Several directions for future work can be identified:

- Data Expansion and Feature Engineering: Incorporating additional feature such lifestyle factors, genetic data, or advanced imaging techniques help improve model accuracy. More diverse datasets, including different demographics and geographical populations, will ensure the generalizability of the model.
- Model Optimization: Hyperparameter tuning for the models used, including the number of layers and neurons in neural networks or kernel choice in SVM, could potentially enhance performance. Additionally, exploring ensemble techniques like Random Forest or Boosting could help raise prediction accuracy by combining backbone of multiple models.
- Real-time Applications & Deployment: For clinical use, models need to be deployed in real-time systems, where they can continuously learn and adapt to new data. Explainable AI (XAI) techniques will be crucial for gaining the trust of healthcare professionals by providing transparency in decision-making.
- Ethical Considerations: Applications of AI in healthcares grown, it is vital ensure that models are fair and the unbiased, particularly in diverse populations.

Ethical guidelines for AI deployment, along with ensuring patient data privacy, will be paramount in future research.

This research development more accurately, efficient, and trustworthy ML models for predicting WBCD, which can greatly benefit healthcare systems worldwide.

ACKNOWLEDGMENT

This is supported by School of IT, Washington University of Science and Technology, VA, USA

REFERENCES

- [1] Strelcenia, E., & Prakoonwit, S. (2023). Effective feature engineering and classification of breast cancer diagnosis: A comparative study. *BioMedInformatics*, 3(3), 616–631.
- [2] Albadr MAA, Ayob M, Tiun S, AL-Dhief FT, Arram A, & Khalaf S (2023). "Breast cancer diagnosis using the fast learning network algorithm." *Front. Oncol.*, 13:1150840. doi:10.3389/fonc.2023.1150840.
- [3] Hossin, M. M., Shamrat, F. J. M., Bhuiyan, M. R., Hira, R. A., Khan, T., & Molla, S. (2023). "Breast cancer detection: an effective comparison of different machine learning algorithms on the Wisconsin dataset." *Bulletin of Electrical Engineering and Informatics*, 12(4), 2446-2456.
- [4] Bhardwaj, A., Bhardwaj, H., Sakalle, A., Uddin, Z., Sakalle, M., & Ibrahim, W. (2022). "Tree-Based and Machine Learning Algorithm Analysis for Breast Cancer Classification." *Computational Intelligence and Neuroscience*, no. 1, 6715406.
- [5] Rasool, A., Bunterngchit, C., Tiejian, L., Islam, M. R., Qu, Q., & Jiang, Q. (2022). "Improved Machine Learning-Based Predictive Models for Breast Cancer Diagnosis." *Int. J. Environ. Res. Public Health*, 19, 3211. doi:10.3390/ijerph19063211.
- [6] Kadhim, R. R., & Kamil, M. Y. (2022). "Comparison of breast cancer classification models on Wisconsin dataset." *Int. J. Reconfigurable Embed. Syst.*, ISSN 2089-4864.
- [7] Ibrahim, S., Nazir, S., & Velastin, S. A. (2021). Feature selection using correlation analysis and principal component analysis for accurate breast cancer diagnosis. *Journal of Imaging*, 7(11), 225. doi:10.3390/jimaging7110225.
- [8] Abdulkareem, S. A., & Abdulkareem, Z. O. (2021). "An evaluation of the Wisconsin breast cancer dataset using ensemble classifiers and RFE feature selection." *Int. J. Sci., Basic Appl. Res.*, 55(2), 67-80.
- [9] Mohammed Amine Naji, Sanaa El Filali, Kawtar Aarika, EL Habib Benlahmar, Rachida AitAbdelouhahid, & Olivier Debauche (2021). "Machine Learning Algorithms For Breast Cancer Prediction And Diagnosis." *Procedia Computer Science*, 191, 487–492.
- [10] El Rahman, S. A. (2021). Predicting breast cancer survivability based on machine learning and feature selection algorithms: A comparative study. *Journal of Ambient Intelligence and Humanized Computing*, 12(8), 8585–8623. doi:10.1007/s12652-020-02667-5.
- [11] Panwar, N., Sharma, D., & Narang, N. (2020). "Breast cancer classification with machine learning classifier techniques." In *Proceedings of the 4th International Conference: Innovative Advancement in Engineering & Technology (IAET)*.
- [12] Assiri, A. S., Nazir, S., & Velastin, S. A. (2020). "Breast tumor classification using an ensemble machine learning method." *Journal of Imaging*, 6(6), 39.

A Machine Learning-Based Analysis of Tourism Recommendation Systems: Holistic Parameter Discovery and Insights

Raniah Alsahafi¹, Rashid Mehmood^{2*}, Saad Alqahtany³

Department of Computer Science-Faculty of Computing and Information Technology,
King Abdulaziz University, Jeddah 21589, Saudi Arabia¹

Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah 42351, Saudi Arabia^{2,3}

Abstract—Tourism is a cornerstone of the global economy, fostering cultural exchange and economic growth. As travelers increasingly seek personalized experiences, recommendation systems have become vital in guiding decision-making and enhancing satisfaction. These systems leverage advanced technologies such as IoT and machine learning to provide tailored suggestions for destinations, accommodations, and activities. This paper explores the transformative role of tourism recommendation systems (TRS) by analyzing data from 3,013 research articles published between 2000 and 2024 using a BERT-based methodology for semantic text representation and clustering. A robust software framework, integrating tools such as UMAP for dimensionality reduction and HDBSCAN for clustering, facilitated data modeling, cluster analysis, visualization, and the identification of key parameters in TRS. We discover a comprehensive taxonomy of 16 TRS parameters grouped into 4 macro-parameters. These include Personalized Tourism; Sustainability, Health and Resource Awareness; Adaptability & Crisis Management; and Social Impact & Cultural Heritage. These macro-parameters align with all three dimensions of the triple bottom line (TBL) -- social, economic, and environmental sustainability. The findings reveal key trends, highlight underexplored areas, and provide research-informed recommendations for developing more effective TRS. This paper synthesizes existing knowledge, identifies research gaps, and outlines directions for advancing TRS to support sustainable, personalized, and innovative travel solutions.

Keywords—*Recommendation Systems (RS); Tourism Recommendation Systems (TRS); big data analytics; machine learning; unsupervised learning; social; economic and environmental sustainability; Bidirectional Encoder Representations from Transformers (BERT); SDGs; literature review*

I. INTRODUCTION

The tourism industry has undergone a transformative evolution in recent decades, driven by advancements in digital technologies and the proliferation of data-driven systems [1]–[3]. With the global tourism market reaching unprecedented scales, travelers now demand personalized experiences that cater to their unique preferences and requirements [4], [5]. Traditional methods of tourism planning, relying on guidebooks and generic travel advice, have become insufficient in addressing the complexity and diversity of modern travel needs

[6], [7]. In this context, tourism recommendation systems (TRS) have emerged as pivotal tools, enabling travelers to navigate the abundance of information and make informed decisions about destinations [8], accommodations [9], activities [10], and other travel-related services [11], [12].

Recommendation systems in tourism leverage machine learning techniques [13] to deliver tailored suggestions to users [14]. By analyzing diverse datasets -- ranging from user preferences [15] and historical behaviors [16] to real-time contextual information [17] -- these systems aim to enhance user satisfaction and optimize travel experiences. Such systems play a dual role: improving the decision-making process for tourists [18] and offering a competitive edge to tourism providers by increasing customer engagement and loyalty [19].

The academic and industrial interest in tourism recommendation systems is growing, leading to a wealth of research addressing various aspects such as collaborative filtering [20], [21] content-based filtering [19], hybrid models [22], and the integration of emerging technologies such as deep learning [23], natural language processing (NLP) [23], [24], and generative AI [26], [27]. Despite the progress, challenges persist, including issues related to data sparsity [27], cold-start problems [28], interpretability of recommendations [29], and ethical concerns such as privacy and bias [30]. Addressing these challenges requires a more advanced and systematic approach to analyzing the TRS landscape.

In response, this paper presents a data-driven methodology that systematically extracts and classifies key research themes in TRS. While prior works have focused on specific aspects, our study addresses a significant gap (as outlined in Section II) by presenting a holistic taxonomy, meaning a structured and comprehensive classification of parameters and macro-parameters that captures the full breadth of TRS research. By analyzing an extensive dataset spanning 24 years, this study provides detailed insights into Personalized Tourism; Sustainability, Health & Resource Awareness; Adaptability and Crisis Management; and Social Impact & Cultural Heritage, helping to address gaps related to fragmented knowledge, evolving research trends, and emerging challenges. This structured approach allows for a deeper understanding of TRS developments while ensuring scalability and adaptability to future advancements.

Unlike existing literature, this paper incorporates a BERT-based (Bidirectional Encoder Representations from Transformers) methodology integrated with a machine learning pipeline to systematically analyze an extensive dataset spanning 24 years, from 2000 to 2024. BERT enables deep semantic analysis, allowing for more accurate extraction of key research themes and relationships across studies, thus overcoming limitations of traditional keyword-based methods. This dataset, constructed using the Scopus database, includes data from 3,013 research articles, refined through pre-processing steps such as tokenization, lemmatization, and duplicate removal. By systematically analyzing this large-scale dataset, our study ensures a broad yet structured understanding of TRS developments, reducing bias from smaller-scale literature reviews and enabling a more data-driven taxonomy.

To enable a robust analysis, we developed a comprehensive software framework consisting of four core modules: data acquisition and storage, preprocessing, modeling and parameter extraction, and validation with visualization. The system utilizes pre-trained BERT embeddings for contextual text representation, UMAP (Uniform Manifold Approximation and Projection) for dimensionality reduction, and HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) for identifying meaningful clusters. These clusters are analyzed using a class-based TF-IDF (Term Frequency–Inverse Document Frequency) scoring method to rank word importance and derive parameters, which are then categorized into macro-parameters through expert validation. Visualization techniques, including taxonomy and similarity matrices, are employed to facilitate interpretation and ensure clarity. Python libraries such as Plotly and Matplotlib were used extensively for these purposes, supporting both analysis and presentation.

By addressing these limitations and presenting a unified framework, this study not only synthesizes current knowledge but also identifies underexplored areas and interconnected themes. It seeks to offer valuable insights for researchers and practitioners aiming to develop more effective and ethical tourism recommendation solutions. Furthermore, it explores potential directions for future work, emphasizing the need for systems that enhance personalization while aligning with sustainable tourism practices and inclusivity.

The rest of this paper is organized as follows: Section II provides a detailed literature review, highlighting prior works and identifying key gaps. Section III outlines the methodology, including data collection, pre-processing, and the design of our analytical framework. Section IV presents the findings, detailing the quantitative and qualitative analyses of the results and the taxonomy of macro-parameters. Section V summarizes the state-of-the-art in tourism recommendation systems, the challenges facing the field, and directions for future work. Finally, Section VI concludes the paper with key insights and recommendations.

II. LITERATURE REVIEW

TRS have been examined through various focused studies addressing specific aspects of their development and application. Hamid et al. [31] emphasize the importance of robust data management in TRS, highlighting the integration of IoT and hybrid models to handle large datasets and enable real-

time, scalable recommendations. Santamaria-Granados et al. [32] focus on emotion recognition in TRS, using a scientometric review to explore how wearable devices and physiological sensors can enhance personalization by capturing emotional states. Menk et al. [33] examine the integration of social networks in TRS, showing how platforms such as Facebook and Twitter provide user-generated data to improve personalization and accuracy.

In addition to these focused studies, broader reviews of TRS offer general insights into approaches, developments, and issues within the field. Sarkar et al. [34] survey the evolution of TRS from traditional methods, such as collaborative filtering and content-based filtering, to advanced AI-driven techniques. The paper highlights how AI techniques can contribute to improving both traditional filtering methods and overall recommendation accuracy. It also emphasizes the need for innovation to address challenges such as system scalability and diverse data integration. Solano-Barliza et al. [35] offer a review of TRS trends and techniques, categorizing existing approaches and addressing challenges such as sparse data availability in emerging destinations. Their work highlights the importance of hybrid systems and data integration in enhancing system performance and user engagement. Khan et al. [36] review contextual suggestion systems within e-tourism, emphasizing the role of contextual factors such as location, time, and environmental conditions in tailoring recommendations. Their work highlights the importance of sustainability-focused recommendations while examining the methodologies and applications of context-aware TRS. Huda et al. [37] review smart tourism recommendation models, focusing on the integration of smart ICT technologies to enable real-time adaptability and personalization. Their work underscores the importance of enhancing tourist experiences through dynamic and context-aware recommendations.

The existing literature reviews provide valuable specific and general insights into TRS, offering important contributions to understanding the field. However, they fail to provide a holistic view due to limitations in scope, dataset size, methodology, findings, and the timeframes they cover. Except for Solano-Barliza et al. [35], all reviews are approximately three to five years old, making them less reflective of recent developments.

In contrast, our study addresses these limitations by employing a BERT-based methodology integrated into a novel software tool to systematically analyze a large and up-to-date dataset spanning 24 years (2000–2024). This approach allows for an in-depth exploration of the field, uncovering both a holistic taxonomy of parameters and macro-parameters and identifying interconnected themes and underexplored areas. Key areas covered include Personalized Tourism, Sustainability, Health and Resource Awareness, Social Impact & Cultural Heritage, and Adaptability & Crisis Management, addressing all three dimensions of the triple bottom line (TBL). The use of BERT enables advanced semantic analysis of academic literature, providing a significant improvement over traditional keyword-based methods. By offering a comprehensive framework for understanding TRS, this study not only addresses the shortcomings of prior works but also facilitates a deeper and more integrated understanding of the field, paving the way for future research and practical advancements.

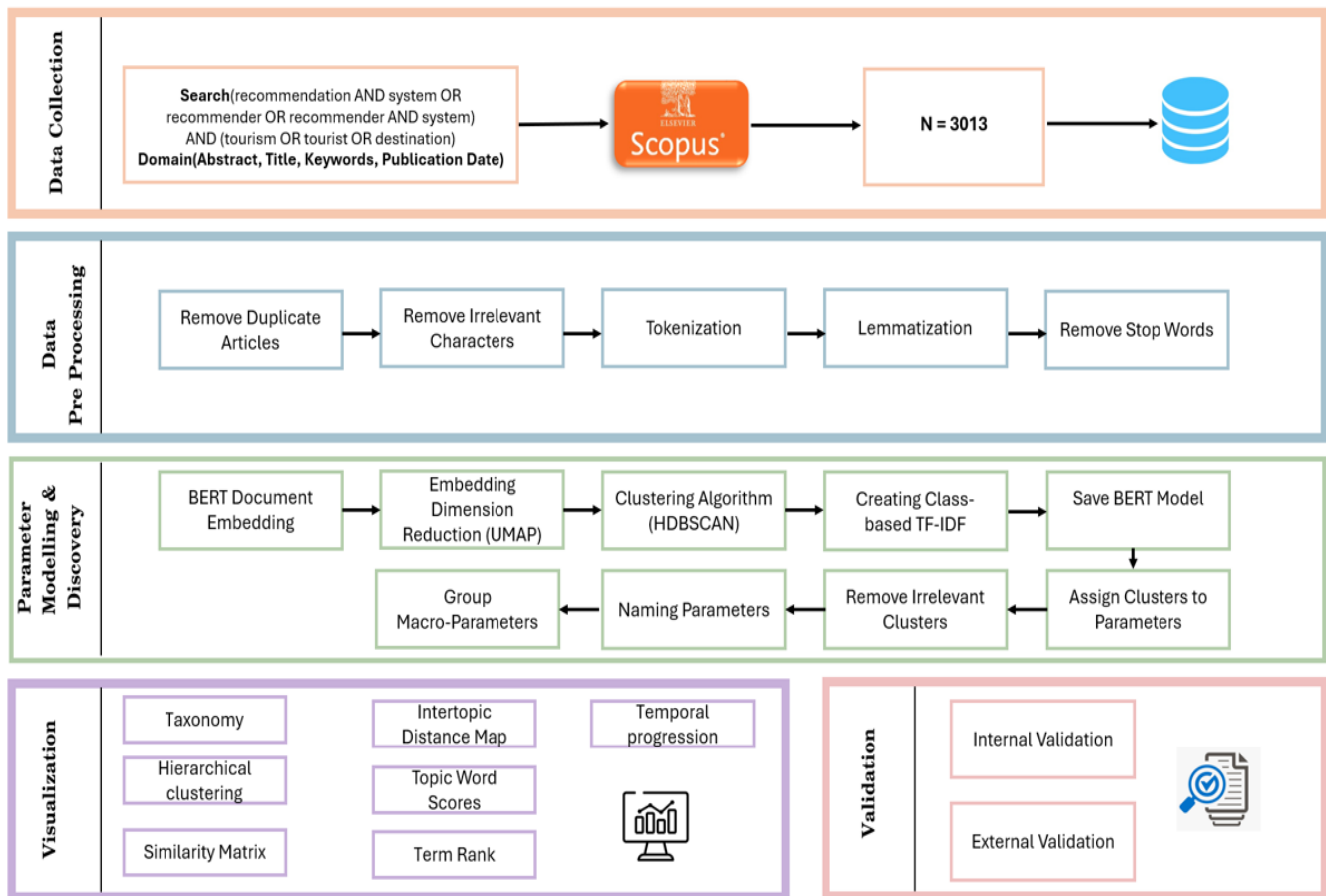


Fig. 1. System methodology and architecture.

III. METHODOLOGY

We present here the methodology and design of our system for machine-learning-based analysis and parameter discovery from academic literature on tourism recommendation systems. Further details of the broader methodology can be found in our earlier work [5]. The system architecture is illustrated in Fig. 1, detailing data collection, preprocessing, embedding creation, dimensionality reduction, clustering, and visualization.

To gather relevant data for this study, we formulated a specific search query: TITLE-ABS-KEY ((recommendation AND system OR recommender OR recommender AND system) AND (tourism OR tourist OR destination)). This query was used to extract data from Scopus, a comprehensive database containing an extensive range of academic literature across multiple disciplines. The initial dataset included publications from the years 2000 to 2024, with document types restricted to conference papers and journal articles, all written in English. After applying these filtering criteria, data from a total of 3,013 research articles were selected for further analysis.

The preprocessing phase involved several crucial steps to ensure the dataset was clean and suitable for further analysis. Initially, the collected articles were saved in CSV format and loaded into a Pandas DataFrame. Redundant and irrelevant entries were removed, including duplicate records and articles without abstracts. Subsequently, text-cleaning techniques were

applied, which involved eliminating unnecessary characters and performing tokenization. The tokenization process was executed using the 'gensim' Python package, which facilitates breaking text into meaningful words.

To enhance the quality of extracted information, we employed stop-word removal techniques using the Natural Language Toolkit (NLTK) predefined stop words list. Additionally, the text data was lemmatized using the WordNetLemmatizer, which converts words into their base forms while preserving their meanings. These preprocessing steps ensured that only meaningful and well-structured data were retained for the next phase.

For topic modeling, we implemented the BERTopic approach, which leverages transformer-based word embeddings to identify and cluster significant topics within the dataset. The first step in this process involved generating word embeddings using BERT, a deep learning-based model designed for natural language processing. Specifically, we used the 'distilbert-base-nli-mean-tokens' sentence transformer model to convert each document into a dense numerical representation.

Given that high-dimensional embeddings require dimensionality reduction for efficient processing, we applied the UMAP technique. The UMAP model was fine-tuned by setting key parameters such as $n_neighbors = 20$ and $n_components = 7$, which were determined to provide optimal clustering results.

Following this step, the HDBSCAN algorithm was used to group documents into clusters based on their semantic similarities. The most critical parameters for HDBSCAN, including `min_cluster_size` and `min_samples`, were optimized to ensure high-quality clustering.

To determine the significance of words within each topic, we calculated the class-based `c-TF-IDF` scores. These metric measures word importance by comparing the frequency of a term within a cluster to its overall occurrence across the entire corpus. The resulting scores enabled us to derive meaningful keyword-based descriptions for each identified topic.

The final number of clusters was determined through an iterative fine-tuning process using the `nr_topics` parameter in `BERTopic`, leading to a final selection of 20 distinct clusters, including one outlier cluster. After this refinement, each cluster was carefully evaluated by the authors to ensure relevance and coherence. This process, guided by the domain expertise of the authors, involved removing any irrelevant clusters if present, merging thematically similar ones when necessary, and assigning appropriate labels. The labeled clusters were then referred to as parameters, representing distinct research themes in tourism recommendation systems (TRS). To improve interpretability and provide a structured understanding of TRS research, these parameters were further aggregated into broader macro-parameters. The concept of parameters is designed to facilitate their integration into autonomous systems, enabling dynamic updates either periodically or in response to specific events, ensuring that the latest understanding of TRS or any related topic is continuously maintained. For further details, see our earlier work [38]–[40].

To ensure the reliability and validity of our results, we conducted both internal and external validation. Internal validation involved assessing the relevance of each document assigned to a given cluster, ensuring a meaningful relationship between texts and their respective clusters. External validation was carried out by comparing the parameters with established research findings. Multiple visualization tools, including, term ranking plots, hierarchical clustering dendrograms, and similarity matrices, were used to interpret the results effectively. These visuals were generated using Python libraries such as `Matplotlib` [41], `Seaborn` [42], and `Plotly` [43], enabling a detailed analysis of the dataset and topic structures.

Through this methodological approach, we successfully extracted, processed, and analyzed data from research articles to identify key parameters and macro-parameters within the domain of tourism recommendation systems. The rigorous validation and visualization techniques ensured the robustness of the findings, providing a reliable foundation for further analysis and interpretation.

IV. RESULTS

We now discuss the results obtained through our machine-learning-based tool, which used academic data to dissect the field of tourism recommendation systems and highlight its

cutting-edge advancements through quantitative and qualitative analysis.

A total of 2,991 documents were processed by the tool for clustering, following the removal of duplicates ($n = 22$). The model identified 19 clusters, one of which was irrelevant ($n = 11$) and removed along with an outlier cluster ($n = 1,156$), leaving 1,824 documents. The remaining 18 clusters were used to identify 16 parameters, with two parameters created by merging two clusters in each case due to thematic similarity. These parameters were further grouped into four macro-parameters: Personalized Tourism; Sustainability, Health & Resource Awareness; Adaptability and Crisis Management; and Social Impact and Cultural Heritage.

Fig. 2 illustrates the taxonomy of these parameters and macro-parameters identified by our BERT model. The first level of the taxonomy represents the macro-parameters, while the second level specifies the parameters, including their associated cluster numbers and document counts. For instance, “Travel Recommendation Algorithms (0, 1076)” refers to a parameter associated with cluster 0, containing 1,076 documents. These clustering results provided a structured framework for understanding and advancing the field.

These findings are detailed in the following sections. Section 4.A presents the quantitative analysis of the results, followed by a qualitative analysis of the four macro-parameters in Sections 4.B to 4.E.

A. Quantitative Analysis

Our analysis involves several quantitative methods, including term score, Intertopic Distance Map, hierarchical clustering, and a similarity matrix. While the clusters are linked to specific keywords, not all keywords effectively represent the parameters. As shown in Fig. 3, it reveals the required number of keywords to describe each cluster adequately. The analysis indicates that only the top seven to ten terms per parameter are truly representative.

Fig. 4 shows the hierarchical clustering of 19 recommender system clusters in tourism, organized by similarities in functionality or focus. Clusters 11, 12, 2, 5, and 7 formed a distinct group, along with Cluster 10, labeled as Personalized Tourism, reflecting high similarity.

Fig. 5 visualizes the similarity matrix between different parameters of recommendation systems in tourism, where dark blue represents the highest similarity score, and light green indicates the lowest. For example, cluster 0 (Travel Recommendation Algorithms) has a high similarity score with cluster 2 (Context-Aware Mobile Apps), as indicated by a darker cell at their intersection. This suggests that these two clusters share common features, making them closely related. Both focus on providing personalized recommendations to travelers based on their context, such as location, preferences, and behaviors. These visualizations highlight conceptual relationships between various themes of the field.

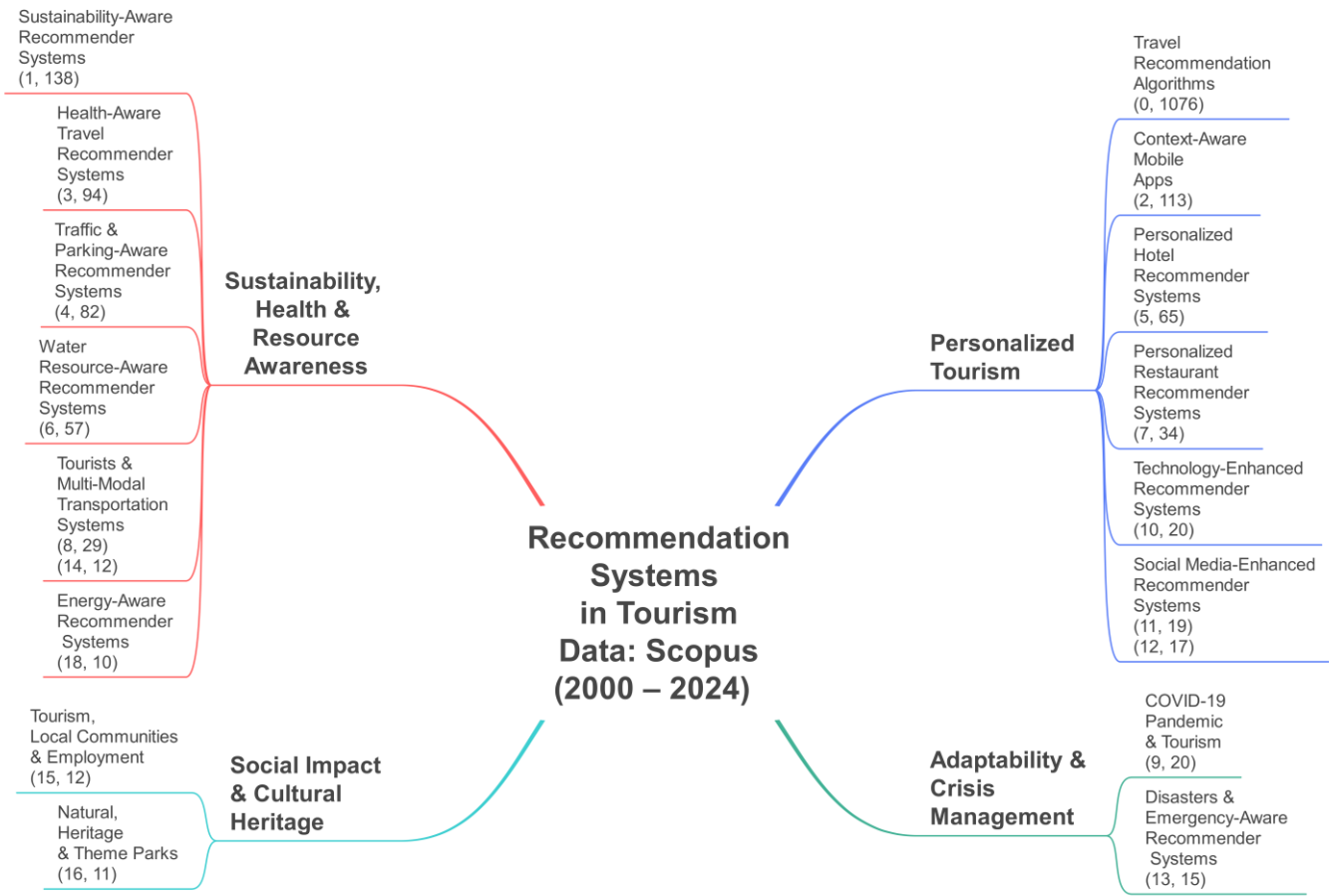


Fig. 2. Taxonomy for parameters of Tourism Recommendation System.

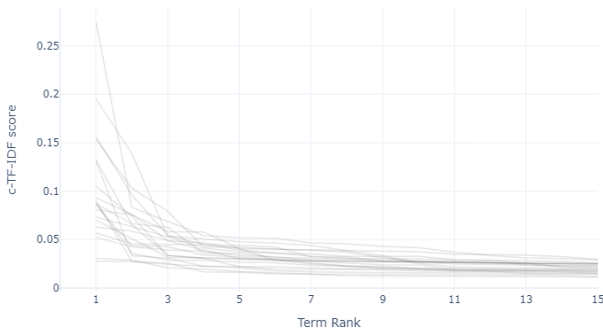


Fig. 3. Cluster term ranks.

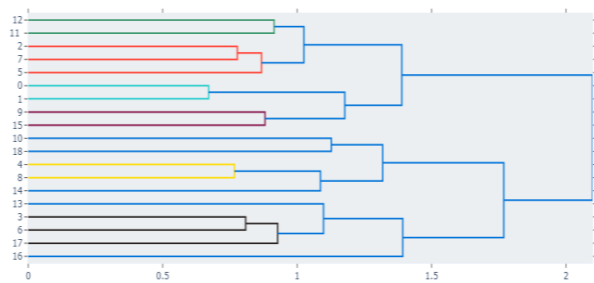


Fig. 4. Hierarchical clustering diagram.

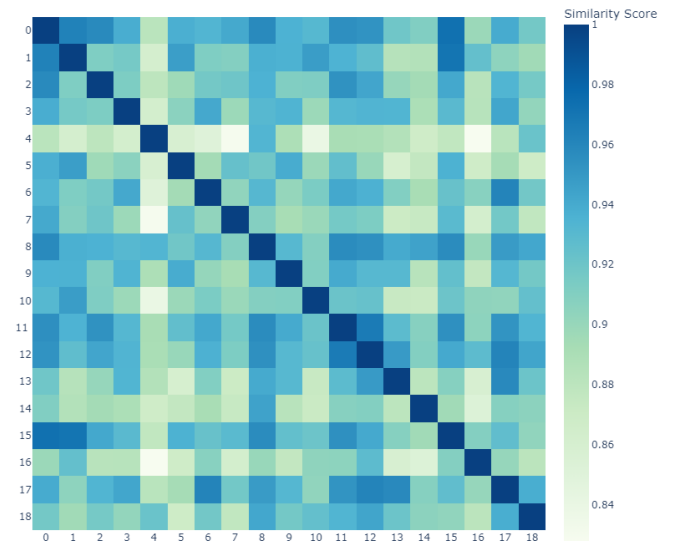


Fig. 5. Cluster similarity matrix.

Fig. 6 displays the top 10 keywords for each parameter, ranked using their c-TF-IDF importance scores. The 16 subfigures feature horizontal lines representing importance scores and vertical lines listing the parameter keywords.

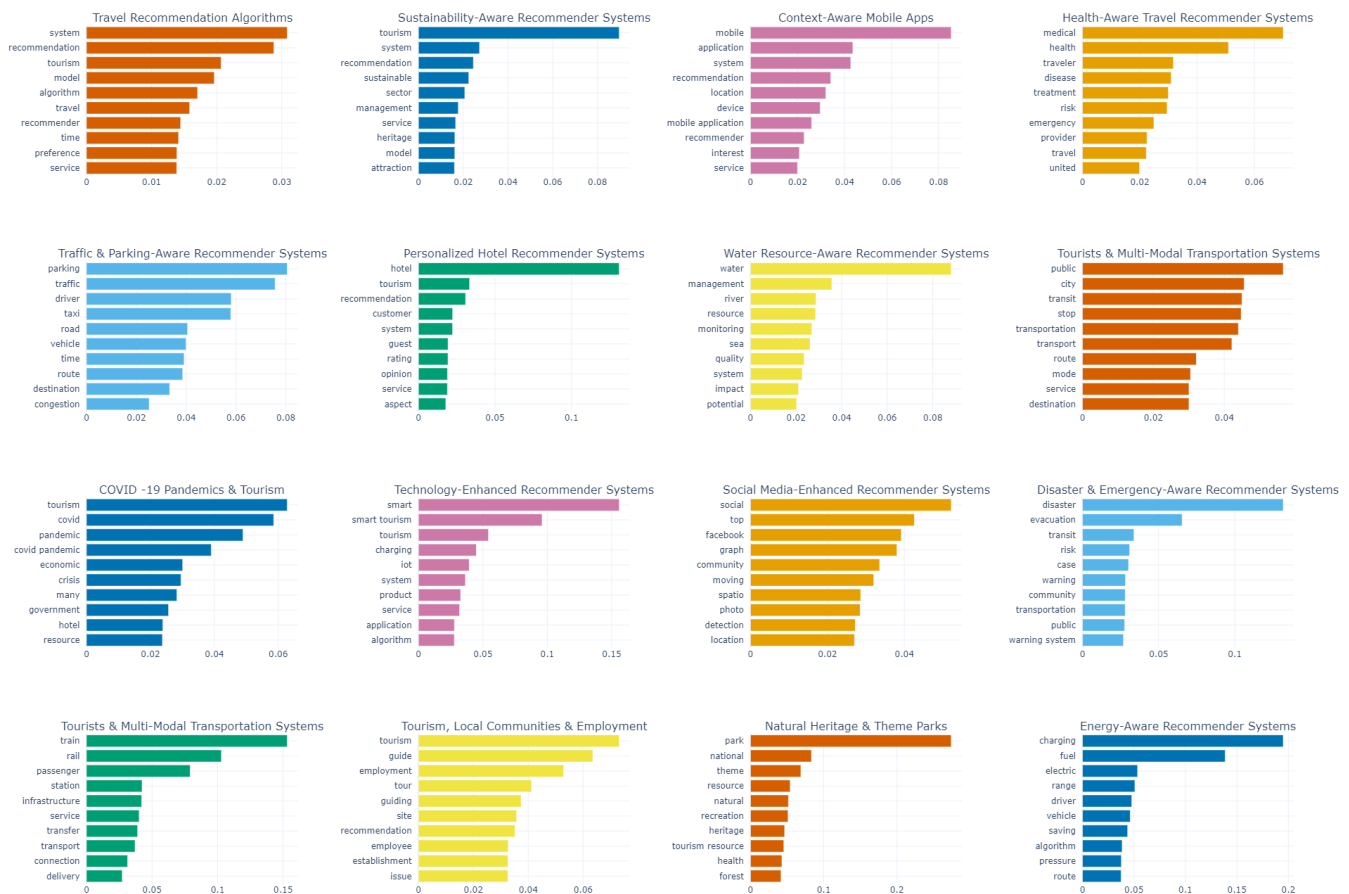


Fig. 6. Keyword c-TF-IDF scores for parameters.

Note that, to ensure a comprehensive and structured analysis, we included all 19 clusters in the quantitative analysis presented in Fig. 3 to Fig. 5. This analysis allows for the validation and refinement of the clusters by assessing their coherence, identifying irrelevant or redundant clusters, and examining their semantic relationships. Through this process, we systematically refined the clusters, ultimately leading to the discovery of parameters and macro-parameters. This approach not only ensures a coherent classification of research themes in TRS but also facilitates the extraction of the underlying knowledge structure and taxonomy of the field. However, Fig. 6, which displays the top 10 keywords for each parameter, ranked using their c-TF-IDF importance scores, contains subfigures for only 16 parameters because it reflects the post-cluster analysis phase, after the discovery and labeling of the final parameters.

B. Personalized Tourism

Personalized Tourism focuses on customizing the travel experience to individual preferences and contexts. It employs algorithms and technologies to offer customized travel suggestions, accommodations, dining options, and social experiences. By leveraging data from various sources, including mobile apps and social media, personalized tourism recommendation systems aim to enhance the satisfaction and convenience of travelers, making their experiences uniquely suited to their interests and needs. This macro encompasses six

key parameters designed to refine and customize the travel experience for individuals. The parameters are discussed below.

Travel Recommendation Algorithms are pivotal in analyzing traveler preferences to suggest destinations and activities. These recommendation algorithms and systems have notably evolved to utilize hybrid models that combine content-based and collaborative filtering with real-time, contextual, and social data inputs, greatly improving the personalization of travel suggestions [22], [44]. These systems effectively adapt to the dynamic nature of tourist needs, incorporating real-time data processing and location-based services [10], which are essential for travelers making spontaneous decisions or needing to adjust plans due to unforeseen circumstances [13], [45]–[47]. The incorporation of visual data and social media inputs further enhances the ability of these systems to deliver highly relevant and visually engaging recommendations that align with modern travel behaviors [48], [49].

Context-aware mobile Apps take situational factors into account, providing recommendations that are relevant to the user’s current location and circumstances. These mobile apps are increasingly pivotal, utilizing real-time data such as location, time, and user preferences to offer hyper-localized suggestions that enhance the immediacy and relevance of the information provided to the users [17], [50], [51]. These apps are not just enhancing user satisfaction but are also raising significant privacy and security considerations, necessitating the

development of privacy-by-design principles [52]. The use of advanced machine learning algorithms helps these apps learn from each interaction, progressively refining the recommendations to suit individual preferences better.

Personalized Hotel Recommender Systems focus on aligning accommodation options with the traveler's specific preferences, such as budget, amenities, and location [53]. Similarly, Personalized Restaurant Recommender Systems tailor dining suggestions to match the traveler's dietary needs and taste preferences [21]. These recommender systems leverage complex data analyses of user preferences, past behaviors, and online interactions to predict and meet individual needs, significantly enhancing customer satisfaction and operational efficiencies. These systems are increasingly using semantic analysis and contextual data to offer deep suggestions that consider dietary preferences, specific amenities, or desired experiences, which are critical in the highly competitive hospitality industry [25], [54].

Technology-Enhanced Recommender Systems leverage cutting-edge technology to offer highly personalized travel insights and options. The role of technology, particularly the integration of the Internet of Things (IoT) [55], and augmented reality (AR) [56], is transforming the tourist experience by enabling smarter, more connected environments that cater to the detailed needs and preferences of tourists. These technology-enhanced systems not only improve personalization but also ensure that services provided are efficient and timely, adapting to the individual's current context [57].

Social Media-Enhanced Recommender Systems utilize social networks to offer travel suggestions influenced by friends' recommendations, trends, or influencers, enhancing the travel planning process with a social dimension. These systems use data from platforms such as Instagram and Twitter to analyze user behaviors, preferences, and social interactions, enabling the delivery of personalized travel experiences that resonate well with users' tastes and preferences [58]. The use of big data analytics in these systems allows for a deeper understanding of individual preferences, significantly transforming the way destinations and activities are marketed and presented to potential tourists [59].

Collectively, these parameters aim to deliver a travel experience that is as unique as the travelers themselves, enhancing satisfaction through personalization. The integration of advanced algorithms, real-time data analytics, and user-centric technologies across these parameters is crafting a highly sophisticated landscape of personalized tourism recommendation systems. These systems are not only enhancing the travel experience by providing timely, relevant, and personalized recommendations but are also facing challenges such as data privacy, the need for continuous learning, and the integration of diverse technological solutions. The continuous evolution of these systems is crucial for sustaining innovation and growth in the tourism sector, promising a future where technology profoundly shapes the way tourist services are conceptualized and delivered.

C. Sustainability, Health and Resource Awareness

Sustainability, Health & Resource Awareness emphasizes promoting travel options that are environmentally sustainable, health-conscious, and resource-efficient. It includes systems designed to minimize the ecological footprint of tourism, recommend health-oriented travel options, and optimize the use of resources such as water and energy. These recommendation systems aim to support responsible tourism that respects the planet and the well-being of both tourists and local communities.

It encompasses six vital parameters designed to promote environmentally friendly and health-conscious travel practices. Sustainability-Aware Recommender Systems prioritize eco-friendly travel options, helping to reduce the ecological footprint of tourism. Health-Aware Travel Recommender Systems focus on health considerations, suggesting destinations and activities that align with travelers' health needs and preferences. Traffic & Parking-Aware Recommender Systems aim to alleviate congestion and improve efficiency by providing real-time traffic updates and parking information. Water Resource-Aware Recommender Systems emphasize conservation by promoting destinations and practices that minimize water usage. Energy-Aware Recommender Systems focus on reducing energy consumption through recommendations that favor energy-efficient options. Lastly, Tourists & Multi-Modal Transportation Systems facilitate seamless travel by integrating various modes of transportation, promoting ease of movement and reducing environmental impacts. Together, these systems strive to enhance travel experiences while being mindful of health, resource conservation, and sustainability.

Sustainability, Health and Resource Awareness in tourism recommendation systems represent a sophisticated integration of environmental stewardship, health optimization, and resource efficiency, underpinned by advanced technology and data analytics. The systems within this macro collectively address the triple bottom line of sustainability: environmental, economic, and social aspects. For example, Sustainability-Aware systems encourage visits to lesser-known sites [8], distributing economic benefits more evenly, and reducing environmental pressures on over-visited locations. Similarly, Health-Aware systems promote safety and health by integrating real-time health data, enhancing traveler well-being [60]. These approaches are mutually reinforcing. For instance, promoting less frequented sites helps manage the capacity and preserves the integrity of natural resources, aligning with the goals of Water Resource-Aware systems to manage environmental impacts effectively [61], [62].

Across all parameters, there is a heavy reliance on AI and machine learning to process real-time data and provide dynamic, context-sensitive recommendations [63]. This technological backbone enables Traffic and Parking-Aware systems to offer real-time routing adjustments just as Energy-Aware systems optimize resource use, demonstrating a cross-application of similar technological frameworks to solve varied problems within the tourism sector [64]. The integration of various types of data (e.g., traffic flow, water resource levels, energy consumption, health statistics) into a cohesive recommendation engine exemplifies a holistic approach to managing both expected and emergent challenges in tourism [65], [66].

The recommendation systems are increasingly adept at offering personalized travel suggestions that consider environmental conditions, health requirements, and individual preferences. For instance, Multi-Modal Transportation systems that recommend optimal travel modes based on user preferences and local traffic conditions overlap with Health-Aware systems that consider individual health needs [67]–[69]. This personalization extends to ensuring that recommendations are sensitive to local cultural norms and practices, enhancing the social sustainability of tourism by fostering respect and appreciation for local traditions, which is a focal point of both Sustainability-Aware and Health-Aware systems [11], [63].

The macro showcases a robust adaptability to global and local challenges, such as health emergencies or environmental crises. Systems rapidly adjust to new data, whether it's shifting health advisories during a pandemic or updating environmental regulations and conditions [70]–[73]. This resilience is critical in maintaining the trust and safety of tourists, ensuring that the tourism sector can quickly respond to and recover from disruptions, thereby supporting long-term sustainability goals [74].

The interconnected nature of these systems suggests significant policy implications, particularly in the need for coordinated action across health, environmental, and urban planning departments. The findings advocate for a policy framework that supports integrated data sharing and collaborative decision-making processes, enabling a more unified response to the multifaceted demands of sustainable tourism. Moreover, these systems serve as a model for other sectors, demonstrating how technology can bridge diverse data sources and operational goals to create more sustainable and resilient infrastructures.

In summary, Sustainability, Health & Resource Awareness illustrate a complex yet harmonious integration of multiple tourism-related aspects, driven by advanced technology and comprehensive data analytics. This integration not only enhances the efficiency and responsiveness of tourism recommendation systems but also significantly contributes to the broader objectives of sustainable development, public health, and economic equality.

D. Adaptability and Crisis Management

Adaptability and Crisis Management focuses on the flexibility of tourism recommendation systems to adapt to changing circumstances, such as global pandemics or natural disasters. It involves offering travel advice that considers safety guidelines, emergency preparedness, and the overall impact of crises on tourism. The goal is to ensure that travelers remain informed and safe, while also supporting the recovery and resilience of the tourism industry during and after crises.

Our analysis of the literature on Adaptability and Crisis Management reveals how both areas require robust, adaptable frameworks that integrate technology, policy, and localized approaches to manage crises effectively. It underscores the critical role of integrated, technology-driven solutions in managing and recovering from tourism-related crises. The effective combination of advanced recommender systems, supportive policy environments, and localized, customizable

strategies facilitate not only immediate crisis management but also contribute to the long-term sustainability and resilience of the tourism industry.

The COVID-19 pandemic forced a drastic rethinking of tourism practices, highlighting the need for adaptive crisis management strategies [75]. Key findings indicate a shift towards localized, safety-focused tourism, supported by digital and smart tourism solutions [70]. This shift was not merely reactive but also strategic, leveraging information systems to promote safer travel options and to adjust to a new tourism economy severely impacted by global restrictions [76]. The necessity for adaptation was evident in the rapid integration of sustainable practices and smart technologies, which were crucial in managing the downturn in traveler numbers [71].

Parallel to the pandemic's challenges, the use of recommender systems in managing disasters and emergencies showcases a proactive use of technology to enhance safety and efficiency [12]. These systems are crucial in real-time crisis management, offering optimized evacuation routes and strategies, and facilitating the rapid adaptation of transportation and local services to emergent needs [76]. The systems' capability to utilize local data for tailored community advisories further underscores the importance of localized responses in crisis management [77], [78].

Across both domains, the integration of advanced technology with supportive policy frameworks forms a backbone for effective crisis management [74]. During the pandemic, technology helped navigate economic shocks through targeted recovery strategies, while in disaster scenarios, technology optimized real-time responses [79]. This synergy suggests that robust, flexible digital infrastructures, capable of adapting to varied and sudden changes, are essential in sustaining tourism during crises [80], [81].

A recurring theme is the focus on localized and customized solutions, whether adapting tourism practices during a pandemic or responding to a localized disaster [79]. This approach maximizes the relevance and effectiveness of the response, illustrating how tailored information and strategies can significantly impact community resilience and crisis recovery [82]. Both areas highlight the need for systems that are not just reactive but highly adaptive and resilient [74]. The ongoing evolution of tourism practices in response to the pandemic, and the dynamic adjustments in disaster management, reflect a complex interplay between immediate crisis response and longer-term, strategic planning [71]. This dual approach is vital for the development of a resilient tourism sector capable of withstanding future crises.

E. Social Impact and Cultural Heritage

In tourism recommendation systems, Social Impact & Cultural Heritage encapsulate critical aspects including Tourism, Local Communities & Employment, and Natural Heritage and Theme Parks. These parameters highlight the intersection of tourism with local societal and environmental facets. For Tourism, Local Communities and Employment, recommendation systems play a vital role in promoting tourism experiences that benefit local economies and create job opportunities. Natural Heritage and Theme Parks focuses on

leveraging recommendation technologies to balance visitor numbers and preserve natural sites. These systems can suggest off-peak visit times and less-explored parks, thus managing foot traffic and reducing environmental impact.

The macro-parameter showcases a complex interaction in tourism between economic stability, cultural integrity, and environmental conservation [20]. It reveals the dynamic ways in which technology facilitates sustainable tourism, enhancing both local economies and cultural experiences [83], [84]. Recommendation systems play a crucial role in driving economic benefits by connecting tourists with local cultures and natural settings, diversifying income sources for local communities and stabilizing employment through culturally and ecologically respectful tourism [85], [86].

These systems manage both human resources, such as local guides [87], and natural resources, such as conservation sites, with a strategy that optimizes tourist flows to prevent overexploitation and ensures sustainable interactions between tourists and local resources [49]. They enhance the visitor experience by integrating local culture into tourism offerings, and educating tourists about local traditions and history while promoting respect and preservation for these cultures. In natural settings, the incorporation of cultural narratives enriches the visitor's engagement, fostering a deeper appreciation for both natural and cultural heritage [83], [88], [89]. Furthermore, directing tourists to less frequented sites mitigate environmental impacts on heavily visited locations, aiding conservation efforts and ensuring that economic benefits are broadly distributed [80], [90]. Recommendation systems also exemplify a commitment to social equity by proactively including diverse demographic groups in tourism employment, which benefits local economies and enhances the social fabric by making tourism more inclusive [89].

The focus on wellness tourism, such as forest bathing and other nature-based activities, not only promotes health benefits for tourists but also opens new economic avenues for local development, particularly in rural areas [91]. The systematic approach to enhancing both cultural and ecological tourism settings reflects a holistic view of tourism development, where various aspects of the tourist experience are assessed and integrated. This sophisticated approach supports long-term destination sustainability by meeting diverse visitor expectations and aligning with broader trends in health, conservation, and cultural engagement.

V. DISCUSSION

We now summarize the state of the art in tourism recommendation systems, the challenges facing the field, and directions for future work.

The integration of advanced technologies and data-driven strategies within the tourism sector is profoundly reshaping the landscape of travel recommendation systems, as demonstrated by the diverse range of macro-parameters analyzed. These macro-parameters reveal a complex and interconnected framework that aims to enhance tourist experiences [23], [92], promote sustainability [89], adapt to crises [12], [76]–[78], and preserve cultural heritage while fostering social impacts [47].

Central to tourism is the profound influence of technology in reshaping tourism experiences. Advanced algorithms and machine learning techniques play a critical role, enabling the delivery of highly personalized recommendations that adapt to the changing needs and contexts of travelers [65], [93]. Real-time data processing [94], [95], location-based services [96], and the integration of social media inputs are pivotal [97], [98], transforming the way travelers interact with destinations and services. These technological advancements are not just about enhancing user satisfaction; they also introduce significant considerations for privacy and security, prompting the development of sophisticated solutions such as privacy-by-design principles [30].

Simultaneously, there's a marked shift towards integrating sustainability and health into the core of tourism recommendation systems. These platforms strive to balance personalization with environmental consciousness and health awareness. For instance, sustainability-aware systems advocate for visiting lesser-known locales, thus alleviating the burden on popular destinations and promoting environmental preservation [8]. Health-aware systems enhance traveler safety by incorporating real-time health data [63], which is especially crucial in a post-pandemic world. This commitment extends to resource management, where systems intelligently recommend travel options that optimize energy use and minimize ecological impacts [99].

Adaptability and crisis management also feature prominently in this integrated approach. The recent global upheavals, such as the COVID-19 pandemic, have tested the flexibility and responsiveness of tourism infrastructures [71], [75], [100]. Recommendation systems have quickly adapted, offering solutions that prioritize local and safe travel options [101], thereby supporting the tourism industry's recovery. These systems demonstrate resilience [74], [82], adjusting to new health advisories and environmental conditions swiftly, and ensuring the safety and trust of tourists.

Furthermore, the role of recommendation systems in promoting cultural heritage and social impact cannot be overstated. By steering tourists towards culturally significant sites and engaging them with local traditions, these systems play a crucial role in cultural preservation [83]. They also support local economies by diversifying income sources and promoting equitable tourism practices [86]. Such initiatives not only enrich the visitor's experience but also ensure that tourism contributes positively to local communities [88].

A. Challenges and Future Work

The journey towards enabling smart tourism through recommendation systems is fraught with multifaceted challenges that must be addressed through innovative research and collaborative efforts. Personalized tourism recommendation systems aim to create a seamless travel experience by deeply understanding individual preferences and adapting to real-time contexts [94], [95]. The future of these systems lies in the development of advanced hybrid algorithms that blend machine learning with semantic technologies, enabling more precise and context-aware recommendations [44]. By leveraging extensive datasets and real-time environmental factors [102], these systems can offer dynamic, personalized travel itineraries that

adapt to changes in user mood and preferences [103]. However, achieving this level of personalization presents significant challenges. Ensuring the privacy and security of personal data is paramount, as these systems rely heavily on sensitive information [30]. Adherence to global privacy standards and regulatory compliance across jurisdictions is necessary to maintain user trust. Additionally, the complexity of managing heterogeneous data from diverse sources such as social media, user interactions, and IoT devices requires sophisticated data management strategies to ensure data integrity and timely recommendations [24], [51].

Sustainability and health awareness are critical dimensions that future tourism recommendation systems must incorporate [9]. Advanced machine learning models can predict environmental and social impacts with greater accuracy, providing actionable insights for sustainable tourism practices [81], [104], [105]. The optimization of multi-modal transportation systems using AI can enhance urban mobility and tourist satisfaction, while localized and personalized recommendations can align with local sustainability goals [68], [69], [106]. Despite these advancements, significant hurdles remain. Data availability and quality, particularly in underdeveloped regions, pose a major bottleneck. Ensuring the privacy of health data within travel recommenders necessitates advanced security frameworks such as federated learning, which allow for the private sharing of data insights while maintaining individual privacy. Additionally, the scalability of these systems to accommodate diverse tourist demographics and the integration of data from heterogeneous sources present technical challenges [104], [107].

The capability of tourism recommendation systems to adapt during crises is a critical area of future research. The development of advanced predictive analytics leveraging AI techniques such as machine learning and deep learning can enhance the accuracy and timeliness of crisis response strategies [71]. Robust models that handle dynamic, real-time data streams from diverse sources are essential for providing real-time, personalized travel recommendations during emergencies [100]. However, crisis-adaptive systems face several key challenges. The availability and reliability of data during crises are often compromised, affecting the operational efficiency of AI-driven tools. Ensuring these systems are robust enough to withstand data scarcity, potential cyber threats, and high user demand during critical periods is essential. Additionally, building and maintaining user trust through transparent systems that adhere to ethical standards and regulations is fundamental [75].

Tourism recommendation systems must also address the social impact and cultural heritage of travel destinations [3]. Advanced personalization techniques that cater to cultural interests can offer unique, culturally enriching experiences. AI can analyze extensive datasets on visitor interactions and cultural engagement patterns, enabling recommendations that resonate deeply with tourists [76]. Moreover, comprehensive tools for assessing the long-term impacts of tourism on local communities and cultural sites are necessary to ensure sustainable development [88], [89]. The challenges in this domain are significant. Data privacy and ethical concerns must be managed carefully to avoid violations and ensure that these systems benefit the communities they intend to support [30].

Ensuring cultural sensitivity and appropriateness in recommendations is crucial, as is avoiding the perpetuation of stereotypes or the misrepresentation of cultural heritages. Co-designing systems with input from local communities can help maintain cultural integrity. Balancing tourism growth with community welfare and avoiding over-reliance on technology that detracts from authentic cultural interactions presents additional challenges.

To address these challenges and realize the full potential of tourism recommendation systems, a holistic and integrated approach is required. This involves developing sophisticated and dynamic systems that leverage advanced AI techniques for deeper personalization, sustainability, and adaptability. It also necessitates fostering interdisciplinary collaborations among technology providers, tourism operators, local governments, and communities to enhance system responsiveness and efficacy. Ensuring transparency and adherence to ethical standards and regulations is crucial for building and maintaining user trust. Furthermore, aligning recommendations with sustainability goals and community welfare is essential to ensure that tourism growth does not negatively impact local residents or cultural heritage. By addressing these future research areas and challenges, the field can progress toward more resilient, responsive, and personalized tourism recommendation systems. This integrated approach will ensure that these systems contribute positively to both tourists and local communities, while respecting and enhancing the cultural heritage they aim to promote. The transformative potential of AI in tourism lies in its ability to create enriching, sustainable, and adaptive travel experiences that cater to the evolving needs and preferences of travelers worldwide.

VI. CONCLUSION

This paper aimed to develop and apply a machine-learning-based tool to analyze academic literature in the field of tourism recommendation systems, providing a structured taxonomy of parameters and macro-parameters to guide future research. The taxonomy offers a systematic framework for organizing the field, breaking it into clearly defined categories that facilitate understanding, highlight gaps, and direct future exploration. By identifying key parameters and their relationships, it enables researchers to prioritize areas for development, foster thematic alignment, and address emerging challenges. Despite the journal's page limit, we provided a detailed discussion of the parameters and macro-parameters, demonstrating their practical applications and aligning research priorities with real-world needs such as sustainability, health, and adaptability. These contributions provide a foundation for advancing the field and ensuring that future research and innovations are both cohesive and impactful.

This study provides a comprehensive analysis of TRS, but some limitations remain. Our analysis relies on academic literature from the Scopus database, which may exclude relevant industry reports, white papers, and non-English sources. Expanding the dataset to include other academic and non-academic sources could provide a broader perspective on TRS research. For future directions, integrating real-time data sources such as social media trends and user-generated content could enhance the adaptability of TRS. Additionally, incorporating

personalization techniques based on user intent, sentiment analysis, and contextual factors could improve recommendation accuracy. Finally, addressing ethical concerns such as data privacy, fairness, and algorithmic transparency is crucial for responsible TRS development.

ACKNOWLEDGMENT

This article is derived from a research grant funded by the Research, Development, and Innovation Authority (RDIA), Kingdom of Saudi Arabia, with grant number 12615-1U-2023-IU-R-2-1-EI-.

REFERENCES

- [1] P. D. Vecchio, G. Mele, V. Ndou, and G. Secundo, "Creating value from Social Big Data: Implications for Smart Tourism Destinations," *Inf. Process. Manag.*, vol. 54, no. 5, pp. 847–860, 2018, doi: 10.1016/j.ipm.2017.10.006.
- [2] A. Kontogianni and E. Alepis, "Smart tourism: State of the art and literature review for the last six years," *Array*, vol. 6, no. September 2019, p. 100020, 2020, doi: 10.1016/j.array.2020.100020.
- [3] W. Z. Li and H. Zhong, "Development of a smart tourism integration model to preserve the cultural heritage of ancient villages in Northern Guangxi," *Herit. Sci.*, vol. 10, no. 1, 2022, doi: 10.1186/s40494-022-00724-3.
- [4] H. Li, M. Hu, and G. Li, "Forecasting tourism demand with multisource big data," *Ann. Tour. Res.*, vol. 83, no. March, p. 102912, 2020, doi: 10.1016/j.annals.2020.102912.
- [5] R. Alsahafi, A. Alzahrani, and R. Mehmood, "Smarter Sustainable Tourism: Data-Driven Multi-Perspective Parameter Discovery for Autonomous Design and Operations," *Sustain.*, vol. 15, no. 5, 2023, doi: 10.3390/su15054166.
- [6] A. Fronzetti Colladon, B. Guardabascio, and R. Innarella, "Using social network and semantic analysis to analyze online travel forums and forecast tourism demand," *Decis. Support Syst.*, vol. 123, no. January, p. 113075, 2019, doi: 10.1016/j.dss.2019.113075.
- [7] L. Serrano, A. Ariza-Montes, M. Nader, A. Sianes, and R. Law, "Exploring preferences and sustainable attitudes of Airbnb green users in the review comments and ratings: a text mining approach," *J. Sustain. Tour.*, vol. 0, no. 0, pp. 1–19, 2020, doi: 10.1080/09669582.2020.1838529.
- [8] W. Buranasing, P. Meeklai, and P. Pattarathananan, "Recommendation System for Lesser-Known Places to Visit in Thailand," *ACM Int. Conf. Proceeding Ser.*, pp. 24–28, Nov. 2021, doi: 10.1145/3507473.3507477.
- [9] M. Nilashi et al., "Preference learning for eco-friendly hotels recommendation: A multi-criteria collaborative filtering approach," *J. Clean. Prod.*, vol. 215, pp. 767–783, 2019, doi: 10.1016/j.jclepro.2019.01.012.
- [10] K. Li and C. Qu, "Design and Implementation of Tourism Route Recommendation System Based on LBS," *IEEE Adv. Inf. Technol. Electron. Autom. Control Conf.*, pp. 2748–2751, 2021, doi: 10.1109/IAEAC50856.2021.9391036.
- [11] S. Jamshidi et al., "A hybrid health journey recommender system using electronic medical records," *CEUR Workshop Proc.*, vol. 2216, pp. 57–62, 2018.
- [12] A. Charef, Z. Jarir, and M. Quafafou, "Smart System for Emergency Traffic Recommendations : Urban Ambulance Mobility," *IJACSA Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 10, p. 2022, Accessed: Oct. 16, 2024. [Online]. Available: www.ijacsa.thesai.org.
- [13] M. UmmeSalma and C. Yashiga, "COLPOUSIT: A Hybrid Model for Tourist Place Recommendation based on Machine Learning Algorithms," *Proc. 5th Int. Conf. Trends Electron. Informatics, ICOEI 2021*, pp. 1743–1750, Jun. 2021, doi: 10.1109/ICOEI51242.2021.9452746.
- [14] P. Yuan, Q. Chen, Z. Wang, and J. Yang, "Personalized tourism recommendation algorithm integrating tag and emotional polarity analysis," *Proc. - 2022 10th Int. Conf. Adv. Cloud Big Data, CBD 2022*, pp. 163–168, 2022, doi: 10.1109/CBD58033.2022.00037.
- [15] C. Srisawatsakul and W. Boontarig, "Tourism Recommender System using Machine Learning Based on User's Public Instagram Photos," in *InCIT 2020 - 5th International Conference on Information Technology*, 2020, pp. 276–281, doi: 10.1109/InCIT50588.2020.9310777.
- [16] S. J. Miah, H. Q. Vu, J. Gammack, and M. McGrath, "A Big Data Analytics Method for Tourist Behaviour Analysis," *Inf. Manag.*, vol. 54, no. 6, pp. 771–785, 2017, doi: 10.1016/j.im.2016.11.011.
- [17] J. H. Yoon and C. Choi, "Real-Time Context-Aware Recommendation System for Tourism," *Sensors* 2023, Vol. 23, Page 3679, vol. 23, no. 7, p. 3679, Apr. 2023, doi: 10.3390/S23073679.
- [18] X.-K. Wang, S.-H. Wang, H.-Y. Zhang, J.-Q. Wang, and L. Li, "The Recommendation Method for Hotel Selection Under Traveller Preference Characteristics: A Cloud-Based Multi-Criteria Group Decision Support Model," *Gr. Decis. Negot.*, vol. 30, no. 6, pp. 1433–1469, 2021, doi: 10.1007/s10726-021-09735-0.
- [19] N. W. P. Y. Praditya, A. E. Permanasari, I. Hidayah, M. I. Zulfa, and S. Fauziati, "Collaborative and Content-Based Filtering Hybrid Method on Tourism Recommender System to Promote Less Explored Areas," *Int. J. Appl. Eng. Technol.*, vol. 4, no. 2, pp. 59–65, 2022.
- [20] Y. Cai, H. Gao, J. Liao, X. Li, Y. Xu, and J. Xiong, "A Personalized Recommendation Model based on Collaborative Filtering and Federated Learning for Cultural Tourism Attractions in Fujian-Taiwan," *Proc. - 2023 Int. Conf. Softw. Syst. Eng. ICoSSE 2023*, pp. 69–77, 2023, doi: 10.1109/ICOSSE58936.2023.00020.
- [21] R. Alabduljabbar, "Matrix Factorization Collaborative-Based Recommender System for Riyadh Restaurants: Leveraging Machine Learning to Enhance Consumer Choice," *Appl. Sci.* 2023, Vol. 13, Page 9574, vol. 13, no. 17, p. 9574, Aug. 2023, doi: 10.3390/AP13179574.
- [22] Y. Hao and N. Song, "Dynamic Modeling and Analysis of Multidimensional Hybrid Recommendation Algorithm in Tourism Itinerary Planning under the Background of Big Data," *Discret. Dyn. Nat. Soc.*, vol. 2021, no. 1, p. 9957785, Jan. 2021, doi: 10.1155/2021/9957785.
- [23] J. C. Cepeda-Pacheco and M. C. Domingo, "Deep learning and Internet of Things for tourist attraction recommendations in smart cities," *Neural Comput. Appl.*, vol. 34, no. 10, pp. 7691–7709, May 2022, doi: 10.1007/S00521-021-06872-0/TABLES/7.
- [24] O. Artemenko, V. Pasichnyk, N. Kusanets, and K. Shuneych, "Using sentiment text analysis of user reviews in social media for e-tourism mobile recommender systems," in *CEUR Workshop Proceedings*, 2020, vol. 2604, pp. 259–271, [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85085173260&partnerID=40&md5=eb679e9bbfea37208660e8240cda1c7e>.
- [25] M. Godakandage and S. Thelijjagoda, "Aspect Based Sentiment Oriented Hotel Recommendation Model Exploiting User Preference Learning," in *2020 IEEE 15th International Conference on Industrial and Information Systems, ICIIS 2020 - Proceedings*, 2020, pp. 409–414, doi: 10.1109/ICIIS51140.2020.9342744.
- [26] K. Al Farami, F. Nafis, B. Aghoutane, A. Yahyaouy, J. Riffi, and A. Sabri, "Hybrid recommender system for tourism based on big data and AI: A conceptual framework," *Big Data Min. Anal.*, vol. 4, no. 1, pp. 47–55, Mar. 2021, doi: 10.26599/BDMA.2020.9020015.
- [27] M. Nilashi et al., "A Hybrid Method to Solve Data Sparsity in Travel Recommendation Agents Using Fuzzy Logic Approach," *Math. Probl. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/7372849.
- [28] Z. Bahramian, R. Ali Abbaspour, and C. Claramunt, "A Cold Start Context-Aware Recommender System for Tour Planning Using Artificial Neural Network and Case Based Reasoning," *Mob. Inf. Syst.*, vol. 2017, 2017, doi: 10.1155/2017/9364903.
- [29] J. Gao, P. Peng, F. Lu, C. Claramunt, and Y. Xu, "Towards travel recommendation interpretability: Disentangling tourist decision-making process via knowledge graph," *Inf. Process. Manag.*, vol. 60, no. 4, p. 103369, Jul. 2023, doi: 10.1016/J.IPM.2023.103369.
- [30] C. Wang, Y. Zheng, J. Jiang, and K. Ren, "Toward Privacy-Preserving Personalized Recommendation Services," *Engineering*, vol. 4, no. 1, pp. 21–28, Feb. 2018, doi: 10.1016/J.ENG.2018.02.005.
- [31] R. A. Hamid et al., "How smart is e-tourism? A systematic review of smart tourism recommendation system applying data management,"

- Comput. Sci. Rev., vol. 39, p. 100337, 2021, doi: 10.1016/j.cosrev.2020.100337.
- [32] L. Santamaria-Granados, J. F. Mendoza-Moreno, and G. Ramirez-Gonzalez, "Tourist Recommender Systems Based on Emotion Recognition—A Scientometric Review," *Futur. Internet* 2021, Vol. 13, Page 2, vol. 13, no. 1, p. 2, Dec. 2020, doi: 10.3390/FI13010002.
- [33] A. Menk, L. Sebastia, and R. Ferreira, "Recommendation Systems for Tourism Based on Social Networks: A Survey," Mar. 2019, Accessed: Sep. 21, 2024. [Online]. Available: <https://arxiv.org/abs/1903.12099v1>.
- [34] J. L. Sarkar, A. Majumder, C. R. Panigrahi, S. Roy, and B. Pati, "Tourism recommendation system: a survey and future research directions," *Multimed. Tools Appl.*, vol. 82, no. 6, pp. 8983–9027, Mar. 2023, doi: 10.1007/S11042-022-12167-W/METRICS.
- [35] A. Solano-Barliza et al., "Recommender systems applied to the tourism industry: a literature review," *Cogent Bus. Manag.*, vol. 11, no. 1, p. 2024, doi: 10.1080/23311975.2024.2367088.
- [36] H. U. Rehman Khan, C. Kim Lim, M. F. Ahmed, K. L. Tan, and M. Bin Mokhtar, "Systematic Review of Contextual Suggestion and Recommendation Systems for Sustainable e-Tourism," *Sustain.* 2021, Vol. 13, Page 8141, vol. 13, no. 15, p. 8141, Jul. 2021, doi: 10.3390/SU13158141.
- [37] C. Huda, A. Ramadhan, A. Trisetayrso, E. Abdurachman, and Y. Heryadi, "Smart Tourism Recommendation Model: A Systematic Literature Review," *IJACSA Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 12, p. 2021, Accessed: Sep. 21, 2024. [Online]. Available: www.ijacsa.thesai.org.
- [38] A. A. Alaql, F. Alqurashi, and R. Mehmood, "Multi-generational labour markets: data-driven discovery of multi-perspective system parameters using machine learning," *Sci. Prog.*, vol. 106, no. 4, Nov. 2023, doi: 10.1177/00368504231213788.
- [39] N. Alahmari, R. Mehmood, A. Alzahrani, T. Yigitcanlar, and J. M. Corchado, "Autonomous and Sustainable Service Economies: Data-Driven Optimization of Design and Operations through Discovery of Multi-Perspective Parameters," *Sustain.* 2023, Vol. 15, Page 16003, vol. 15, no. 22, p. 16003, Nov. 2023, doi: 10.3390/SU152216003.
- [40] S. Alswedani, R. Mehmood, I. Katib, and S. M. Altowaijri, "Psychological Health and Drugs: Data-Driven Discovery of Causes, Treatments, Effects, and Abuses," Jan. 2023, doi: 10.20944/PREPRINTS202301.0415.V1.
- [41] "Histograms — Matplotlib 3.6.0 documentation." <https://matplotlib.org/stable/gallery/statistics/hist.html> (accessed Oct. 09, 2022).
- [42] M. Waskom, "seaborn: statistical data visualization," *J. Open Source Softw.*, vol. 6, no. 60, p. 3021, Apr. 2021, doi: 10.21105/JOSS.03021.
- [43] "Plotly: Low-Code Data App Development." <https://plotly.com/> (accessed Oct. 23, 2022).
- [44] M. V Murali, T. G. Vishnu, and N. Victor, "A Collaborative Filtering based Recommender System for Suggesting New Trends in Any Domain of Research," in 2019 5th International Conference on Advanced Computing and Communication Systems, ICACCS 2019, 2019, pp. 550–553, doi: 10.1109/ICACCS.2019.8728409.
- [45] J. T. Joseph and N. Santiago, "An Intelligent Image Based Recommendation System for Tourism," 2021 IEEE Conf. Norbert Wiener 21st Century Being Hum. a Glob. Village, 21CW 2021, Jul. 2021, doi: 10.1109/21CW48944.2021.9532512.
- [46] R. Sharma, S. Rani, and S. Tanwar, "Machine learning algorithms for building recommender systems," in 2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019, 2019, pp. 785–790, doi: 10.1109/ICCS45141.2019.9065538.
- [47] C. Trattner, A. Oberegger, L. Marinho, and D. Parra, "Investigating the utility of the weather context for point of interest recommendations," *Inf. Technol. Tour.*, vol. 19, no. 1–4, pp. 117–150, Jun. 2018, doi: 10.1007/S40558-017-0100-9/FIGURES/12.
- [48] W. Grossmann, M. Sertkan, J. Neidhardt, and H. Werthner, "Pictures as a tool for matching tourist preferences with destinations," *Pers. Human-Computer Interact.*, pp. 337–353, Aug. 2023, doi: 10.1515/9783110988567-013.
- [49] L. Zhang et al., "Visual analytics of route recommendation for tourist evacuation based on graph neural network," *Sci. Reports* 2023 131, vol. 13, no. 1, pp. 1–15, Oct. 2023, doi: 10.1038/s41598-023-42862-z.
- [50] C. S. Fun, Z. F. Zaaba, and A. S. Ali, "Usable Tourism Application: Malaysia Attraction Travel Application (MATA)," 2021 Int. Conf. Inf. Technol. ICIT 2021 - Proc., pp. 888–892, Jul. 2021, doi: 10.1109/ICIT52682.2021.9491757.
- [51] S. Missaoui, F. Kassem, M. Viviani, A. Agostini, R. Faiz, and G. Pasi, "LOOKER: a mobile, personalized recommender system in the tourism domain based on social media user-generated content," *Pers. Ubiquitous Comput.*, vol. 23, no. 2, pp. 181–197, 2019, doi: 10.1007/s00779-018-01194-w.
- [52] P. S. Efraimidis, G. Drosatos, A. Arampatzis, G. Stamatelatos, and I. N. Athanasiadis, "A privacy-by-design contextual suggestion system for tourism," *J. Sens. Actuator Networks*, vol. 5, no. 2, 2016, doi: 10.3390/jsan5020010.
- [53] H. C. Wang, A. Justitia, and C. W. Wang, "AsCDPR: a novel framework for ratings and personalized preference hotel recommendation using cross-domain and aspect-based features," *Data Technol. Appl.*, vol. ahead-of-p, no. ahead-of-print, 2023, doi: 10.1108/DTA-03-2023-0101/FULL/XML.
- [54] C. Dursun and A. Ozcan, "Sentiment-enhanced Neural Collaborative Filtering Models Using Explicit User Preferences," *HORA 2023 - 2023 5th Int. Congr. Human-Computer Interact. Optim. Robot. Appl. Proc.*, 2023, doi: 10.1109/HORA58378.2023.10156719.
- [55] W. Wang et al., "Realizing the Potential of Internet of Things for Smart Tourism with 5G and AI," *IEEE Netw.*, vol. 34, no. 6, pp. 295–301, 2020, doi: 10.1109/MNET.011.2000250.
- [56] S. Kalloori, R. Chalumattu, F. Yang, S. Klingler, and M. Gross, "Towards Recommender Systems in Augmented Reality for Tourism," *Springer Proc. Bus. Econ.*, pp. 267–272, 2023, doi: 10.1007/978-3-031-25752-0_29/FIGURES/1.
- [57] H. Hu and C. Li, "Smart tourism products and services design based on user experience under the background of big data," *Soft Comput.*, vol. 27, no. 17, pp. 12711–12724, Sep. 2023, doi: 10.1007/S00500-023-08851-0/METRICS.
- [58] S. Han, C. Liu, K. Chen, D. Gui, and Q. Du, "A Tourist Attraction Recommendation Model Fusing Spatial, Temporal, and Visual Embeddings for Flickr-Geotagged Photos," *ISPRS Int. J. Geo-Information* 2021, Vol. 10, Page 20, vol. 10, no. 1, p. 20, Jan. 2021, doi: 10.3390/IJGI10010020.
- [59] K. K. Ranga, C. K. Nagpal, and V. Vedpal, "Trip Planner: A Big Data Analytics Based Recommendation System for Tourism Planning," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 11, no. 3s, pp. 159–174, Accessed: Nov. 01, 2024. [Online]. Available: https://www.academia.edu/102022310/Trip_Planner_A_Big_Data_Analytics_Based_Recommendation_System_for_Tourism_Planning.
- [60] B. KC et al., "Types and outcomes of pharmacist-managed travel health services: A systematic review," *Travel Med. Infect. Dis.*, vol. 51, p. 102494, Jan. 2023, doi: 10.1016/J.TMAID.2022.102494.
- [61] X. Zhou, D. Zhang, J. Tian, and M. Su, "Low-Carbon Tour Route Algorithm of Urban Scenic Water Spots Based on an Improved DIANA Clustering Model," *Water (Switzerland)*, vol. 14, no. 9, 2022, doi: 10.3390/w14091361.
- [62] L. Orlando, L. Ortega, and O. Defeo, "Perspectives for sandy beach management in the Anthropocene: Satellite information, tourism seasonality, and expert recommendations," *Estuar. Coast. Shelf Sci.*, vol. 262, p. 107597, Nov. 2021, doi: 10.1016/J.ECSS.2021.107597.
- [63] M. Torres-Ruiz, R. Quintero, G. Guzman, and K. T. Chui, "Healthcare Recommender System Based on Medical Specialties, Patient Profiles, and Geospatial Information," *Sustain.* 2023, Vol. 15, Page 499, vol. 15, no. 1, p. 499, Dec. 2022, doi: 10.3390/SU15010499.
- [64] M. F. Jaafar Sidek, F. A. Bakri, A. A. Kadar Hamsa, N. N. Aziemah Nik Othman, N. M. Noor, and M. Ibrahim, "Socio-economic and Travel Characteristics of transit users at Transit-oriented Development (TOD) Stations," *Transp. Res. Procedia*, vol. 48, pp. 1931–1955, Jan. 2020, doi: 10.1016/J.TRPRO.2020.08.225.
- [65] S. P. R. Asaithambi, R. Venkatraman, and S. Venkatraman, "A Thematic Travel Recommendation System Using an Augmented Big Data

- Analytical Model,” *Technol.* 2023, Vol. 11, Page 28, vol. 11, no. 1, p. 28, Feb. 2023, doi: 10.3390/TECHNOLOGIES11010028.
- [66] Bhumika and D. Das, “MARRS: A Framework for multi-objective risk-aware route recommendation using Multitask-Transformer,” *RecSys 2022 - Proc. 16th ACM Conf. Recomm. Syst.*, pp. 360–368, Sep. 2022, doi: 10.1145/3523227.3546787/SUPPL_FILE/10.11453523227.3546787.MP4.
- [67] J. W. Adie, W. Graham, R. O’Donnell, and M. Wallis, “Patient presentations to an after-hours general practice, an urgent care clinic and an emergency department on Sundays: a comparative, observational study,” *J. Health Organ. Manag.*, vol. 37, no. 1, pp. 96–115, Apr. 2023, doi: 10.1108/JHOM-08-2021-0308/FULL/PDF.
- [68] H. Liu, T. Li, R. Hu, Y. Fu, J. Gu, and H. Xiong, “Joint Representation Learning for Multi-Modal Transportation Recommendation,” *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, pp. 1036–1043, Jul. 2019, doi: 10.1609/AAAI.V33I01.33011036.
- [69] M. A. Mondal and Z. Reheza, “Designing of A* Based Route Recommendation Service for Multimodal Transportation System in Smart Cities,” *Iran. J. Sci. Technol. - Trans. Civ. Eng.*, vol. 47, no. 1, pp. 609–625, Feb. 2023, doi: 10.1007/S40996-022-00948-0/METRICS.
- [70] S. Ghosh, I. S. Misra, and T. Chakraborty, “Developing an Application for Intelligent Transportation System for Emergency Health Care,” *2022 IEEE Calcutta Conf. CALCON 2022 - Proc.*, pp. 39–43, 2022, doi: 10.1109/CALCON56258.2022.10060474.
- [71] E. Brazález, H. Macià, G. Díaz, V. Valero, and J. Boubeta-Puig, “PITS: An Intelligent Transportation System in pandemic times,” *Eng. Appl. Artif. Intell.*, vol. 114, p. 105154, Sep. 2022, doi: 10.1016/J.ENGAPPAL.2022.105154.
- [72] S. Gkevreki, V. Fiska, S. Nikolopoulos, and I. Kompatsiaris, “Enhancing Sustainability in Health Tourism through an Ontology-Based Booking Application for Personalized Packages,” *Sustain.* 2024, Vol. 16, Page 6505, vol. 16, no. 15, p. 6505, Jul. 2024, doi: 10.3390/SU16156505.
- [73] R. Roy and L. W. Dietz, “Modeling physiological conditions for proactive tourist recommendations,” *ABIS 2019 - Proc. 23rd Int. Work. Pers. Recomm. Web Beyond*, pp. 25–27, Sep. 2019, doi: 10.1145/3345002.3349289.
- [74] L. Chapungu, K. Dube, and I. Mensah, “African Tourism Destinations in the Post-COVID-19 Era: Conclusions, Recommendations and Implications,” *COVID-19, Tour. Destin. Prospect. Recover. an African Perspect.* Vol. 2, vol. 2, pp. 263–277, Jan. 2023, doi: 10.1007/978-3-031-24655-5_14.
- [75] G. Glukhov and I. Derevitskii, “Points-of-Interest Recommendation Algorithms for a COVID-19 Restrictions Scenario in the Catering Industry,” *15th IEEE Int. Conf. Appl. Inf. Commun. Technol. AICT 2021*, 2021, doi: 10.1109/AICT52784.2021.9620251.
- [76] R. Pitakaso et al., “Designing safety-oriented tourist routes for heterogeneous tourist groups using an artificial multi-intelligence system,” *J. Ind. Prod. Eng.*, vol. 40, no. 7, pp. 589–609, Oct. 2023, doi: 10.1080/21681015.2023.2248144.
- [77] B. Yang et al., “A Novel Heuristic Emergency Path Planning Method Based on Vector Grid Map,” *ISPRS Int. J. Geo-Information 2021*, Vol. 10, Page 370, vol. 10, no. 6, p. 370, May 2021, doi: 10.3390/IJGI10060370.
- [78] M. B. Younes, “Safe and Efficient Advising Traffic System Around Critical Road Scenarios,” *Int. J. Intell. Transp. Syst. Res.*, vol. 21, no. 1, pp. 229–239, Apr. 2023, doi: 10.1007/S13177-023-00349-1/METRICS.
- [79] K. V. Daya Sagar, P. S. G. Arunasri, S. Sakamuri, J. Kavitha, and D. B. K. Kamesh, “Collaborative Filtering and Regression Techniques based location Travel Recommender System based on social media reviews data due to the COVID-19 Pandemic,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 981, no. 2, p. 022009, Dec. 2020, doi: 10.1088/1757-899X/981/2/022009.
- [80] T. R. Legrand, K. M. R. A. I. Bandara, J. A. D. Stefania Crishani, L. W. P. Uvindu, N. Amarasena, and D. Kasthurirathna, “TRIPORA: Intelligent Machine Learning Solution for Sri Lanka Touring Access and Updates,” *4th Int. Conf. Adv. Comput. ICAC 2022 - Proceeding*, pp. 24–29, 2022, doi: 10.1109/ICAC57685.2022.10025139.
- [81] S. Becken and J. Loehr, “Asia-Pacific tourism futures emerging from COVID-19 recovery responses and implications for sustainability,” *J. Tour. Futur.*, vol. 9, no. 1, pp. 35–48, Mar. 2023, doi: 10.1108/JTF-05-2021-0131/FULL/PDF.
- [82] I. B. Shem-Tov and S. Bekhor, “Evacuation Scenario Simulator with Location-Based Social Network Data Support,” *Transp. Res. Procedia*, vol. 69, pp. 69–76, Jan. 2023, doi: 10.1016/J.TRPRO.2023.02.146.
- [83] M. Casillo, M. De Santo, M. Lombardi, R. Mosca, D. Santaniello, and C. Valentino, “Recommender Systems and Digital Storytelling to Enhance Tourism Experience in Cultural Heritage Sites,” *Proc. - 2021 IEEE Int. Conf. Smart Comput. SMARTCOMP 2021*, pp. 323–328, Aug. 2021, doi: 10.1109/SMARTCOMP52413.2021.00067.
- [84] U. Pongsuppat, P. Jantarat, D. Kamhangwong, and S. Wicha, “Enhancing Local Tourism Sustainability through a Digital Local Tourism Management System (DLTMS),” *8th Int. Conf. Digit. Arts, Media Technol. 6th ECTI North. Sect. Conf. Electr. Electron. Comput. Telecommun. Eng. ECTI DAMT NCON 2023*, pp. 393–398, 2023, doi: 10.1109/ECTIDAMTNCN57770.2023.10139694.
- [85] N. Bai, M. Ducci, R. Mirzikhshvili, P. Nourian, and A. P. Roders, “Mapping urban heritage images with social media data and artificial intelligence, a case study in Testaccio, Rome,” doi: 10.5194/isprs-archives-XLVIII-M-2-2023-139-2023.
- [86] Y. Yin, “Research on the integration path of cultural creative industry and tourism industry based on collaborative filtering recommendation algorithm,” *Appl. Math. Nonlinear Sci.*, vol. 9, no. 1, Jan. 2024, doi: 10.2478/AMNS.2023.2.00551.
- [87] H. Niu, “The effect of intelligent tour guide system based on attraction positioning and recommendation to improve the experience of tourists visiting scenic spots,” *Intell. Syst. with Appl.*, vol. 19, p. 200263, Sep. 2023, doi: 10.1016/J.ISWA.2023.200263.
- [88] B. K. S. D. Santos, G. A. De A. Cysneiros Filho, and Y. A. Lacerda, “An approach to recommendation systems oriented towards the perspective of tourist experiences,” in *ACM International Conference Proceeding Series*, 2020, pp. 201–208, doi: 10.1145/3428658.3430977.
- [89] P. Banik, A. Banerjee, and W. Wörndl, “Understanding User Perspectives on Sustainability and Fairness in Tourism Recommender Systems,” *UMAP 2023 - Adjunct Proc. 31st ACM Conf. User Model. Adapt. Pers.*, pp. 241–248, Jun. 2023, doi: 10.1145/3563359.3597442.
- [90] L. V. Nguyen, “OurSCARA: Awareness-Based Recommendation Services for Sustainable Tourism,” *World 2024*, Vol. 5, Pages 471–482, vol. 5, no. 2, pp. 471–482, Jun. 2024, doi: 10.3390/WORLD5020024.
- [91] A. Panteli, A. Kompothrekas, C. Halkiopoulos, and B. Boutsinas, “An Innovative Recommender System for Health Tourism,” *Springer Proc. Bus. Econ.*, pp. 649–658, 2021, doi: 10.1007/978-3-030-72469-6_42/FIGURES/2.
- [92] M. T. Cuomo, I. Colosimo, L. R. Celsi, R. Ferulano, G. Festa, and M. La Rocca, “Enhancing traveller experience in integrated mobility services via big social data analytics,” *Technol. Forecast. Soc. Change*, vol. 176, p. 121460, Mar. 2022, doi: 10.1016/J.TECHFORE.2021.121460.
- [93] A. P. Darko and D. Liang, “A heterogeneous opinion-driven decision-support model for tourists’ selection with different travel needs in online reviews,” *J. Oper. Res. Soc.*, vol. 74, no. 1, pp. 272–289, 2023, doi: 10.1080/01605682.2022.2035274.
- [94] O. A. Ofem, M. A. Agana, and E. O. Felix, “Collaborative Filtering Recommender System for Timely Arrival Problem in Road Transport Networks Using Viterbi and the Hidden Markov Algorithms,” <https://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/IJSI.315660>, vol. 11, no. 1, pp. 1–21, Jan. 1AD, doi: 10.4018/IJSI.315660.
- [95] “A Development of Real-time Tourism Information Recommendation System for Smart Phone Using Responsive Web Design, Spatial and Temporal Ontology.”
- [96] L. Ravi, V. Subramaniaswamy, V. Vijayakumar, S. Chen, A. Karmel, and M. Devarajan, “Hybrid Location-based Recommender System for Mobility and Travel Planning,” *Mob. Networks Appl.*, vol. 24, no. 4, pp. 1226–1239, 2019, doi: 10.1007/s11036-019-01260-4.
- [97] M. Kovalchuk and D. Nasonov, “Hashtags: An essential aspect of topic modeling of city events through social media,” *Proc. - 20th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2021*, pp. 1594–1599, 2021, doi: 10.1109/ICMLA52953.2021.00255.

- [98] R. Alhayali, O. Hatem, and Z. Al-Dulaimi, "Image content based topological analysis for friend recommendation on twitter Image Content based Topological Analysis for Friend Recommendation on Twitter 1*," *Artic. J. Adv. Res. Dyn. Control Syst.*, vol. 10, 2018, Accessed: Jan. 14, 2024. [Online]. Available: <https://www.researchgate.net/publication/333043931>.
- [99] L. Zhu, J. Holden, E. Wood, and J. Gender, "Green routing fuel saving opportunity assessment: A case study using large-scale real-world travel data," *IEEE Intell. Veh. Symp. Proc.*, pp. 1242–1248, Jul. 2017, doi: 10.1109/IVS.2017.7995882.
- [100] C. H. Lin, J. Arcos-Pumarola, and N. Llonch-Molina, "Tourism safety on train systems: A case study on electronic word-of-mouth in Spain, Italy and Greece," *Secur. J.*, vol. 37, no. 3, pp. 1033–1059, Sep. 2023, doi: 10.1057/S41284-023-00405-1/METRICS.
- [101] M. E. Syahputra, S. Achmad, F. Fahrain, A. J. MacKenzie, F. Putra Panghurian, and A. A. Santoso Gunawan, "Smart Tourism using Attractive and Safe Travel Recommendation Technology," *2022 IEEE Creat. Commun. Innov. Technol. ICCIT 2022*, 2022, doi: 10.1109/ICCIT55355.2022.10118828.
- [102] K. Meehan, T. Lunney, K. Curran, and A. McCaughey, "Aggregating social media data with temporal and environmental context for recommendation in a mobile tour guide system," *J. Hosp. Tour. Technol.*, vol. 7, no. 3, pp. 281–299, 2016, doi: 10.1108/JHTT-10-2014-0064.
- [103] N. L. Ho and K. Hui Lim, "POIBERT: A Transformer-based Model for the Tour Recommendation Problem," *Proc. - 2022 IEEE Int. Conf. Big Data, Big Data 2022*, pp. 5925–5933, 2022, doi: 10.1109/BIGDATA55660.2022.10020467.
- [104] A. Harinivas, R. Bharathi, C. A. Gowda, P. Mohata, and R. Sharmila, "Knowledge-Based Medical Tourism Recommender System," *2023 IEEE 8th Int. Conf. Converg. Technol. I2CT 2023*, 2023, doi: 10.1109/I2CT57861.2023.10126286.
- [105] E. M. Kryukova, V. S. Khetagurova, L. V. Matraeva, E. S. Vasiutina, and N. A. Korolkova, "Features of the Sustainable Development of the Tourism Economy in the Context of the COVID-19 Pandemic," *Adv. Sci. Technol. Innov.*, vol. Part F1, pp. 85–90, 2023, doi: 10.1007/978-3-031-29364-1_18/COVER.
- [106] L. Liu, J. Xu, S. S. Liao, and H. Chen, "A real-time personalized route recommendation system for self-drive tourists based on vehicle to vehicle communication," *Expert Syst. Appl.*, vol. 41, no. 7, pp. 3409–3417, Jun. 2014, doi: 10.1016/J.ESWA.2013.11.035.
- [107] S. M. Millen, C. H. Olsen, R. P. Flanagan, J. S. Scott, and C. P. Dobson, "The effect of geographic origin and destination on congenital heart disease outcomes: a retrospective cohort study," *BMC Cardiovasc. Disord.*, vol. 23, no. 1, pp. 1–9, Dec. 2023, doi: 10.1186/S12872-023-03037-W/FIGURES/1.